ASSESSING THE QUALITY OF DEMOGRAPHIC DATA ON AGE AND SEX COLLECTED FROM CENSUS 2001, GENERAL HOUSEHOLD SURVEYS (2004-2007), LABOUR FORCE SURVEYS (2005-2007) AND COMMUNITY SURVEY 2007 IN SOUTH AFRICA

By

Germaine Kamleu

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Philosophy in Population Studies

UNIVERSITY of the
WESTERN CAPE

Supervisor: Professor Gabriel Tati

Faculty of Natural Sciences

Department of Statistics and Population Studies

University of the Western Cape

15 November 2012

Declaration

I declare that *Assessing the Quality of Demographic Data on age and sex collected from Census (2001), General Household Surveys (2004-2007), Labour Force Surveys (2005-2007) and Community Survey (2007) in South Africa*, is my own work, that has not been submitted for any degree or examination at any other university and that all the sources I have used or quoted have been indicated and acknowledged by complete references.

Germaine Kamleu November 2012

Signature……..

Acknowledgements

Glory be to GOD for all the courage, strength and his unconditional love and blessings.

I would like to extend my gratitude to my supervisor Professor Gabriel TATI, for his support guidance and understanding. I thank him particularly, for his perceptive comments throughout my research undertaking.

To my lovely father Jacques Kemajou Mbiatat, and my sweet love Bernard Ndouma who always motivated me to pursue my studies and become the best that I can be. Thanks for your prayers, for setting an example, for your encouragement and constant support. You made a big sacrifice over these years. The trust and confidence you displayed in me never faded.

My deepest gratitude goes to Dr Elisabeth Kemajou Teudjeu and Ms Olive Flore Mbianda Noukeu for funding my study.

I am sincerely grateful for having a loving and supportive family who never stopped believing in me – Bernard Ndouma, Vanessa Hilary Ndouma, Marylin Patrice Ndouma, Ledriss Brice de Doria Ndouma, Blaise Aurelien Ndouma.

My appreciation goes to Dr Isabel Schmidt, Mthombolwazi Dimbaza and MrsTasneem Solomons of the Statistics SA Western Cape for providing me useful information and guidance.

Many thanks to Clement Basuayi, Siaka Lougue, Joel Defo, Philomene Nyirasafali, Judith Ndonang, for their friendship and material support when I needed these. You made me enjoy Cape Town.

UNIVERSITY of the
WESTERN CAPE

Abstract

In many countries, an enumeration of all household members remains the most important source of population statistics. According to Statistics South Africa, two population censuses and quite a few household surveys have taken place across the country. The quality of data recorded varies according to the operation. Despite great improvement in data collection and analysis capacities, some of the demographic data provided have not been assessed in terms of quality. The aim of this study was to ascertain the accuracy of demographic data on age and sex collected and the coverage during the population census 2001, General Household Surveys (2004 and 2007), Labour Force Surveys (2005 and 2007) and Community survey 2007 in South Africa. Two methods were applied to assess the quality of data. First, the direct method consists of checking the content and coverage (errors during enumeration, errors of exploitation, concordance in questionnaire). Second, the indirect method lies in the calculation of some indexes, age ratios, sex ratios, graphing of population pyramids and sex ratios curves. The indexes are Whipple's index, Myer's index and the Combined index of United Nations. Therefore, the main variables of interest are age, sex, place of residence and ethnic groups. Differentials in the quality according to declaration on age by gender, by ethnic group, by place of residence have been explored. This study has identified some variations in different indexes between 2001 and 2007 and has also evaluated the ethnic, gender and regional differentials. Comparison between indexes of each instrument has been done to measure some variations over years. Also, time-space comparisons were conducted across indexes of different instruments. The quality of data on age was better at national level compared to provincial level. Therefore, based on the measurements and patterns observed in the census and surveys data, the study has made some recommendations on the need for an integrated approach to reduce the gap and improve the quality of declarations on age and sex.

Key words: Age heaping, Age pyramid, Age ratio, Combined index of United Nations, Myer's index, National household survey, Population census, Sex ratio, South Africa, Whipple's index.

List of Tables

UNIVERSITY of the
WESTERN CAPE

List of Figures

UNIVERSITY of the
WESTERN CAPE

List of Acronyms

AIDS	Acquired Immune Deficiency Syndrome
CIUN	Combined Index of United Nations
CS	Community Survey
DU	Dwelling Unit
EA	Enumeration Area
GHS	General Household Survey
HIV	Human Immune Virus
IMF	International Monetary Fund
LFS	Labour Force Survey
NGO	Non-Governmental Organization
PES	Post Enumeration Survey
PPS	Probability Proportional Sampling
PSU	Primary Sampling Unit
QLFS	Quarterly Labour Force Survey
SA	South Africa
SPSS	Statistical Package of Social Science
Stat SA	Statistics South Africa
UN	United Nations
UNESCO	United Nations Educational, Scientific and Cultural Organization
USA	United State of America

UNIVERSITY of the
WESTERN CAPE

Figures in appendices

Table of Contents

UNIVERSITY of the WESTERN CAPE

UNIVERSITY of the
WESTERN CAPE

UNIVERSITY of the
WESTERN CAPE

CHAPTER ONE

INTRODUCTION

1.1 Background to the Issues investigated

Population statistics for many countries in Sub-Saharan African are derived from a limited data. Few of these countries have empirical data available from a recent census, and several countries have postponed their 2000 round census for a later date (Velkoff & Kowal, 2007). This applies even to the countries which do have recent census data, as the quality has been uneven. Few countries in Sub-Saharan Africa have functioning vital statistic systems which produce usable data on fertility and mortality (Rao *et al.*, 2004; Timaeus & Jasseh, 2004 cited in Velkoff & Kowal, 2007). The lack of vital statistics and gaps in census and survey data are not properly used to produce estimates of development processes and planning of activities which integrate population variables (Cohen *et al.*, 1996 as cited by Roos *et al.*, 2005). These shortcomings have serious implications in policy implementations. In the context of Africa, age remains a variable of major concern. Vital registration systems have been restricted because of many difficulties in such countries (Heron *et al.*, 2010).

UNIVERSITY of the
WESTERN CAPE

In countries with well-developed civil registration systems, census and survey data can be successfully used together with data from administrative records; it was the case in 1990 population census in Singapore. Enumerators had pre-filled basic information from administrative records, for every member of the household. This approach reduced interviewing time, enumeration costs, and some misreporting (Leow & Koh, 2001; UN Statistical Division, 2008).

Censuses in Africa have been characterized by various anomalies. However, there are some countries which have effectively overcome these. In Kenya for example, a census was first conducted in 1948 when it was still a British Colony. Since 1969, censuses have been conducted every ten years. The last one took place in 2009 and it is the first country in Africa to produce a completely processed census within one year after the census (http//www.ubos.org/). In the same vein, Uganda has been carrying out population and housing censuses under standard norms at intervals of about ten years since 1948. The latest of such population census which was

conducted under the supervision of the Uganda Bureau of Statistics in 2002 was the most complete census ever accepted in Uganda (http//www.ubos.org/).

South Africa has a catalogue of census irregularities. The first census of the Union of South Africa was conducted in 1911. Several enumerations occurred, but the Black population was not accurately counted in any of them. In 1950, the black population was assigned to live in homelands which represented only 13 per cent of the land which was excluded from the official census of South Africa (Khalfani *et al.*, 2005). Restricted in informal settlements, the Black population avoided to register births, deaths, marriages and divorces because of complexity of the laws concerning legal residency. The 1980 census count was nearly 23.8 million; 4.6 million were added to compensate for undercounting (four independent homelands were excluded: Transkei, Ciskei, Bophuthtswana and Venda) (Robinson, 1993; Wittenberg, 2003). The 1991 census took place and the government used aerial photography and sample surveys to enumerate residents in 88 inaccessible areas to officials. After being adjusted for under enumeration, the government still excluded the four independent homelands.

In 1994, the South African government estimated the total population at 40.4 million, after all ten homelands had been incorporated into South Africa. Since 1994 Statistics South Africa has undertaken two population censuses and multiple household surveys, Labour force surveys across the nine provinces of the country.

The 1996 census in South Africa represents the first attempt to count the population of post-apartheid South African and to obtain essential planning information through the census process. Three important issues arise from this process which facilitates preparation of the 2001 census; these are undercount, improving coverage and ensuring the quality control (Stats SA, 2003). The government always tried to respect the periodicity of these operations. The quality of data recorded remains controversial (Stats SA, 2007). In terms of method of data collection on age, the question asked from respondents was: “what is (the person’s) age?” It was not clearly understood by many people. Some respondents tended to round their age and others declared their age in completed years which introduce multiple errors during the collection. This subsequently affected the age distribution.

From 1996 to 2010, the question on age has been refined and they are two important questions merging in one which try to capture the exact age of the respondent: "What is (the person's) date of birth and age in completed years? (Dr Isabel Schmidt, interview on 13/06/2011; refer in appendices). Despite great improvements in data collection and analysis capacities, the reliability of the demographic data provided is always a matter of public debate and statistical uncertainty. Some respondents are not able to remember their date of birth because of the insufficiency in the delivery birth certificates. For this special case, during the enumeration there, the use of important events (historical events) helps respondents to estimate their date of birth if no documentation is available (Stats SA, 2007; Khan, 2010).

South Africa has however committed itself to improving the quality of the demographic information about its people; but the lack of awareness of registration procedures among the general population about the necessity or importance of registration and the lack of registration services in many areas of the country (rural in particular) are also contributing to the quality of age reported (Stats SA, 2003).

At the lower administrative level, reliable data is the major challenge for provinces in South Africa. The limited information for programs and plans of administrative work due to no updated size and characteristics of population after five or ten years used to take long. The access and the quality require more attention. The extent of change is engaged to bring such improvements through reliable census, labour force survey and general household survey. Technological instruments have been increasingly introduced to improve the efficiency of data collection operations.

The government is worried about the discrepancy of the results of these censuses and surveys since it is only through credible results that can enable it to describe the exact demographic situation in the country at a given point in time. The absence of good coordination during the census operation or survey procedure might result in the production of different statistics which are inconsistent. An aspect of concern remains the quality of training of fieldworkers. The performance of enumerators might affect also the quality of data collected during census and survey. Age and sex are affected by age misreporting and wrong declarations which are not always properly detected by enumerators (Stats SA, 2010).

1.2 Motivation for the Evaluation of Quality of Data on Age and Sex

1.2.1 Importance of Data on Age and Sex distribution

Census and surveys, monitoring of governments schemes and planning for the future are the basis for reviewing a country's progress every decade, given that they each contribute in providing valuable information for the formulation of policies and effective public administration. The distribution of a population by age and sex is very important for socio-economic and demographic considerations (Poston, 2005) as it provides the basis for deriving most of the demographic indicators, population estimates and projections required for assessing population change and its dynamics. Age and sex composition are closely related to the provision of social services. For instance, a rapidly growing population is likely to experience problems associated with the need for increased school facilities, employment creation or health facilities.

UNIVERSITY of the
WESTERN CAPE

The age and sex data is used by private and public sector, national and international agencies, scholars, business people, and a host of other stakeholders. Age and sex represents the basic type of demographic information collected about individuals in censuses and surveys. This type of data describes the characteristics of the population. In a number of countries, the size of the population determines the allocation of elected political seats in government. The number of elected officials for each governmental administrative unit is determined by the population size of a given locale. For some countries, the information is also used in the allocation of government resources for developmental efforts. In addition to household composition and size, the information determines the needs of different segments of the population.

Age and sex are essential in studies of different age and sex segments of the population, such as children, youths, adults, and the elderly people. This data helps to analyze socio-economic situation of these groups and identifies segments of the population that require different types of needs with regard to different issues such as education, labour force participation, health and living conditions. In terms of education, employment, and migration, there are significant gender variation for different age groups. The differences in age and sex also affect policy formulation, planning and implementation. Sex ratios can be calculated by 5-year age groups to observe

migration, among the working age cohorts in different regions in the country. The data is classified by sex to easily help detect female and male roles in the society.

Moreover, the Platform for Action of the fourth World Conference on Women recommends the presentation of data by age and sex to reflect problems, issues and questions related to women and men in society for utilization in policy and program planning and implementation(Diane Publishing company, 1996; UN, 1996). Studies have shown that age and sex structures are useful for evaluating trends and for assessing demographic and economic impacts (Merry, 2005). A wrong declaration of age or error on sex affects the population size, characteristics, and well-being. The age pyramid for example, is an important tool used by demographers to locate the country's position in the trajectory towards demographic transition (stable population). Frequent and persistent errors in age declaration cause serious impediments ineffectively making use of the age pyramids. Therefore, efforts are needed to present data by age and sex in great details to maximize their relevance. Various studies rely on the same data sources; however, censuses and surveys can produce different results due to differences in the timing of data availability and judgment. According to basics, data should be adjusted in estimation techniques and choice of models.

UNIVERSITY of the

The main approach for evaluating data is employed to calculate demographic indexes using data from census under study or from additional demographic sources (Brass, 1996). The United States Census Bureau uses the census results by applying demographic techniques to detect errors in population distribution by sex and age. These techniques involve analysis of digit preference, population pyramids, age ratios and sex ratios in base population data. In South Africa, the practice is quite different when it comes to assessing the quality of data on age distribution. The statistical agency (Stats SA) does not make use of the conventional techniques recommended in the field (Indirect methods). It rather relies only on such procedures as Post Enumeration Survey (PES) as will be discussed in the section that follows.

1.2.2 UN Recommendations for Improving the Quality of Data

The weaknesses and irregularities of the earliest population censuses in the world, the disparities and differences in the methodology used to collect and process its data all contributed to increase

the international non-comparability between data. The statistical commission of the United Nations requested the need for develop international standards in the world's programme on population and housing censuses. However, the UN Statistics Division has coordinated the international recommendations, giving technical assistance to countries or areas in need.

The last publication of global census recommendations occurred in 1998. The UN secretariat has published "Economic Commission for Africa, African supplement to the UN principles and recommendations for population and housing censuses, revision 2" (UN, 1996). For 2010 world program, the Statistics Commission requested the UN to continue with the revision and update of the guidelines that all countries need to follow in order to achieve the production of detailed statistics for small area domains, at least once around the year 2010.

The UN recommended that all countries have to make great strides in improving every stage of the census (Dekker, 2001). This study will highlight only those areas where there are gaps. In South Africa, these areas are the mapping, codification, questionnaire design, classification of questions, training, recruitment, organization of fieldwork, skills of staff, publicity campaign, census costs, methodology of evaluating the accuracy of data, the techniques to measure the change over the years, the limitation of Post Enumeration Survey (PES).

UNIVERSITY of the
WESTERN CAPE

In regard to cartography, the recommendation is that it must be done back from a deadline of reproducing the last map three months before the census day. Additional time may be given to attend the need to use these documents effectively. Also, attention must be given to have permanent identification on streets and buildings in such a way that any street that is supposed to have a unique name and building has a unique code. Every single address must be clearly defined to avoid confusion. The type of questionnaire and the method of enumeration can affect the quality and amount of data to collect. The request for data by users will be made within their considerations. The questionnaire design should consider the format, content and arrangement of questions and sensitive topics must be excluded while questions should not be offensive.

Furthermore, costs effective strategies need to be emphasized that would prevent the quality of data. However, the governmental department, non-governmental organization and the private sector users should designed in partnership with all political actors their involvement in census procedure. This strategy would ensure the legitimacy and the advocacy for sufficient funding.

The costs for each stage of the census must be improved in certain manner that the choice of the appropriate technology must be applied. However, several costs tend to increase in most countries. The UN had mentioned that the credibility to a census process increase with a clean outcome from a financial audit which allows the government and civil society to likely accept the final results and also promote the future improvements (UN, 2000).

The recruitment of personnel and training purpose has to tackle language issues. The pilot census should run exactly one year before the planned census in order to conform to the expected seasonal patterns of climate and activity. Therefore, a single enumerator should be able to cover the enumeration area unit during the time allowed for the enumeration (Leggieri, 1994). This enumeration area unit should be well defined and small. When dealing with nomadic and semi-nomadic population, great attention should be given to the difficulties of locating such unstable population groups. In the same vein, some arrangements need to be done for homeless individuals.

The UN also recommends that selection of the personnel should be done on the basis of competence knowing that the preparation and processing of work requires individuals with skills (cartographers, coders, data entry operators, programmers, drivers, and so on) while the enumerators need the physical aptitude to reach enumeration areas (EA) and collecting the information according to specific definitions and instructions. More effort should be given to improving the ability of personnel to achieve field operations effectively by knowing the workload limits of enumerators and ensuring that EA boundaries are designed so that they easily follow identifiable features such as roads, waterways, established walking tracks and railway or power lines. The use of features such as village or local government boundaries should be carefully considered, taking into account the difficulty of identifying boundaries using features such as compass bearings or lines of sight.

The timing of the census is advisable only during school holidays; the enumeration period must be short and with a single reference date. The teachers and demographic students should be involved in the process of collection. In some countries, the holidays are prolonged to permit teachers to achieve the set targets.

The UN recommends that an educational campaign must be done with the purpose to interest the general public and obtain its cooperation. The aim must be clearly defined not only to dissimulate any anxiety regarding the purpose of the census but, to provide the reasons for questions in the questionnaire. It will also offer some guidance as these questions should be answered. The publicity may increase the completeness of census coverage. The recommendation also includes logical imputation based on other information for the individual or household, or for the other individuals or households in case sufficient information is unavailable for the specific individual to correct apparent errors; the hot deck imputation method also can be used.

The quality assurance and improvement program provides further recommendations relating to controlling and assessing the quality of census operations. Even if data are accurate, they do not have sufficient quality if they do not fit some requirements. For instance, the production of data is too late to be useful, or cannot be easily accessed, or conflict with other credible data, or are too costly to be produced. These requirements are: relevance, completeness, accuracy, clarity, comparability, coherence, timelines, punctuality, accessibility, availability of information describing sources, definitions and methods. The assessment of accuracy and completeness of data can be issued after the initial census results are disseminated. The evaluation should also be undertaken, by comparing the census results with similar data from other sources. These sources can include surveys in a similar time frame or previous census results. However the process of evaluation should not contribute to delay the dissemination of data.

UNIVERSITY of the
WESTERN CAPE

In addition, the UN recommends two important methods for evaluating census data: demographic analysis and post enumeration surveys (UN, 2008). The advice is to apply both methods to complete information or the overall assessment of the census quality. The demographic analysis examines the overall methodologies used during the census process which consist in determining the graphical analysis of the population pyramid (age-sex distribution). Any census should include the variables age and sex so that, a comparison of the age pyramid and sex ratio for each 10 year age cohort would be basic elements. Age heaping or the tendency of respondents to report a particular digit ending is a useful internal consistency check, as are sex ratios by age and certain summary indices of age and sex data. This includes the United Nations age sex accuracy index which extends age sex ratio analysis by observing deviations of the

observed age gender ratios from the ones expected for each five year age group and combining the results into a single score. These indexes are Whipple's index, and Myers's blended index used for judging age heaping. The study will focus on the calculation of sex ratios and these indexes.

Indeed, the UN mentioned that these techniques are useful and provide overall assessment of census quality but they cannot change the sources of census error in terms of contributions from under coverage, or content error. The information about coverage error is derived from comparative analysis of data from successive censuses by using one of the following methods:

(a) Derivation of an expected population estimate taking account of vital registers of births, deaths and net migrants between censuses, as compared with the latest census;

(b) Cohort component method which the population projections are based on the results of the prior census plus data on fertility, mortality and migration from various sources and comparing the projected estimates with the new census results;

(c) Comparison of two census age distributions based on inter-censal cohort survival rates;

(d) Cohort survival regression method in which the estimates of coverage correction factors used regression methods to make the age results from the two censuses mutually consistent (Siegel, 1980).

UNIVERSITY of the
WESTERN CAPE

It is better to mention that the first two methods tend to be restricted to the evaluation of studies of coverage at the national level, in countries which do not have good sub-national data on migration as in the case of South Africa.

The demographic analyses for census use a strong procedure for evaluating the quality of a census. However, South Africa has applied the second method, the Post Enumeration Survey (PES) with a single system estimation procedure for estimating the true total of population; the coverage error is an undercount. Several reasons explaining the choice include the enormous costs, the timelines to produce all the analysis, the scarcity of experts, the absence of qualified technicians in the field, the policies in place. The UN had defined the objectives of PES as follows: first, to assess the degree of coverage during census enumeration; second, to examine the implications of coverage deficiencies; third, to obtain information for the design of future

censuses and surveys and lastly, to examine the characteristics of individuals who may have been missed during census enumeration. PES is a survey designed to measure census coverage and/or content error. As South Africa, many countries have applied this method in recent decades.

However, PES provides a comprehensive evaluation of coverage and content error when supplemented by and integrated with detailed demographic analysis of census quality. PES is carried out within a few months of the census to measure that the impact of national population changes (births, deaths, and migrations) and gaps in respondents recall.

1.3 Statement of the Research Problem

Age and sex structure is determined by demographic components of fertility, mortality and migration. The use of these variables may not be without difficulties. The difficulties of the above components may arise during data collection. Therefore, the South African instruments do not escape the rules to determine the accuracy of age and sex of the population for the different demographic components. It is in this context that this study investigates to see if the method applied to assess the accuracy of age and sex in the South African instruments provide reliable data. And to compare them with other indirect methods such as Whipple's index, Myer's index, Combined index of United Nations, Age ratios, Sex ratios and Population pyramids. This operation of comparing the indexes is carried out is in order to bring out the differences and similarities in terms of errors, omissions or duplications. Very little is known on the extent to which data on age distribution derived from a population Census; and how it differs from similar data collection like the type of data accumulated from Labour Force Survey. In the same vein, the differences across the provinces and multi-ethnic groups have been rarely investigated within the broad range of data collection operation. This could explain the complexities and dynamics emerged from these areas of empirical research.

1.4. Research Questions

The general research question investigated in these areas of my study revolves around the conception of age and sex, and the accuracy of empirical data collected in the censuses and/or surveys.

The general question of this study is,

How accurate are the age and sex data collected in censuses and/or surveys?

It is however divided into four specific questions:

a) What are the differences in the quality of data on age and sex at provincial and national levels or do we observe any similarities in indicators of age quality at all the geographical levels?

b) Over years, has the quality of data on age and sex been modified for censuses or surveys across gender?

c) To what extent has the quality of declaration of age depended on the type of operation? Do we observe for example better quality for a specific type of survey than for other?

d) How do the differences of such variables as gender and ethnicity associate with the quality of declaration on age?

1.5 Hypotheses

UNIVERSITY of the
WESTERN CAPE

In line with the specific research question above, the following hypotheses have suited the study at varying degree.

a) The quality of declaration of data on age and sex is poorer at provincial compared to national level.
b) The quality of data on age and sex collected from censuses and surveys has improved over years across gender for each operation type (instrument).
c) There are differences across the racial groups. The quality of declaration is poorer among the historically disadvantaged groups (Blacks and Coloureds) than among the two other groups (Whites and Indians).
d) The poor quality of declaration of age is better among males than females.

1.6 Objectives /Aims of the Study

I.6.1 Main Objective

The overall objective of the proposed study is to evaluate the accuracy of the quality of the data collected in censuses and surveys on variables such as age and sex. From this evaluation, the study aims are to demonstrate the data on the age and sex distribution; and assess the quality according to the specific type of operation.

1.6.2 Specific Objectives

The study investigates four specific objectives.

The first objective ascertains the level of quality of data on age and sex using Whipple, Myers, Combined Index of United Nations and other methods at specified dates of data collection.

The second objective is to compare for each operation, the improvements achieved across time and space (national and provincial levels).

The third objective is to identify which operation gets better quality of data.

UNIVERSITY of the
WESTERN CAPE

The fourth objective is to analyze the variations in age declaration associated with such variables as gender, ethnic groups (race) and administrative level.

1.7 Scope of the Study

The geographical coverage is both national and provincial. The evaluation considers the data collection operations conducted in the period 2000 to now.

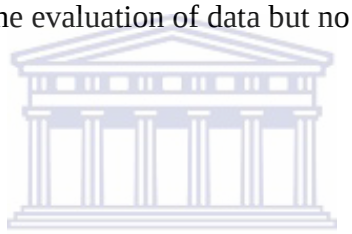
1.8 Data and Methods to be Used and Justification

- Population Census, General Household Surveys, Labour Force Surveys, and Community Survey.
- Non adjusted data

- Several operations in order to assess the similarities, differences and draw lessons for improvement.

1.9 Limitations

- The study cannot detect the urban and rural differences because of the absence of this variable in most operations. In the same vein, the non-available province variable in Census 2001 data sheet and ethnic group in Community Survey 2007 data sheet made impossible to identify differences or similarities at various stages of the study.
- The approach used is inherently dependent on the strengths and weaknesses of the methods used.
- The study was focused on the evaluation of data but no adjustment was done.


UNIVERSITY of the
WESTERN CAPE

1.10 Definition of Keywords

Age Heaping: A tendency of the respondents to declare a certain age at the expense of the truth age.

Age Pyramid: A graphical illustration that shows the population age distribution broad at the base (younger ages) and narrow at the top (older ages).

Age Ratio: A given age group to one-half of the sum of the populations preceding and following group times 100. If there are no external variations in the past events, the age ratio for all age groups should be approximate to 100.

Combined Index of United Nations: The United Nation's age-sex accuracy index that combines the mean deviation of the age ratio for male, the mean deviation of the age ratio for female and three times the mean of the age to age differences in the reported sex ratio. It is computed for five years age group up to 70.

Myer's Index: An index that measures the quality of age declaration or reported for digit ending 0 to 9. Myer's index detects avoidance or preference in the accuracy of responses reported.

National Household Survey: A Survey organized at the national level which aims at capturing living conditions of household respondents.

Population Census: A census is one of the largest single sources of information on the life of all individuals in a given society at a given time. This is an operation in which questions try to capture various sets of information in terms, so as to evaluate the living condition of respondents in the country. It is a total process of collecting compiling, analyzing, evaluating and publishing demographic and socio-demographic data at a specified time involving all persons in a country or in a delimited part of the country.

Sex Ratio: Number of males per 100 females in a given population.

South Africa: Nation of diversity, the population of South Africa is estimated at *51 770 560*according to the result of 2011 census data. The country is divided into nine provinces: Eastern Cape, Free State, Gauteng, KwaZulu-Natal, Limpopo, Mpumalanga, Northern Cape, North West and the Western Cape. Around thirty per cent of the population is aged less than 15 years old indicating that South Africa has a young population.

Whipple's Index: An index of concentration which aims at detecting the extent to which age data shows systematic heaping at ages ending 0 and 5. It is widely used because of its simplicity.

UNIVERSITY of the
WESTERN CAPE

1.11 Thesis Outline

The study is subdivided into six chapters. Chapter 1 provides an intriguing and profound introduction to the research; starting with the background to the study and the importance of data on age and sex. It outlines the UN recommendations and presents the statement of problem, research questions, objectives, and hypotheses. The scope of the study is defined and the justification of data and methods to be used are stated also in this chapter. In chapter 2, the importance of data quality on age and sex distribution is given, UN recommendations, methodologies to be used to ascertain the quality of data distribution, techniques for evaluating age and sex composition, analysis of deficiencies in age and sex data using Whipple's index, Myer's index, CIUN, reporting for extreme old ages and centenarians, age not reported, age

ratios, sex ratios and population pyramids, inspection of some examples of evaluation and progress made in the methodologies are also provided. Chapter 3 explains the research design, sampling and data collection, description of variables, method used in analysis and limitations in the comparisons. Chapter 4 provides the results of the data analysis. Chapter 5 discusses the findings. The chapter 6 concludes and extends some recommendations from the study to the policy makers.

UNIVERSITY of the
WESTERN CAPE

CHAPTER TWO

LITERATURE REVIEW

2.1 Introduction

Censuses and surveys are the basis for reviewing a country's progress, monitoring the ongoing schemes of the government and most importantly, plan for the future. They provide valuable information for planning, formulation of policies, and effective public administration. The data from these sources are used by national and international agencies, scholars, business people, private and public sectors, and many more. Age and sex represent the basic of demographic information collected about individuals in censuses and surveys. The study focuses on census 2001, General Household Surveys (2004 and 2007), Labour Force Survey (2005 and 2007) and Community Survey 2007 in South Africa. Therefore, this review presents firstly the importance of the quality of data on age and sex distributions; secondly, it examines United Nations (UN) recommendations with the methodologies used to ascertain the quality of data on age and sex distribution and at the last stage, it dwells on the inspection of some examples where the assessment of data on age and sex distribution are engaged (indirectly or statistical office) and the progress made in the methodology.

UNIVERSITY of the
WESTERN CAPE

2.2 Importance of Quality of Data on Age and Sex Distribution

Age and sex are the most basic type of demographic information in any census or survey operation. Many studies of population in Africa using census and survey data reveal the unreliability of the data on age and sex; though they tend to use them without further assessment. The various uses of age data recommend commonly the data on age to be collected according to gender. The age misreporting affects acutely the quality of age data.

Sex is the best fundamental differentiation in humans, since its distinction is based purely on biological organs. It is useful to have information about the extent to which things are different for women and men. Considering the factors of population change in fertility, mortality and migration, it is useful to have separate data for females, males and age groups, to have information on age and sex at all levels of government to implement and evaluate programs

according to the age and sex structure of the population. Such data classified by age and sex enhance the ways to analyze social and economic authenticity of these groups. Some issues have to be considered as labour force participation, women's educational equity, health services, education of children and young adults and so on.

Feminist movements have recently reinforced political issues at many political arenas. The platform for Action of the fourth World Conference on Women in 1995 (Beijing, China) recommends the presentation of data by age and sex to reflect problems, issues and questions related to women and men in society for use in policy, program planning and implementation. Sex refers to the biological and physiological characteristics that define men and women. The population data classified by sex highlight easily female and male roles in the society. The imbalance in the number of women and men can affect the labour force participation and the sex role in the society (Gupta, 2006). For instance, the analysis of military manpower requires separate information of men and women.

Age data are needed for calculating estimates and building projections in public and private sectors (school, health services, labour force, food, housing, etc.). Studies have shown that age and sex structures are useful for making population projections at national and provincial levels, evaluating trends, and for assessing demographic dynamics and economic impacts on the living conditions of its people (Haslett & Jones, 2008; Ansari, 2008). The case of Philippines is an example that age distribution for 1990 census presents some distortions and further investigation reveals some forms of misreporting with strong preference for digits 0 and 1 (Opiniano, 2004). A sex distribution is useful for effective analysis of different types of data.

Age and sex data contribute to support a variety of legislative and program requirements. These data are useful in allocating for example, funds from health programs, in particular to programs targeting specific age groups. For instance, age data are used to calculate the proportion of school aged children in each province in order to properly allocate funds for education. Considering old population, data on age provide information such as measurement of people eligible for social security and Medicare beneficiaries. These data can be applied in the research field when analyzing trends related to mortality for instance, population aging.

The quality of census data on age in studies of changes in population is even appreciated when adequate vital statistics from a registration system are not available (Mathews *et al.*, 2011). In the case of South Africa, the vital registration system is not completely achieved as in several countries in Sub-Saharan Africa. Many errors arise in their data during collection and processing due to the lack of registration services in many areas (rural) of the country, and lack of awareness among population about the importance of accurate data and its contribution to planning implementation (Cutler *et al.*, 2006).

Some studies (e.g. Setel *et al.*, 2005) demonstrate that failure to make registration compulsory, the lack of qualified staff (vital registration) and the non-reporting of births constitute contributive factors to the misreporting of population. Rustein (2008) outlines the fundamental distinction of coverage errors and misreporting errors. However, age and sex have a vital role in population studies.

2.3 The United Nations (UN) Recommendations and Methodologies to be used to ascertain the Quality of Data on Age and Sex Distribution

UNIVERSITY of the
WESTERN CAPE

The UN provides recommendations to ascertain the quality of data on age and sex as these are critical variables on demographic studies (Poston, 2005). This section therefore highlights the recommendations and methodologies to be used in research that deals with these variables.

2.3.1 Recommendations on Age and Sex

Age is a more complex demographic characteristic than sex. The United Nations recommendations define age as the interval between the date of birth and the date of census expressed in completed years (Poston, 2005). Age data collected from census or survey can be tabulated in a single year of age, 5-years age groups, or broader groups (Shryock *et al.*, 1980). The UN recommendations for population and housing censuses insists on the tabulation of the national total, urban and rural population, for each civil division, for each principal locality, in single years of age to 100. However if any particular geographical area is not possible to tabulate in single age, the age data should at least be tabulated in 5-year age groups, by sex. Data on sex are collected by asking each person to report either being female or male. The UN also

recommends that the principal tabulations of birth statistics for geographic areas within countries be made according to place of usual residence (Becht *et al.*, 2003).

However, to process direct reports on age is easy to highlight inaccurate information on age than reports on date of birth, hence no valuable prove (authentic document). Questions on age are rarely fully addressed because most answers are approximations. Women often forget births that occurred in the distant past and make systematic errors when estimating events (Dandona *et al.*, 2006). Prior to 2009, the missing date of birth information is not imputed for GHS in South Africa. It practices limit age imputation based on the reported age of the partner or spouse whether married or living together in the household. This is done using the age of the parents in the case of the missing child age information (Stats SA, 2009).

In Sub-Saharan countries, the proportion of population which unfilled age reports is high because of the existence of illiterate population. These errors may result from incorrect recording of responses during enumerations; misunderstanding of questions on age, mistakes during data entry, or in our context, respondents not knowing their exact age signifying that under enumeration in censuses is largely caused by mis-reporting (Adam, 2007). Misreporting has serious implications on the estimates in demographic issues, as well as heaping age of respondents. Froen *et al.* (2009) prove that in South Africa, only 50 per cent of deaths and also 18 per cent of births in the first year of life are registered, for example. This reveals the non-consistency of the civil registration system in the country.

The United Nations recommends that attention must be given to the date of birth of children reported as 1 year of age by asking normally their date of birth and the rest of the population have to complete the question on age. Respondents may also be assigned to form age groups on the basis of birth or some historical events affecting the population; this is the case in South Africa (Joubert *et al.*, 2012). The major problem relating to the quality of data on sex collected in censuses reside in the difference observed in the completeness of the coverage of the two sexes. Malley *et al.* (2007) and Lester *et al.* (2010) assert that parents may report boys as girls so they may avoid the attention to be overlooked when their cohort is called up for military purposes. These factors contribute to differentials in the enumeration of both sexes. The problem of identification is acute where the coverage of annual industrial surveys in Africa has been extended to small-scale industrial activities (engaging 10 or less persons). These activities

exclude from the coverage of industrial surveys mainly because they are difficult to locate. Techniques like area sampling could be used to solve the lack of information and technique for managing the survey especially in small-scale activities (Adam, 2007). Some countries in Africa experience that a sampling unit may have different names when used in different contexts and the names may change over time or differ in spelling. When evaluating estimates of omissions and counting errors for the whole nation or for large geographic areas, many of these errors cancel out a significant number of people who are included in the number of omissions because they are missing at the correct location. They are also included in the number of counting errors because they are counted but at the wrong location. Anderson & Fienberg (2002) argue that about 40 to 50 percent of counting errors represent people who are also counted as omissions or duplications for which the two classifications balance out. However, for small areas, these kinds of geographic location errors may make a difference.

Household mobility or migration of sampling units is a complicating factor in the execution of African censuses and surveys. Mobility tends to group into three categories as follows:

The first group is that of Nomads: they are people who move together with their animals in search of new pastures and water points. It presents serious problem of enumeration of the people in many countries of Africa (Sahel), Sudan, Ethiopia, Mauritania, Somalia, Lesotho, northern Cameroon, etc. The United Nations have brought different methods to capture the data but the result is not satisfactory and the problem remains unsolved (Zeleza, 1997; Adam, 2007).

The second group represents international and internal migrants (mobility in search of opportunities; education, health and work). Notably rural-urban, rural-rural, and urban-rural movements of people pose great problems; even seasonal labourers contribute to errors in counting of people. These are the characteristics in countries like South Africa, Botswana, and Malawi. Statistics SA suggests that this pattern of under enumeration reflects different levels of urbanization and difficulties in achieving comprehensive coverage in rural areas (Stats SA, 1998a).

Shifting cultivation experienced among sedentary population: In countries facing this situation, it is easy for holders to abandon old and non-productive fields for new ones so that

sampling frames of fields become outdated. This problem occurs in the northern parts of Uganda and in the northern parts of Kenya.

A post enumeration survey in South Africa in November 1996 indicates that the undercount in the census varies by province. According to Statistics SA, infants and young adult men are particularly prone to under enumeration. On the other hand young women are more geographically mobile than younger girls or older women (Wright *et al.*, 2009) and as a result they are also often missed in censuses. Arnab & Serumaga (2006) argues that new migration patterns develop as a result of HIV/AIDs where adult children succumbing to the disease move from urban areas to rural areas to be cared for by their parents and families. In the same vein, Richter & Desmond, (2008) postulate that HIV/AIDs related deaths are contributing to the disintegration of households, resulting in orphaned children being forced to relocate in rural areas. People tend to be undercounted in population census all over the world; but undercounting is not the only problem in population censuses. People are also erroneously included, but the net effect generally is undercounting.

In the 1996 census, South Africa experienced patterns of undercount by age that is common in other parts of the world. One may have the true age of a respondent reported. However, in other situations, it is hard to define. For example a sick person may rate himself/herself fit depending on the circumstances. Under-enumeration is also linked to census fieldwork and procedures related to complexity in social transactions among individuals (Welzel & Kligemann, 2003). In reality, the value reported or observed is always irrespective of who reports it and under what circumstances it is obtained. In some category, there are omitted people from the census in their residences, enumerated people are more than once, babies born after census, people whose existence are ascertained but whose characteristics are missing and so are taken from another enumerator records (substitutions). The combination of erroneous and substitutions can be added together to produce the total number of counting errors, for example, people who miss the census or are counted at the wrong location (omissions). Citro & Kalton (2007) estimate the net undercount in the 1990 census of USA to be about 4 million which is quite close to the estimates of 4.7 million on the basis of demographic analysis.

According to the UN, the first population census in the USA for instance was taken in 1770, but it was in 1960 that demographic analysis is applied to verify the results of the population census.

Strong digit preference occurs in the reporting of ages (UN, 1964). Phillips *et al.* (2003) set out that the quality of the 1970 census data for blacks is not as good as in the 1996 census in South Africa. Arnold (1990) notes errors in the coverage based on the wrong declaration of children ages. He also observes that many women ages have on digits ending in 0 and 5 which signals that they tend to round their ages. Blacks and Coloureds are less likely than Whites and Indians to be enumerated. Statistics SA suggests that this pattern of under enumeration reflects different levels of urbanization and difficulties in achieving comprehensive coverage in rural areas (Stats SA, 1998a).

Wittenberg (2004) claims that the surveys are badly designed to pick up mining employment. The fact that the census picks up much more employment in construction and in private households creates more problems for the quality of the household surveys than for the census. It is also hard to prove how workers in the manufacturing sector could erroneously place into these categories (Robbins, 2005). Instead, one can envisage how the household surveys may miss some domestic workers and small-scale construction workers. Robbins (2005) discovers that individuals with stable jobs are more likely to start families and become head of households than other individuals. This should increase rather than decrease the probability of these individuals becoming captured in the census or in standard household surveys. When a worker is illegal in the country, the employer tends to keep him informally. However this alone cannot explain the large differences between the survey and the population census. In a study by Klassen & Wooland (2001) cited in Wittenberg (2004), manufacturing workers are captured in the census but not recorded as being manufacturing workers.

UNIVERSITY of the
WESTERN CAPE

Roberts (2005) demonstrates that the unemployment issue that dominates in the apartheid era continues to affect demographic data and also explains the overlaps occurring between census data and survey data. Moller (2007) investigates in South African shack settlements on the periphery of urban centre populate by jobseekers increase while job opportunities do not keep swiftness with demand.

For the 1973 population census of Sudan, the census cartographic units use data on listings which is incorrectly done; hence, the demarcation of zones is also inaccurate (Taha, 2001). It is common experience that African census fails to provide adequate addresses of sampling units especially in the rural areas where housing units are not numbered and where any numbering

during a census operation is invariably too temporary (Adam, 2007). An adequate or accurate sampling frame in Africa cannot be over emphasized because it is rare to find adequate lists of households as sampling frames, even updating old frames (Session IV UN, 2007). Cartographic work is inaccurate in many cases either because maps are not accurate or the existing documents are not reflecting the reality. In some countries in Africa settlements are dispersed; for example, in forests in central Africa and in mountains of Ethiopia where physical difficulties are recurrent, settlements become removed from communication networks and other social services so that it turn out to be difficult to construct complete sampling frame and to investigate enumeration areas systematically (Adam, 2007).

The recruitment of interviewers and supervisors require a selective method. In the context of Africa, the interviewers are limited or some of them have no prior experience. Jones *et al.* (2010) suggest that appropriate training has to be achieved with particular challenges in various settings; every single question must be explained to avoid difficulties on the field of duty. Most of sub-Saharan Africa countries adopt the metropolis language as their official language (English, French, Spanish or Portuguese) on attainment of independence. The majority of the people however do not understand the official language as it is acquired through formal education to which the majority of the people are not exposed. The questionnaire is printed in English or French, responses are recorded in the language of the questionnaire. There are errors involved in leaving the work translation in the hands of field staff, not to mention lack of consistency and uniformity in such a way questions are translated into local language by different enumerators (UN, 2005). For example, Tanzania experiences this situation; its report mentions some perplexities (Kiregyera, 1982). Feskens *et al.* (2006) suggest that it is more useful to make questionnaire more understandable to reduce the bias in the response.

McCarty *et al.* (2007) demonstrate that some political issues have a significant role on the informant and they can influence the response rate. Storms & Loosveldt (2004) as cited in Graham *et al.* (2006) explain that different cultural backgrounds present challenges because of cultural relevance which also affect the response rate.

UNIVERSITY of the
WESTERN CAPE

2.3.2 Recommendations on Methodologies to be used

The UN also provides recommendations on methodologies to ascertain the quality of data on age and sex distribution. Many of these errors are apparent, measurement techniques are developed to identify errors and to refine the reports data. Tabulation of data may present errors of coverage, failure to record age and misreporting of age. Many of the measures of errors do not serve to adjust the errors in the data and there is no distinction between the measurement of errors in census data and procedures for adjusting the census data to eliminate or reduce these errors. Sampling biases may arise from inadequate or wrong specified probability sample or faulty methods of estimation of the universe values.

The former includes defects in frames, wrong selection procedures, and partial or incomplete enumeration of selected units. Control of processing errors can be achieved through verification procedures (Groves, 2004). The UN sex-age adjusted birth rate is particularly applicable to some statistically under-developed countries because only data on the total number of births and the population by age and sex which fit the requirement for the calculation. Also, a special standard does not have to be selected, and the rate is simpler to compute than the conventional birth rate adjusted by the indirect method (Siegel, 1980).

UNIVERSITY of the
WESTERN CAPE

Measure of errors in age data are indexes describing the relative level or magnitude of error for the total distribution. The indexes may refer to particular classes of the age distribution, to various ages, for example ages with certain terminal digits. Other procedures provide only estimates (ε) relative error for age groups. An index for a particular age or age group or an alternative estimate of the actual population or of its relative size may then serve as the basis for adjusting the erroneous census count.

2.4 Techniques for Evaluating and Analyzing Data on Age and Sex Composition

These methods relate to the analysis of sex composition, analysis of deficiencies in age data and analysis of age composition.

2.4.1 Analysis of Sex Composition

There are three measures namely the percentage of males in the population, the sex ratio and the ratio of the excess or deficit of males to the total population. These are useful to make comparison to one population group with another, for different areas over time. They set out the effect of variation in population size. United Nations recommendations define these measures in the masculinity proportion or percentage males. The masculinity is the measure of sex composition. This indicator represents $(P_m/P_t)^* 100$; where Pm is the number of males in the population, P_t is the total number of population.

Another important measure of sex composition is the sex ratio of population. Sex ratio is also a measure of sex composition. It is the number of males per females times 100, $(P_m/P_f)^*100$. A sex ratio above 100 shows excess of males and below 100 demonstrates an excess of females in the population. Some countries such as India, Bulgaria calculate the contrary, the sex feminist (percentage of females, female sex ratio…). According to some situations that countries face, their sex ratios can range between 90 to105 (conflicts, massive migrations, epidemics).
Sex ratio = [(Percentage males) / (1 - percentage males)] * 100.

UNIVERSITY of the
WESTERN CAPE

2.4.1.1 Analysis of Sex Ratio

The analysis of sex ratio presents all differences in sex composition from group to group in a given time for a particular group. In terms of subgroup population the sex ratio varies among regions. The sex ratio for age group varies around the sex ratio for total population. At the very young ages the sex ratio tends to be high and then tends to decrease with increasing age. Siegel *et al.* (2004) find that young population and population with high birthrate tend to have higher overall sex ratios than old population and population with low birth rates because of the excess of boys among births and the excess of male deaths at the older ages.

The sex ratio of birth is 100 for nearly all countries for which complete data is available and between 104 and 107 for developing countries where the registration system is not completed (Siegel *et al.*, 2004). The access to identify the gender of a fetus is influencing the sex ratio

especially in those countries with a strong preference for sons (Chaurasia, 2005). Banister (2004); Ebenstein (2007) demonstrate the consequence of sex selective abortion in the increase of the observed sex ratio at birth in Korea and India. Guilmoto, (2009) insists on this pattern of sex selective and highlights potential areas such as Western China, Central, Eastern India, Pakistan, Nepal and Bangladesh that remain so far affected with the sex selection. These areas may also fail to increase the sex selection. Areas with incomplete births registration reflect the suspecting observed sex ratio.

In some developing countries with low level of literacy, a low percentage of the population living in urban area, a low percentage of births occurring in hospitals male births are more likely to be registered than female births (Siegel *et al.*, 2004).

Death tends to depress the sex ratio of most population. These indicators are different from country to country because of two important reasons. First, the difference is matching in the age-sex structure of the population from country to country and second, the differences occur in death rates for each age-sex-group. For the difference between male and female death, Franks *et al.* (2006), Amann *et al.* (2008) prove that biological and cultural factors contribute to the sex differential in mortality. Many women are exposed to the risks of childbearing, the weight of biological forces reflects in the higher mortality of male infants and fetuses compared to females.

Trovato & Heven (2006) explain that since 1970s the sex differential in mortality increase in some developed countries. The effect of war, political instability in most of developing countries affects also the sex ratio. Generally men are directly participating in conflicts (Burundi, Chad, Rwanda in 1994, South Africa during apartheid period,…). Some practices may affect also the sex ratio, the provision of better care to children of one sex than the other, for example in Afghanistan the death rate for girls is higher than for boys.

The HIV/AIDS is a great factor affecting sex ratio of death and age composition. Salomon *et al.* (2005), Bongaarts *et al.* (2008) highlight that more women than men are HIV positive in Sub-Saharan countries; therefore, considering South Africa with the high prevalence rate, the projection is showing that by 2020, the mortality for women will peak during the ages 30 to 34 while for male the projected peak will correspond to ages group of 40 to 44 years.

The sex ratio of an area may affect some specific features of the area that selects some categories of migrants. This is confirmed by the recent immigration reports of many countries showing that more skilled males and females are received (Docquier *et al.*, 2012). Different aspects are observed in Asia where women are dominating migration streams to large cities (Elmhirst & Resurreccion, 2008; Bhagat, 2008; Siridhar & Reddy, 2011). The cultural practice of a woman's moving to her husband's village at marriage contributes to the migration of females from urban to rural areas. This mobility also affects the sex ratio especially during enumeration period (Visaria, 2004). Firebaugh (1979) examines that in Latin America, women are dominating in the internal migration streams to urban areas. The factor of internal migration is fundamental in the different sex ratios of the rural and urban population of the United States. Women are dominating the migration of the rural to urban areas. However, the missing of males or females impacts on the sex ratio of a specific area accordingly.

UNIVERSITY of the
WESTERN CAPE

2.4.1.2 Excess of Males or Deficit of Males

The excess or deficit of males defined as a percentage of the total population is equal to the difference between male and female population $(P_m-P_f) / P_t) * 100$. According to this measure the positive value assume an excess of males and the negative value denotes an excess of females and the balance in sexes is zero. In the absence of basic data on the number of males and females, it is possible to obtain the percentage of males by using the following formula: Percentage males = ((Sex ratio) / (1 + Sex ratio)) * 100. Percentage excess or deficit of males = (Percentage males – (1 –Percentage males)) * 100.

2.4.2 Analysis of Deficiencies in Age Data

The compilation of births may suffer from inaccuracy or deficiencies with respect to the geographic allocation of the birth, the allocation to the year of occurrence, demographic and social characteristics of the births of their parents (e.g. age of the mother, education of mother or father, etc…). Deficiencies in the geographic allocation of births usually take the form of

excessive allocation of births to central cities at the expense of the rural areas and result in the concentration of hospitals and clinics in the central cities and the consequent tendency of mothers to go to the city to have their babies. The additional question on place of birth in 1970 round of African Censuses follow the recommendations of the United Nations Economic Commission for Africa. The UN is mindful of the particular problems of recall that may arise in time specific questions in African censuses. The place of residence is accurately recorded by the enumerator, but the specification of the place of birth presents a series of potential problems for no internal check is possible on the accuracy of this answer. Considering the non- response, birthplace may not be known by some of the population. This problem occurs among nomadic group engaged in regular movements crossing regional, international boundaries for which administrative units are not easily identifiable. Some women leave their place of residence temporarily to deliver for example of their parental homes. Delivery can clearly affect the response to a place of birth question. Some census authorities have to phrase the birth place question in terms of the mother's normal residence at the time of birth.

In some cases, response error may arise as a result of deliberate misspecification of the place of birth by the respondent. De Moor & Van Zanden (2010) investigate the relative numeracy of women during the middle ages and Clark (2008) reviews the evidence. For example, in 1970, the Ghanaian census coordinators gave instructions to enumerators to stress on clarity on the place of birth. The perception of regional units by respondents may affect census area boundaries if they do not coincide and where major changes in administrative boundaries have recently taken place. This bias may occur in those cities of either the less or more industrialized countries where a large percentage of births occur in hospitals. It is offset by the availability of hospital facilities. A comparison of births tabulates by place of occurrence and births tabulate by place of usual residence of the mother for certain cities would illustrate the extent of the problem.

The errors are grouped in different categories: errors in single years of age, errors in grouped data, reporting of extreme old age and failure to report age.

2.4.2.1 Patterns in Single Year of age

The respondents tend to report certain ages at the expense of others. These distortions are age heaping or digit preference. Digit preference refers to preference for the various ages having the same terminal digit and the age heaping is most pronounced among one population or population subgroups. Crayen & Baten (2006) find that United Kingdom registered a substantial age heaping during the early 19th century. In the same manner, Manzel (2007) identifies that age heaping increased in Spain during the famine period at the middle of 19th century. The patterns of age differ from one culture to another. But preference for digit ending 0 and 5 is common. For example, in Korea, China, and some other countries in East Asia, there is sometimes a preference for ages ending in the numeral 3 because it sounds like the word or character for life. In some cultures certain numbers and digits are avoided, for example, 13 is frequently avoided in the West because it is considered unlucky. The numeral 4 is avoided in Korea and in China because it has the same sound as the word or character for death.

UNIVERSITY of the
WESTERN CAPE

The quality of age reporting is measured by means of age heaping indices to detect the degree of preference or avoidance of specific digits in age reporting (Micklin & Poston, 2005). Crayen & Baten (2006) also find that the level of age heaping correlate with illiteracy. Irregularities in reporting age are detected by using graphs and indexes. As far as graphs are concerned the study emphasizes graph of age ratios, graph of sex ratios and graph of population pyramids. Concerning indexes, there are Whipple Index, Myers Index and combined index of United Nations.

2.4.2.1.1 Whipple's Index

Whipple Index is mostly applied to determine the attractiveness or repulsiveness of particular ages using data from census or survey (Hobbs, 2004). This study measures the heaping on terminal digits 0 or 5 in the range 23 to 62 arbitrary (the choice of range is arbitrary). This technical method measures the variability in the quality of age ending by 0 or 5 reporting

between sex distributions. United Nations selects in the Demographic Yearbook of 1955 the Whipple's Index because of its simplicity and wide usage in other sources (UN, 2008). For example, the calculation of the Whipple's index for the 1987 Mali population census reveals better quality data for males than females according to age declaration. According to the scheme for estimating the reliability of the data, the index varies between a minimum of 100, indicating no concentration at all at digits 0 and 5 and maximum of 500; if only digit 0 and 5 are selected. Between these extreme values, the data quality is regarded as highly accurate if the Whipple's index is calculated in a particular manner (Hobbs, 2003) as supported by the United Nations (1990) recommendation that if the values of Whipple's Index are less than 105, then the age distribution data deems to be ''highly accurate''. Also, if the values are between 105 and 109.9, the distribution of age is ''fairly accurate''; if between 110 and 124.9, ''approximate''; if between 125 and 174.9, ''rough''; and if 175 or more, ''very rough''. Generally, the ages of childhood and old ages are excluded because they reflect another type of errors of reporting than of preference when comparing indexes.

UNIVERSITY of the
WESTERN CAPE

2.4.2.1.2 Myers's Blended Index

This approach is to investigate a preference for terminal digits ranging from 0 to 9. Myers checks the incidence of each of the digits 0, 1, 2, 3… 9 separately and comparing with the sum of population in the total age range between 10 and 69. Shryock & Siegel (1976) argue that Myers has a limitation due to different assumptions at the base of the computation. This approach also tests age accuracy for single year distribution. The index shows the preference for or the avoidance of, each of these digits to determine preferences. With advancing terminal digits of age, this is not appropriate because the sum tend to increase. The Myers index solves this problem by combining the population in such a way that each digit has almost an equal sum. The blended totals for each of the ten digits expect to be nearly 10 per cent of the grand total. The extent of the over selection or avoidance of a particular digit shows up in the deviation from 10 per cent of the proportion of the total population reporting on the given digit. The deviation of each sum from 10 per cent of the grand total are added together, ignoring the signs (absolute value). Their sum is the Myers Index.

A summary index of preference for all terminal digits is derived as one half of the sum of deviations from 10 per cent. The index is 0 where all terminal digits are equally chosen and 90 where a single terminal digit is chosen by everyone. The theoretical range of Myers' index is from 0 to 90. An index of 0 represents no heaping and an index of 90 represents a heaping of all reported ages at a single digit. For example, Opiniano (2004) points out that: the age distribution of the Philippines 1990 from ages 0 to 99 shows some form of misreporting and the investigation discovers strong preference for 0 and some for 1,2, and 5 (Siegel & Swanson, 2004).

2.4.2.2 Patterns in Grouped Data

2.4.2.2.1 Overview

Age misreporting and under enumeration continue to affect the 5 or 10 years age data. Absolute net age misreporting error and the percentage of net age misreporting error alternate the positive and negative value over the age scale, dropping to zero for the total population sum of all ages. Therefore the net census error is identical to the net under enumeration. In general the net age misreporting is decreasing and the net age under enumeration is increasing. These errors vary from country to country, from census to census. Ewbank (1981), Siegel & Swanson (2004) identify many studies of age misreporting patterns in developing countries and demonstrate that these patterns for the age groups 0 to 14, 15 to 29 and 30 to 34, the reported ages of children tend to be more accurate than for the adults, even children ages show decreasing accuracy with increasing age of the child. Enumerators introduce bias in the reporting of age for women aged 15 to 29, by estimating age on the basis of union/marital status, or parity of the women. It is difficult to measure the errors in grouped data on age with any precision.

UNIVERSITY of the
WESTERN CAPE

2.4.2.2 2 Combined Index of United Nations (C.I.U.N.)

The CIUN is the best method and it is more relevant. It considers the variations in the quantity of omissions by age group for which declaration is not accurate and preference for ages ending by one of the digits ranging from 0 to 9. This index is developed on the basis of sex ratios and age ratios, computed for five year age groups rather than single year age data, up to age 70. The adjustment of data is done only if the quality of data on sex is assessed. In accessing the accuracy of reporting by sex and five year age groups, the sex ratio score and the United Nations age-sex

accuracy index are used (UN, 1952; Hobbs, 2004). The history of the population under study needs considering the encountering events which influence the structure of age and sex distribution. The United Nations joint scores or the age-sex accuracy index quantifies the accuracy of the overall age-sex data when the data identify digital preferences, but it is also sensitive to the omissions of charge in the vital rates (Arriaga *et al.*, 1994).

The index computes by simply adding the sum of the males' age ratio score and the females' age ratio score to three times the sex ratio score. In defining criteria for data quality, according to the UN (1983) in Gunasekera (2009) stipulates that if the index is under 20, then data are accurate, if the index lies between 20 and 40, then data are inaccurate and if the index is above 40, then data are highly inaccurate. CIUN index presents advantages over other similar indices. Firstly, it is applied to the 5 year age groups rather than single year of age data. Secondly, CIUN index takes into account the differential omission of persons in various age groups in addition to the age misstatements. Therefore, it reflects the general accuracy of age sex data. However, according to (Shryock & Siegel, 1973) in Gunasekera (2009), CIUN has some disadvantages, which are for instance, the failure of the index to consider the expected decline in sex-ratio with increasing age and of discrepancies in age distribution due to epidemics, migrations and wars; the age ratio does not consider the central age group and which has an upward bias; and considerable weight is given to the sex-ratio component in the formula.

2.4.2.3 Reporting for Extreme Old Age and Centenarians

Census age distribution for those aged 85 and over has common reporting problems of exaggeration in older ages. The ignorance of the true age by the members of the household contributes to this distortion of age distribution. Ewbank (1981) as cited Siegel and Swanson, (2004) find that this problem occurs among the reported age group of 95 to 99 and 100 and over. Many considerations should be taken into account: the probability of dying after 100 years is high even in the case where improvements is observed in the health sector, the life expectancy decrease with high rate.

The number of survivors (100 years and over) at a given census date, of the population 90 years old and over the earlier census tend to decrease than the current census count of the population 100 years old and over. For instance, this situation occurs in the 1990 US census where the count of the population of 100 years old and over was 37306 exceeding the number expected on this basis by 8 per cent (Siegel & Swanson, 2004). The number of centenarians is often inexplicably greater for groups with lower overall levels of life expectancy at birth. Siegel and Swanson argue that about 16 per cent of the 37306 reported at age 100 and over in the 1990 US census were black whereas the number of blacks represents only 12 per cent which means 7 per cent of the population 85 years and over.

Using the Vincent's method of extinct generation to evaluate the census count of old age, Gupta (1991) estimates 15236 centenarians in the United States in 1980 compared with an enumerated total of 32194. The use of administrative records (medical care records) is another method to estimate the number of centenarians and to evaluate the report census counts of this group of population. Jdanov *et al.* (2008) and Miller *et al* (2010) using the medical care files identify that the reported census total of the centenarians represent an over enumeration of this group. It is necessary to compute the age rate to develop certain indicators for the whole population. For example, computation for the life expectancy at 40 (age) even in such a case the rate would not be correct. There is a great challenge to measure the number of centenarians knowing that the growing number of this population group entails an increase in the need of public health costs.

UNIVERSITY of the
WESTERN CAPE

2.4.2.4 Age not reported

The problem of age not reported occurs in many censuses. During collection of data on census or survey, the enumerators have the ability to assign an estimate age while enumerating if the age is not reported. Generally, the category of unknown ages is distributed arithmetically prior to the publication. United Nations Demographic yearbooks present census age distributions without category of unknown ages as a result but the paradox is that the method used in census to eliminate the frequencies in this category is always unknown and does not appear on the United Nations tables.

Demographic yearbook (1997) shows that about the half of 75 countries with census age distribution reported have frequencies in a category of age not reported. Siegel and Swanson

(2004) explain that ages have to be assigned to persons, whose age is not reported on the basis of related information on the marital status for the person and other members of the family (partner, spouse), for ages based on data from questionnaire, using education's information, employment status. For instance Siegel and Swanson (2004) show that in the 1990 census of the US, age is allocated 2.4 per cent of the enumerated population on the basis of other information regarding the same person, other person in the household, or person with similar characteristics reported in the census questionnaire. Age substitutes for an additional 0.6 per cent of the population; hence, the total of 3.0 per cent of the 1990 census population has a computer-generated age.

2.4.3 Analysis in Age Composition

Age is one of the most important characteristics of population composition, hence all kinds population vary significantly with age. The usefulness of age data is more visible especially when it cross-classifies by variables such as gender, province, population groups, education, economy activity which vary with age in different patterns.

UNIVERSITY of the
WESTERN CAPE

2.4.3.1 Nature of Age Distribution

Some analysis requires data of single year's age because change can affect rapidly its characteristics as labour force status, marital status and so on. But generally most of the data are tabulated and published in 5-years groups. For other purposes age data can be a combination of 5-years groups and 10-years groups. It is useful to have the number of people of 18 and over, 21 and over for example.

2.4.3.2 Percentage Distribution

The absolute numbers distributed by 5-years age groups are converted to percentages. To compare the percentage distribution for different countries, we have to consider the size and other characteristics.

2.4.3.3 Variation by Age

The measurement of change over time is necessary in the analysis of age data. The method of analysis of age focuses not only in the comparison of different populations but also to the comparison of the same population at different date.

2.4.3.4 Measure of Old Ages

The median age divides the population into two groups: the young population and the aging population. When the median is increasing, the population is aging and the decrease is observed in the young population. The proportion of aged persons is considered as an indicator of young or aging population. The proportion of people under age 15 is also an indicator of the degree to which a population is young or old. The ratio of the number of elderly persons to the number of children, or the age child ratios consider the number and change at both ends of the age distribution. Developing countries have small proportion of people aged 65 and above, with very large proportion of children under 15. However, in some cases population can youngling and aging at the same time which means that the decline observed in the intermediate ages can compensate the increase occurred in the proportion 65 and above and the increase in the proportion under 15.

2.4.3.6 Age Dependency Ratio

The age dependency ratio is a measure of age composition. Its total represents the ratio combined child population and aged population to the population of intermediate age (15 to 64). Variations in the age dependency ratio reflect in a general way the contribution of variations in age composition to variations in economic dependency.

2.4.3.7 Age-Sex Structure or Population Pyramid

Population pyramid is a diagram illustrating the age distribution of a population. It can be used to find the number of economic dependent being supported in a particular population. Demographic

transitions highlight some issues of population structure. However, the decline in fertility provokes substantial changes in the future population structure. Another measure for age-sex structure can clarify issues that arise with this process. There is a need to incorporate age-sex structure policies with population policies.

2.5 Inspection of some Examples where the Evaluation of Data on Age and Sex Distribution are Engaged and Progress Made in the Methodologies

2.5.1 Cases Studies

2.5.1.1 Example 1: The reliability of Lesotho's census data (1986 and 1996 censuses)

Mba, (2003) assesses the reliability of 1986 and 1996 censuses of Lesotho. The main purpose is to assess the quality of age and sex distribution reported. He uses Whipple's index Myers index and Bachi's index. He finds that the preference of even numbers occurred in the census of Lesotho. The Whipple's index of 1986 is 115; it decreases to 106 in 1996. Myers index also declines from 11 in 1986 to 9 in 1996 and, Bachi's index in the same vein decreases from 7 to 6. However, there is an improvement in the quality of its data. This study presents some inconsistencies in the reported age and sex distribution of Lesotho. In regard to single year age distribution, there is evidence of distortions in reporting age. The patterns of digit preference occur in both censuses of Lesotho. In 1986 census, digit 0 is most preferred by males and females follow by 8, 6 and 2. In 1996 the digit 6 is most preferred by males and females preferred the digit 0 more than anyone. Both sexes prefer digit 0, follow by digit 6. The magnitude of preference and avoidance in Lesotho is very small when compared to African standards. The sex ratios are fluctuating in both the 1986 and 1996 censuses. The age group 35-39 with sex ratio of 102 in 1986 and the sex ratio of 105 at the age group 45-49 presents inconsistencies in the age data. In the same vein, having 103 of sex ratio at the age group of 45-49 in 1996 while preceding age group has lower sex ratio demonstrates deficiencies in the data. However, the two censuses indicate that the number of females is greater than that of males. Mba (2003) assumes that the low sex ratio over the decade may result in under reporting of females, high emigration of males, low sex ratio at birth and high male mortality at subsequent ages. In some case after working in South Africa, males migrate back into Lesotho, this for retirement

UNIVERSITY of the WESTERN CAPE

purpose. The derived age ratios and the age accuracy affect probably at the age group 45 and over. The distortion observed in 1986 census data indicates that there is over enumeration at the age groups 5-9, 10-14, 50-54, and 65-69 for both sexes while other age groups register under enumeration.

In 1996, the highest over enumeration is observed at age groups 60-64 for both sexes, while the highest under enumeration occurs at age group 65-69 for both sexes. These findings indicate that old ages are more likely to state their ages incorrectly than young ages. This issue links to poor level of education of old ages. Considering the Age-sex accuracy index, a slight improvement occurs during the study period. In 1986, the index is 35 and in 1996 it is 32. Despite the fact that the data quality is not accurate, these results are significantly lower than what obtains in many Sub-Saharan countries. The study reveals that the reported age sex distribution of Lesotho's population suffers from inconsistencies but the exact degree of errors cannot be determined because of complexities in the patterns of these errors. Civil registration systems are still developing in Lesotho, it is impossible to determine the correct age distribution in Lesotho without appropriate birth certificates even when migration is considered.

UNIVERSITY of the
WESTERN CAPE

2.5.1.2. Example 2: Estimating Adult Mortality in Cameroon from Census Data on Household Deaths: 1976-1987

Bangha (2008) assesses the age and sex distribution data of child and adult mortality using 1976 and 1987 censuses data of Cameroon. The percentage distribution of single year for both censuses presents a general decline as age advances with some patterns of strong preference for digits 0 and 5 in the reporting of ages. In regard to 1976 census, the patterns of preference decrease in 1987. The heaping declines with the processing of grouping age. When comparing 5-years and 10-years in parallel lines, there is no crossover to find inter-censal migration movement or a large under enumeration in one census relative to other. The age specific sex ratios are consistent indicating more male than female births.

The sex ratio estimates at 96 per 100 for 1976 and 1987 with a constant fraction of males of the total population. The sex ratios by age groups vary from 104 for the youngest age group from 0-

14, to below 100 for age groups 15 and above. The patterns of age specific sex ratios are consistent with the whole picture for Africa (UN, 2004) where the sex ratios are around 100 for the age group 0-14, decrease to 90 for the broad age group 15-59, and then drop below 80. The overall sex ratio of household deaths reported for the 12 months preceding the 1987 census is 117 males per 100 females which is close to 116 reported for the 12 months prior to the 1976 census. The observed age specific patterns reveal an excess male mortality.

The sex ratios of reported deaths by age group start at 120 and remain above 100 across age except for ages 15-44 where the male mortality drop consistently and reappear at the end of the reproductive span (45-49). The pattern is the same for the two censuses, except the one observed between ages 15-49 which is deeper for the first census than for the second. This illustrates high maternal mortality at the first census which declines by the 1987 census. The young adults and old adults are generally missing during a census enumeration. In this study, the migration age selectivity is not referring for males since this pattern is observable for both sexes; the missing can be a potential shifting of individuals as a result of misstatement.

The 1976 census is more complete compared to the 1987. The statistics show that the 1987 is about 93 per cent complete relative to the 1976 census, with the male enumeration estimates at 95 per cent complete and females at 90 per cent complete. The under enumeration is largely experiencing for the second than the first census. The population is experiencing decline adult mortality between 1970 and 1980. The results decrease by sex showing that the adult's female mortality is improving over the inter-censal period. The influence of various factors (socio-political, economic and environmental) occurred in Cameroon during 1980, has disturbing and producing the negative effect on the health of adult males. These difficulties may have inducing a differential male-female mortality response. However, the third enumeration round in 2005 would help to evaluate this study once its data will be available.

2.5.1.3. Example 3: Investigation Pattern of Preference in the Census Data (1994-2009),

2.5.1.3.1 Philippines 1990

Rodriguez (2009) assessed the reliability of the age distribution of the Philippines in 1990 from ages 0 to 99 which appears in Siegel & Swanson (2004), p. 103. The distortions seem to

represent some form of age misreporting. The researcher investigates pattern of preference and obtains strong preference for 0, and some for 1, 2 and 5 while strong avoidance appears for 8, 9 and 6. Philippines, as many developing countries, is facing problem of registration of births and is also observing high rate of mortality. Considering the findings, the expected people should be more at age 0 than 5 or 9, and more at 10 than 15 or 19 and so forth. Using Myers method, Rodriguez (2009) found that with data cover the range 10-89, Myers's blended index was 1.93. According to this result there is a need to reclassify 2 per cent of the cases to have uniform distribution. Rodriguez (2009) continues to sum all the range 10 to 99 (excluding 100 and above, which is left out of the dataset which means that the assumption of having all frequencies above 99 are all 0); applying that assumption, the Myers's blended index becomes 2.33. The preference is most pronounced at age 0, 5 and 9 while avoidance is more observed at age 3, 6 and 1. The problem of education could be another possible reason affecting this population of Philippines to not report properly their ages.

2.5.1.3.2 Breastfeeding Duration: Bangladesh 1975-1976

UNIVERSITY of the
WESTERN CAPE

Rodriguez (2009) studied the duration of breastfeeding in the last closed interval between the next to last and last child from Bangladesh. He found that the duration of breastfeeding was related with the age of mother. He also identified that preference for multiples of a year and multiple of 6 months are closed and common to wishes of women. Generally women tend to wean the child at age one, two or three. Using the month's option, it could extent to 60 months, an exact multiple of 12, and put down 11 more months of data for blending. Rodriguez (2009) found the Myers index with a range of 0 to 59, he observed 63 per cent of the cases report an exact multiple of 12 months (equal to a year) and 25 per cent report a multiple of 6 months (but not 12). Myers blended index was equal to 70.58 indicating the pattern of preference was more pronounced in this population. The quality of distribution was totally inaccurate. There was a great need to reclassify 71 per cent of the cases to obtain a uniform distribution by month. The value of Myer's index (70.58) was considered very high compared to the full range of this index which varies from 0 to 59. This index confirms that the reporting of the duration of breastfeeding in the population of Bangladesh is significant. A possible reason explaining why respondents often report wrong their duration of breastfeeding with multiple of 12 might be that the children born in rural areas were breastfed for shorter periods than those born in urban areas of

Bangladesh. Also, the duration of breastfeeding decreases systematically according to the level of education of the mother despite the fact that she tends to declare inaccurate responses.

2.5.2 Progress Made in the Methodologies

Many aspects of individual life, public policy, ranging from schooling, health care, and social security to labour force participation and productivity relate to age and gender showing the matter of great interest. The collection of data on age and sex distribution from censuses or surveys presents serious difficulties which affect the quality of these data in many developing countries. Age is an important demographic characteristic with an impact on the dynamic of the whole aspect of the size of the population under study. The completeness and the accuracy of reporting the age or the date of birth are very critical issues. The additional question on date of birth, the birth history, and the important events occurred in the past and the characteristics of other members in the household contribute to achieve the approximate age of those respondents unable to provide their exact age. The evaluation of these data require a consensus about methods, content of censuses or surveys, civil registration systems, sampling procedures and training procedures to produce valuable data. Literacy of population is showing a positive expectation.

UNIVERSITY of the
WESTERN CAPE

Nowadays, South African government is paying attention to the vulgarization of vital registration systems and they are also providing good awareness to its populations. There is a slight improvement, but in many of Sub-Saharan countries, the lack of progress towards a complete civil registration system remains critical. According to the age heaping in Africa, there is no change occurring. The mobility of population is an obstacle to the improvement of these data. For instance, the emigration of well-educated Whites, productive and skilled workers highlights the sluggish economic growth in South Africa. There is a need of influx of skilled workers to fill this gap. Huge efforts also are needed in translation of instruments into local languages implying extensive gains in the accuracy of responses and possibly reducing` the incidence of interviewer error. Another issue which arises in this study is that after age 40, the rate of male mortality is significantly higher than for female which reveals female lives longer in retirement years. Their less participation in Labour Force and the quality of their work cannot help them to afford the family needs.

In many African countries, the availability of data on age and sex distribution is published with delays in the UN Demographic Yearbook. For example, the data on family and household of Morocco resulted in long delays 1975-1992 (New York, the population council, 1992) as cited (Cleland, 1996). Greater funds are available for data collection but during the analysis process the agents involve in this operation always facing problem of funding. These days, Statistics SA are given more attention in the training for enumerators and supervisors in such a way to improve their effectiveness on the field during October 2011 (census).

UN recommendations are aware of the particular problems of recall that might arise in time specific questions in African censuses. The variations in age and sex distribution occur as result of differential mis-enumeration. The problem of age misreporting is particularly serious in Africa and may be compounded by different patterns of misreporting for males and females. Age and sex help to evaluate the importance one should assign to explicit consideration of such other dimensions as labour force, region, educational status, health facilities…

Disparities in data on age and sex persist. The continued assessment of age is essential. No country follows international recommendations to the letter; many national statistical agencies respect only the basic concepts. As a result, the scientific quality of enumerations improves spatial comparability increases (UN, 1998). The population census becomes a global phenomenon, the making of data samples; harmonization of census is an obvious way of enhancing the use of these data. Each year brings more relevant data by which these analyses can be made more accurate. Post-censal field checks are increasing. It is now possible to evaluate the data much better and to use them more effectively despite their deficiencies. However, two important challenges remain to deal with, first, the problem of unbalance sex-ratio in developing countries with its implications when the excess of females are adults and second, the growing number of extreme-age with its implications in the future policies of these countries.

In conclusion, despite the systematic checking in all steps (procedures) during the census, the researcher has to examine all difficulties encountered during the fieldwork. In this case, a post enumeration survey is recommended to assess the quality of data collected. The extent to which they provide an adequate indication of patterns of population census gives the possibilities for limitation of their use in population analysis. The next chapter deals with methods applied in the research.

CHAPTER THREE

METHODOLOGY

3.1 Introduction

This chapter discusses the key methods applied in this study, the ways in which information utilized were acquired and methods of analysis. In particular, it emphasizes the techniques used for data collection; sampling design, variables and procedures. The instruments employed were Census 2001, General Household Surveys (2004 and 2007), Labour Force Surveys (2005 and 2007) and Community Survey (2007). The chapter highlights how data were analyzed, using indirect methods such as Whipple's index, Myer's index, CIUN, age ratios, sex ratios and population pyramids to evaluate the quality of declaration on age and sex. The procedure involves the comparison of statistics for each instrument, by type at national, provincial or per population groups and also for the same instrument over years for different instruments over years. The chapter ends with the discussion on the limitation of this study.

3.2 Type and Perspective of the Study

UNIVERSITY of the
WESTERN CAPE

The design of this research on assessment of the quality of data on age and sex follows a quantitative perspective and more specifically the study makes use of a normative comparison design. This type of research is important to utilize since it facilitates the purpose of this study, that is, to examine some structural changes in the age and sex structure of the population in South Africa between 2001 and 2007. The main objective of this design is not only referring to explaining and detecting fluctuations observed in the statistics but to contribute innovating new facts or techniques to adjust in the future. The study is based on demographic characteristics such as age, gender, ethnic group and province of residence.

The measurement of variables in this study is also defined and the statistical methods used to compare the observed findings and the expected findings. Thus the significance of the study is descriptive analysis which highlights different variations observed. By bringing together the demographic variables and other factors related, the study captures the structural changes between 2001 and 2007 and also allows the researcher to see the difference in terms of the

quality of response on age. The following section describes the methods used for data collection, sample design and procedures of each instrument applied in this research.

3.3 Source of Data Evaluated

Full data sets (Census 2001, GHS 2004 and 2007, LFS 2005 and 2007, Community survey 2007) were acquired by requesting them from Statistics South Africa. This study was conducted in all the nine provinces of South Africa specifically by looking at some structural changes in the age and sex structure of population between 2001 and 2007 which is the period of interest. These data were accessed according to the methodology and pro-coding underlined by Statistics South Africa. The files were obtained in SPSS format and this made it possible to (cross-tabulate) run statistical methods on the data files which refer to sections of the questionnaire. The data files considered in this study were that of personal file.

3.3.1 Census 2001

Censuses are the most logistical operations to be carried out by any country. The Census 2001 was conducted on a de facto basis using the same reference point 9-10 October. Each individual was counted in respect of his usual place of residence. The approach of enumeration was face to face and self-completion was applied.

However, the count or figures as indicated in the 2001 census was wrong due to the under count of blacks for reason of accessibility of black areas assumed to be dangerous and misinterpretation arising from the translation of the questionnaire into the 11 official languages. In cases of missed count, the PES presented an opportunity for correction or they simply resort to statistical imputation to correct the census figures. In census, EA boundaries were reevaluated in 2001 and PES was organized after the census. The results from PES were compared with those from the census and where there was any gap or uncertainty, the solution was to revisit the particular household. Then according to the census the required adjustments to the figures were done.

3.3.2 General Household Surveys (2004-2007)

In 2004, more than 30 000 households across the country were sampled for this purpose. The census frame used to draw the Primary sampling units (PSUs) for the master sample removed all EAs with less than 25 households. A multi-stage stratified sample was applied, using probability proportional to size principles to make it more representative. The master sample was drawn from the enumeration areas (EAs) established during the demarcation phase of Census 1996. Small EAs consisting of fewer than 100 households were pooled together to form primary sampling units (PSUs) of at least 100 households in such a way it will be possible to repeat sampling of dwelling units within each PSU.

The method of sampling for the master sample comprised stratification by province and within each province, by urban and non-urban areas. The measure of size was the number of households in the PSU. 3 000 PSUs were selected. Only ten dwelling units were drawn systematic in each selected PSU, thus, resulting in approximately 30 000 dwelling units. All households in the sampled dwelling units were enumerated. The PSUs were allocated to the district councils (strata). The probability proportional to size principles was applied to sample PSUs. The number of households in a PSU as planned in the census was the measured of size. The systematic sampling of dwelling units were drawn from sampled PSUs which formed clusters. The master sample was divided into five independent clusters.

UNIVERSITY of the
WESTERN CAPE

However, small modifications appeared in the 2007 GHS. This latest applied a multi-stage design based on a stratified design with probability proportional to size selection of primary sampling units (PSUs) at the first level and sampling of dwelling units (DUs) with systematic sampling at the second stage. After assigning the sample to the provinces, the sample was then stratified by geography and non-metropolitan geographic area type, and by individual attributes using the Census 2001 data. The adjustment of weighting and benchmarking were done according to the new provincial boundaries that came into effect in December 2006. However, the data for the GHS 2004 to GHS 2007 are therefore comparable.

3.3.3 Labour Force Surveys (2005-2007)

The International Monetary Fund (IMF) decided to revise all aspects of the LFS. All reports, methodologies and techniques relating to the LFS were reviewed, before a completed report on the labour was presented to Stats SA in June 2005. The Quarterly Labour Force Survey (QLFS) was the redesigned Labour Force Survey. The QLFS frame can be applied by all other household surveys regardless of sample size requirement of the survey.

The sampling methodology of LFS was consistent in each round of the survey. The sample for the redesigned Labour Force Survey was based on a stratified two-stage design with probability proportional to size (PPS) sampling of primary sampling units (PSUs) in the first stage, and sampling of dwelling units (DUs) with systematic sampling in the second stage.

The sample for the QLFS, which was based on information collected during the 2001 Census, was designed to be representative at the provincial level and within provinces at the metro/non-metro level. The sample was also distributed by geography type, and there were four geography types, namely urban formal, urban informal, farms and tribal areas (Stats, 2008a).

UNIVERSITY of the
WESTERN CAPE

The PSUs were 3080 divided into 4 rotational groups. Each sampled dwelling will remain in the sample for four consecutive quarters (round of survey). The sample size was roughly 30 000 dwellings per quarter. The sample was drawn from the 2001 population census conducted by Stats SA. However, the country was divided into 80787 enumeration areas. A master sample of PSUs includes EAs which were drawn across the whole country. Each of them was representative and had the same distribution pattern as which was observed in the whole sample. They were attributed numbers from one to four in terms of the quarter of the year to be rotated within the sample group (Stats, 2008a).

According to the Master Sample, the EAs that contained fewer than 25 households were excluded from the sampling frame, and those that contained between 25 and 99 households were combined with other EAs of the same geographic type to form Primary Sampling Units (PSUs). The number of EAs per PSU was classified into four groups. The very large EAs represented two or more PSUs.

Every quarter represents a quarter of the sampled dwellings. The rotation of each quarter was made out of the sample and its dwellings were substituted to new dwellings from the same PSU or the next PSU on the list. The dwelling was the sampling unit and the unit of observation was the household. If a household moves out of a dwelling after being in the sample may be for two quarters, and a new household moves in, the new household will be counted for the next two quarters, but if no moves into the sampled dwelling, the dwelling is considered as unoccupied.

The non-respondent households were assigned to adjustment. The imputation was applied for item non-response (refusal, blank questionnaire, no contact). In fact, the edit failure was applied for invalid or inconsistent responses.

3.3.4 Community Survey (2007)

The Community Survey was designed to be conducted between censuses due to the long gap of 10 years across all provinces by Stats SA in 2007. The sample design considered the sampling frame and the challenges that were experienced during the census 2001. The sampling approach consisted of two stages, the selection of enumeration areas and the selection of dwelling units. Each municipality was considered as a unique stratum. The sample was stratified by municipalities as demarcated at the time of census 2001.

UNIVERSITY of the
WESTERN CAPE

At the first stage, the EAs for each municipality were selected by using systematic random sampling. The following criteria were applied; in municipalities with less than 30 EAs, all EAs were selected but, in municipalities with 30 or more EAs, the sample selection used a fixed proportion of 19 per cent of all sampled EAs. At the second stage, the frame required a full relisting of dwelling units, then the selection of dwelling units to be visited was done only for those structures that were classified as dwelling units whether vacant or occupied. The selection of the dwelling was based on a fixed proportion of 10 per cent of the total listed dwellings in an EA.

However, in regard to the size of EA, where the listed dwelling units were less than 10 dwelling units, all households within the selected dwelling units were covered.

3.4 Variables to Analyze

The main variables which were explored for each instrument are age, sex and ethnic group. In line, it has been given details per instrument, on how age and sex were being collected. Furthermore, it indicated how data on sex and age were analyzed per instrument and how the researcher intents to examine the quality of declaration of age and sex.

Instruments: These are the techniques used for the assessment of the variables stated above and they are Census, LFS, GHS and Community Survey.

3.4.1 Census Variables

The study selected variables according to variables used in Census. There were based on socio-demographic characteristics as follow: age, gender, population group.

3.4.1.1 Age

Age was determined by the question: "What is (the person's) date of birth?" the instruction given to enumerators was that if date of birth is not known; provide as much information as possible. The intention was to obtain the exact age of the respondents. The person's age must be given in years or give an estimate. For babies under one year, write "0" years.

UNIVERSITY of the
WESTERN CAPE

3.4.1.2 Gender

The question asked in both censuses to derive the gender of respondents is: "Is (the person) male or female?" and two options were available namely 1 = Male, 2 = Female. The instruction was not to assume the sex of the respondent just by using the name or referring to his physical build.

3.4.1.3 Ethnic Group

Ethnic group is one of meaningful variable which help to obtain a more nuanced picture of changes occurred in the population. This variable can be attached to the national picture. This allows for the development of a complex approach to monitoring and measuring a self-defined racial identity. This study is looking at all population groups 1= African/Black; 2 = Coloureds; 3

= Indian/Asian; 4 = Whites; 5 = other, specify any structural changes amongst them in quality of declaration on age.

3.4.2 Labour Force Survey Variables

The variables of Labour force are defined along the lines of the level of measurement used by Statistics South Africa. The variables analyzed were age, sex and ethnic groups.

3.4.2.1 Age

The question used to define the age of participants is: "what is (the person's) date of birth and age in completed years?" This question was asked for each individual in the household. The second question is: "What is (the person's) age in completed years?" If the date of birth is unknown, just provide the approximated age in completed years and the response must be written in whole numbers, not in words. For children aged less than a year, the respondent or the enumerator would write 000. If age is still unknown, the enumerator has to refer to historical events in terms to help respondent to estimate. If there is no age found, the non-response will be solved at the processing stage when applying the imputation technique. Then, the age is transferred and recorded into groups as: 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64.

3.4.2.2 Sex

The question used to determine the gender is: "Is (person) a male or female?" If the person was absent at the time of interview, the enumerator had to ask other family members whether the person is male or female and not to decide on the person's gender based on name. The variable was coded as follows: (1) Male, and (2) Female. The instruction gave to enumerators was not to assume the sex of respondent just by referring to his name or his physical build. Enumerators had to ask the respondent the gender of each member of the household who had stayed in the households in selected dwelling units at least four nights a week in the four weeks prior to the interview.

3.4.2.3 Population Group

The question asked regarding all population groups was "What population group does … belong to?" The responses were 1 = African/Black; 2 = Coloured; 3 = Indian/Asian; 4 = White and 5 = Other, specify any structural changes amongst them in quality of declaration on age. The motive for asking this question was to find out the differences on trends and link to the LFS along the racial groups for example, White and Black, and also Coloured and Indian groups. Racial differentials in patterns of participation had been highlighted. The respondent had to answer for each member of the household and the enumerator was instructed not to make any assumptions or not to come to any conclusions which may be influenced by his observation or using people's names during the interview.

3.4.3 General Household Survey Variables

3.4.3.1 Age

UNIVERSITY of the
WESTERN CAPE

According to this variable, the question asked to find out the age of the household members is: "How old is ...?" (In completed years - in whole numbers). For babies aged less than 1 year (00), the instruction was to write the years in whole numbers and not in words. These years must be completed, for instance if a person is two years and six months, the instruction was to write completed years in whole numbers and not in words. In 2007, the question was refined and became "what is (…'s) date of birth and age in completed years?" These combined questions were asked for each individual in the household. The respondents supposed to provide answers as following day of birth, month of birth, year of birth and age in completed years. The age was captured and recorded into groups using SPSS as follows: 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75-79, 80 years and above.

3.4.3.2 Sex

The question used to record the gender of the member is: "Is… a male or female?" The gender or sex of the members was recorded as male, female or unspecified. This question remains the same in 2007. The enumerators were instructed not to assume the gender of the members of the

households by just looking at people's names or physical appearances. The enumerator had to ask the respondent the gender of each member of the household without any assumptions.

3.4.3.3 Ethnic group

The question asked to determine the population group of the members from the selected dwelling is: "What population group does (the person) belong to?" The respondent had to answer for each member without any assumption. However, the enumerator was not supposed to make any conclusion influenced, for instance, by using people's names during the interview. This question seems ambiguous but very important in sense that there is a need to determine the composition of South African population. The ethnic groups were coded in four groups following African/Black (1), Coloured (2), Indian/Asian (3), White (4). This study is looking at all population groups in terms to specify and capture changes amongst them in the declaration on age.

3.4.4 Community Survey 2007 Variables

3.4.4.1 Age

UNIVERSITY of the
WESTERN CAPE

Two separate questions to define the age of member of the household in the 2007 Community Survey were set out. The first one was based on the date of birth "What is (the person)'s date of birth?" The answer was recorded as DD/MM/YYYY, DD was for day; MM was for month and YYYY was for year. The instruction was to leave the box blank if the information was not known. The second question was: "What is (the person)'s age in completed years?" If the age of respondent was not known, the enumerators were instructed to ask for an estimate of age and if no one was able to estimate, they should write 998. They were to write 000 for age of babies less than 1 year or stated the exact age. For instance, a person 7 years and 10 months writes 007 for age.

3.4.4.2 Sex

The question to determine the sex of the member is: "Is (the person) male or female?" No assumptions were supposed to be considered. Two options are applied 1 = Male and 2 = Female.

3.4.4.3 Ethnic group

3.5 The geographical level of evaluation

The evaluation concerned the whole country and particular place of residences to ensure that all of the key concepts and processes are covered in such a way the assessment would set out the changes observed in any each factor.

3.5.1 National level

The interest of this study is to evaluate the national trends for each instrument per year, per population groups in such a way it would be easy to compare the trends for different instruments per year, per population groups and see the progress achieved over years.

3.5.2 Provincial level

Considering location variable, place of residence was the main focus. The question to derive respondents' usual place of residence is: "Where does the person usually live?" This was recorded according to the nine South African provinces following Western Cape (1), Eastern Cape (2), Northern Cape (3), Free State (4), KwaZulu-Natal (5), North West (6), Gauteng (7), Mpumalanga (8) and Limpopo (9). The purpose for asking this question was to evaluate the trends according to the distribution of residence allowed comparing distribution within neighborhoods. The geographic pattern was assessed. There were some inconsistencies between place of residence which the main cause could be the respondent error, or the enumerator error or some unforeseen events. These are the South African provinces as at December 2005 released by the Municipal Demarcation Board in January 2006.

3.6 Methods of Assessing the Quality of Data

Although there are various methods that have been developed to assess the quality of demographic data, those methods are either direct or indirect. The indirect methods are useful in way that they assist the researcher to determine or examine something that was done with the direct methods. For instance, for testing the quality of data, most of the researchers use enhances

the calculation of Whipple's Index, Myer's Index, and Combined Index of UN. However, this study also is willing to apply these methods with objective to assess the quality of data on age and sex from censuses and surveys in South Africa from 2001 to 2007. Apart from the above mentioned methods, it has also been applied for the calculation of sex and age ratios techniques. However, to provide an accurate determination of coherences in the methodology for different instruments, it has used the time-space comparison per year for each type of data and the cross-sectional comparison or comparison between type of data and per year.

3.6.1 Comparisons

3.6.1.1 Time-Space (T) Comparisons or Variation for Each type of Data

This technique was employed in the analysis of each instrument mentioned above. The operation of comparison consists of a simple check to see what the progress has been over years. However, comparison of data was based on the comparison of the results obtained in following indexes, Whipple's, Myer's and Combined index of United Nations. It also focused on the Age ratios, Sex ratios and Population pyramids. This technique indicated that there is an improvement observed which mean decline of index value or a deterioration which implies an increase in the value of index. The change in the age structure during two surveys was characterized by a gradual decline in relative size for instance of the young population, or the working age or older population age. The comparison indicated differences in indexes from GHS, broken down by provinces, by population groups. It also highlighted the challenges faced by the country with reducing sex ratio for instance, working population group, and so on. Somewhat, the future demographic events are influenced, to a large extent, by the present sex age structure, considering other factors (components) being constants.

UNIVERSITY of the
WESTERN CAPE

Apart from differential under enumeration in various ages, the comparison helps to set out if the age data suffers from distortion owing to preferences for certain ages and digits due to social, cultural and legal habits as well as norms observed in a society. The change described the variation between the current index and the previous index at a point in time. If the value of the variation is positive it reveals deterioration indicating an increase of the index over time while if the value of variation is negative, it indicates improvement has been achieved showing a decrease of the index over time. For instance, it is well-known that Whipple's index is designed

for single age data to capture heaping on ages for two digit endings 0 and 5; comparing age heaping among provinces could offer a different interpretation. Indeed, Myer's index as CIUN are planned to detect the preference or avoidance in all digit endings but the variation in age ratios or sex ratios indicates the quality in the age reporting of gender, reveals also the level of fertility or mortality of the population.

3.6.1.2 Cross-Sectional Comparisons (LFS vs GHS, Census vs Community Survey)

Cross-sectional analysis involves the observation of all the population, or a representative subset, at one specific point in time, or without regard to differences in time. This method consisted of comparing the differences between the indexes of two instruments for the same type and the variations that occurred amongst different type of instruments. The comparison was done using the three types of indexes Whipple, Myers and Combined Index of United Nations. This operation also included age ratio, sex ratios and population pyramids as well. The interpretation considered the slight change which can occur during a year or more in regard to the baseline. The difference between the indexes of GHS, LFS, Census and Community Survey has shown the level of improvements realized for every category of instrument. However, if the difference is negative and close to zero, the quality of age reporting is relatively poor for that particular category of instrument. The same application was done between different instruments. For instance, the Whipple's index resulting from Labour force survey was compared to the Whipple's index obtained from General Household survey.

UNIVERSITY of the WESTERN CAPE

Moreover, cross-sectional approaches may be perplexed by secular changes such as the progressive increase of index which reveal the deterioration in the quality of index over the time. This method usually estimates different instruments. There are possible explanations including age, period and cohort effects. The variation for instance between the Whipple's index of LFS 2005 and GHS 2004 is equal to: $D = W_{LFS} - W_{GHS}$; Where D is the difference between the two indexes, W_{LFS} is the Whipple's index derived from Labour Force Survey, W_{GHS} is the Whipple's index derived from General Household Survey. In the same vein, the researcher had estimated the variation between LFS 2005 and GHS 2004, LFS 2007 and GHS 2007, Census 2001 and Community Survey 2007 applying all the indirect methods which are explained in the following section 3.6.2.

3.6.2 Methods applied

3.6.2.1 Direct Method

According to South African census or survey, age declaration result in ambiguity for demographic data users. Due to problems in data age capturing, it is standard procedure that every individual should provide his date of birth on reference night and for those who experience problems in remembering or producing the correct information regarding poor declaration of age impact other variables were applied in regard to the limitations.

Special circumstances applied to individuals who had spent the period of reference in the household. For instance, in the context of the LFS, every individual (male or female) who has stayed in the household for at least four nights on average per week during the last four weeks must be enumerated. The variety of checks in the study was made in such a way to ensure that the declaration of age of respondents is consistent. However, the direct method considers the observance of law and practices relating to limitation of age for instance, at school entry, labour force participation, marital status and household headship to check the quality of age declared. The variety of checks in the study had been made in such a way to ensure that the declaration of age of respondents is consistent.

UNIVERSITY of the
WESTERN CAPE

The legal minimum entry age at grade one in South Africa is seven years old. This entry age was treated as an exogenous variable. This concept can change in terms of demographic area (urban or rural). The survey has found that the age in rural areas is sometimes higher which implies that the age of completion may be several years above the nominal age for the last grade (UNESCO, 2005). As a result, the completion of secondary schooling may not occur until after the age of 15 years. These problems sometimes could affect the assessment with precision of the entry age to obtain the correct age of the respondents in the case of misreporting, wrong statement or unable to provide the correct information (UNESCO, 2003 & Lewis, 2005).

In many African countries, accepted marital status of women was in their twenties. However, in South Africa, the minimum legal age of marriage was 21 years (marriage Act 1961) but has since been changed to 18 years (Recognition of customary marriage Act, 12 of 1998) as has mentioned (Andrews, 2007). Yet, parental consent is required for the marriage of a party under the

minimum legal age. Act 1998 deals with customary marriage and has to be registered within a given period of (03 months). For instance, if the data presents a case of 12 years old married, it means there is a problem or the information is not consistent which could affect the quality of declaration of age. This problem can arise for misinterpretation of concept especially when the couple is living together as partners, the responses provided to the enumerators are sometimes not correct.

Labour Force refers to all individuals aged 15-65 who are available, willing and desiring to do economic work. For census purpose, the minimum age selected was 10 or 12 years in the majority of countries. However in South Africa the legal minimum working age was 15. A control check has to be operated in regard to the age of participants of Labour Force if the norms prescriptions at the national level are respected by the respondents. For example when controlling the data, it was found out that someone aged 10 was reported as worker or employed; great attention should be focused on that particular case because of the irregularity which occurs. The content check of the working age contributes to evaluate directly the quality of age declared.

In the context of South Africa the reference person may be the oldest person in the household. Generally, the household members tend to name the oldest man as its head even he is not economical active or even he is not in regular resident. The only difference is that the authority is variably distributed in male headed households but in female headed, the authority is dominant in one person. However, women assumed better the roles of head than men. In South Africa women assume headship in men's absence. Also in South Africa, the living arrangements favour children to become the household head in presence of an adult who he is incapable to fulfilling this role in case of illness for example. It is better to mention that the conceptual definition of the child-headed household is not adequate. The UN conventions on the rights of the child (1989) declare that the legal age to apply to the child headed household is 18 years. The check of the age head of household supposed to be consistent which increase the quality of age declared.

3.6.2.2 Indirect methods (Statistical methods)

Some technical procedures were applied to determine some indexes which allow detecting attractiveness and repulsiveness for particular ages using data from a population census or survey. Respondents tended to round or heap their ages. The magnitude of this type of error is

variable by sex. The indexes were applied in order to determine the degree of accuracy of data by age that is used in comparing population structures at different dates within the same country, between different sub-population within the same country: male, female, urban, rural. However, the application of these techniques requires population distribution by year or a single age data. The indexes to ascertain in this study are Whipple's index, Myer's index and Combined index of United Nations.

3.6.2.2.1 Whipple's Index

The Whipple's index is the most applied measure of the quality of age reporting. This index measures the attractiveness or the repulsiveness of ages ending either by 0 or 5.The calculation may be done for each sex. The formula is:

$$W = \frac{\sum_{i=5}^{12}(P_{25} + P_{30} + P_{35} + ... + P_{55} + P_{60})}{\frac{1}{5}\sum_{j=23}^{62}(P_{23} + P_{24} + P_{25} + ... + P_{61} + P_{62})} * 100$$

$$= \frac{\sum_{i=5}^{12} P(5i)}{\frac{1}{5}\sum_{j=23}^{62} P(j)} * 100$$ P= Population of the age group; i = range from 5 to 12; j = range from 23 to 62.

UNIVERSITY of the WESTERN CAPE

Fundamentally, the Whipple's index was obtained by summing the age returns ages 23 and 62 inclusive and finding what percentage is born by the sum of the returns of years ending with 0 and 5 to one fifth of the total sum. The result should lie between 1 and 5. However, the choice of age limits 23 and 62 is arbitrary, but it is convenient in practical.

In the same vein using a 10-year range, the heaping may be evaluated in the range 23 to 62 inclusive, comparing the sum of the population at the ages ending in "0" or "5" and in this range with 1/10 of the total population in the range.

$$W = \frac{\sum_{23}^{62}(P_{30} + P_{40} + P_{50} + P_{60})}{\frac{1}{10}\sum_{23}^{62}(P_{23} + P_{24} + ... + P_{61} + P_{62})} * 100$$

$$= \frac{\sum_{i=3}^{6} P(10i)}{\frac{1}{10}\sum_{j=23}^{62} P(j)} * 100$$

However, in this study the application will be made only for 5-year range. The interpretation of Whipple's index (W) in regard to the recommendations of UN is: If the value of W is less than 105, the age distribution data is highly accurate. There was no attractiveness or repulsiveness for all declarations of age, the age was properly declared by all respondents; if the value is between 105 and 109.9, the age distribution data is fairly accurate; if the value of index lies between 110 and 124.9, the age distribution is approximate; if it lies between 125 and 174.9, it is rough; and if it is 175 or more, the age distribution data is very rough (UN, 1990). The Whipple's index is simple to calculate; but, the real problem of Whipple is that it focuses its checking only on two digit endings 0 and 5.

3.6.2.2.2 Myers Blended Method

UNIVERSITY of the
WESTERN CAPE

Due to certain influences of culture in the community, some individuals tended to report their age with digits ending with 1, 2, 3, 4, 5, 6, 7, 8, 9, 0, leading to abnormal data distribution. The phenomenon of age heaping of response in certain ages has to be controlled through Myer's index which analyses all digits to check the preference or avoidance. Myer's index examines the age misreporting and indicates the preference for certain terminal digits in age reporting. The approach is used as a mechanism to count each of the 10 digits in turn and then to average the results. It calculates the proportion that the population ending in a given digit in comparison to the total population 10 times, by varying the particular starting age for any 10 year age group.

The method of calculation would be done for each sex.

Step1: sum the population ending in each digit over the whole range, starting with the lower limit of the range for example 10, 20, 30,…, 11, 21, 31, … and so on. The formula is for the age range 10-69:

$$S_u = \sum_{d=1}^{9} P(10d + u)$$

Where d represents the deciles digit for age and u, the unit digit for age,

Su is the absolute figure of individuals aged 10 years and over for which age ends with each of the numbers beginning from 0 to 9.

Step2: Ascertain the sum excluding the first population combined in step1 for example 20, 30, 40, … , 21, 31, 41, … , 22, 32, 42, …,

$$S'_u = \sum_{d=1}^{9} P(10d + u)$$

Step3: Weight the sums in steps 1 and 2 and add the results to obtain a blended population for example weights 1 and 9 for the 0 digit, weights 2 and 8 for the 1 digit. Myers considers the absolute figures named T_u which represents the blended population.

$T_u = (u + 1)S_u + (9 - u)S'_u$

$T_o = S_o + 9S'_o$

$T_1 = {}_{(1+1)}S_1 + {}_{(9-1)}S'_1 = 2S_1 + 8S'_1$

$T_2 = {}_{(2+1)}S_2 + {}_{(9-2)}S'_2 = 3S_2 + 7S'_2$ and so on

UNIVERSITY of the
WESTERN CAPE

$$T = \sum_{u=0}^{9} T_u$$

Step 4: Convert the distribution in step 3 into percentages

If consider A is a new distribution resulting from this formula:

$A_u = 100T_u/T$ you will have for instance, $A_1 = 100T_1/T$; $A0 = 100T_0/T$, and so on.

Step5: Take the deviation of each percentage in step 4 from 10.0, the expected value for each percentage. The sum of these expected values represents the Myers index (I_2).

$$I_2 = \sum_{u=0}^{9} \{(100T_u / T) - 10\}$$

The results in step 5 indicated the extent of concentration on, or avoidance of a particular digit. The weights in step 3 represented the number of times the combination of ages in step 1 or 2 was

included when the starting age varied from 10 to 19. If blending is not applied, the results are very sensitive to the choice of the particular starting age, and the frequency of the digits shows a substantial decline from 0 to 9. With blending, the frequency of the digits is almost equal.

The weights for each terminal digit will differ if the lower limit of the age range covered were different. This approach expresses that an index of preference for each terminal digit representing the deviation from 10 per cent of the proportion of the total population reporting ages with a given terminal digit. A sum total index of preference for all terminal digits is derived as one - half (1/2) the sum of the deviation from 10 per cent each taken without regard to sign.

Myer's considers the absolute number of the sum of deviation. If age heaping is nonexistent, the index would approximate zero. This index is an estimate of the minimum proportion of persons in the population for whom an age with an incorrect final digit is reported. The theoretical range of Myers's index is 0, representing no heaping, to 90 which would result if all ages were reported at a single digit say zero.

However, slight deviations demonstrated by different measures of heaping are not automatically indicative of heaping. In reality, the population in a single year of age is by no means equal to exactly one-fifth (1/5) of the 5-year age group centering around that age (nor one-tenth of the 10-year age group centering around the age) nor is there necessarily a gradual decline in the number of persons from the youngest to the oldest age in a broad group as is assumed in the formulas.

UNIVERSITY of the
WESTERN CAPE

The age distribution may have small irregular fluctuations, depending largely on the past trend of births, deaths and migrations. The errors in the data for extreme old age do not yet show up. Past fluctuations in the number of births, deaths and migrants may still affect the figures, it is not possible to measure digit preference precisely because there is no precise distinction between the error due to digit preference, other errors and real fluctuations cannot be made.

3.6.2.2.3 Combined Index of United Nations

Shryock & Siegel (1976) found that the quality of age and sex data can be assessed by the age - sex accuracy index which is the sum of age ratio score for males, age ratio score for females and three times the sex ratio score. The United Nations (1983) brought out some limitations, the permissible values of the age ratio score for males is 2.6 for females 2.4 and the permissible

value for the sex ratio score is 1.5. The combination of these three values brought the permissible limit for the age-sex accuracy is 9.5 which represent [(2.6+2.4)+(3* 1.5)]. Based on the empirical analysis of the age and sex data available from population census in different developed and developing countries, the United Nations recommended that the age and sex structure of a population may be classified as accurate if the age-sex accuracy index is under 20, inaccurate if the index is between 20 and 40, and highly inaccurate if the index is above 40 (UN, 1983). Arriaga (1994) found that the age-sex accuracy index is useful mainly in international and historical analysis. The variation of the index over years shows whether the quality of population reports on age and sex is improving or deteriorating. Arriaga (1994) also demonstrated that the age and sex accuracy index and its components quantify only the quality of the age and sex distribution.

Procedure of calculation age-sex accuracy index:

Step 1

The age ratio is calculated for each age group and for each sex. For age group x, x+4 is:

$$C_i = \frac{P(x,x+4)}{[P(x-5,x-1)+P(x+5,x+9)]/2} * 100 \qquad i = 1, \ldots, n.$$

UNIVERSITY of the
WESTERN CAPE

$$C_1 = \frac{P(0-4)}{P(5-9)} * 100 \quad C_2 = \frac{P(5-9)}{\frac{1}{2}[P(0-4)+P(10-14)]} * 100 \quad C_3 = \frac{P(10-14)}{\frac{1}{2}[P(5-9)+P(15-19)} * 100$$

Step 2

Compute deviation $D_i = 100 - Ci$ where I represents the age group (x_i , $x_i + 4$). Then just consider the absolute number of $D_i = / Di /$, sum up for each sex which means $\sum_{i=1}^{n} / Di / = D$

If D_f represents the sum of deviations for females, the formula is $D_f = \sum_{i=1}^{n} / Dif /$

Also for males is $D_m = \sum_{i=1}^{n} / Dim /$

Step 3

Calculate the male ratio (E_i) using the number of males dividing by the number of females for each group (i), times 100 with i = 1, 2,…,n.

$E_i = \frac{Males}{Females} * 100$ $E_1 = \frac{Males1}{Females1} * 100;$ $E_2 = \frac{Males2}{Females2} * 100$

Step 4

Compute $Fi = E_{i+1} - E_i$ for i = 1, 2,…, n. $F_1 = E_2 - E_1;$ $F_2 = E_3 - E_2;$ $F_3 = E_4 - E_3$; and so on.

Notably the age group 70-74 was not included in the computation of Fi. The United Nations recommended considering the age groups below 70 years old because above this category of population, they are facing different patterns in their declaration of age. However, the old ages are not included in the calculation of index because they don't have the same characteristics, and they are often affected for instance by overestimation of its ages.

Step 5

Compute the index of male ratio (M_i); $M_i = \frac{\sum_{i=1}^{n} Fi}{n}$

UNIVERSITY of the
WESTERN CAPE

Step 6

$CIUN = D_{males} + D_{females} + 3M$

However, the serious limitation of the index is that it does not consider the expected decline in the sex ratio with increasing age and the real irregularities in the age composition due to normal fluctuations in fertility, mortality and patterns of migration. The greater weight allocate to the sex ratio is arbitrary, there is no reason and logic in estimating the age-sex accuracy index. Hobbs (2004) demonstrated that because of these limitations the age-sex accuracy index is rough in comparisons among populations. The usefulness of this index appears in its ability to draw attention to extreme values due to under enumeration and misreporting in the age and sex data. This index also captures some of inaccuracies due to unusual changes in the levels of fertility and mortality, and patterns of migration.

3.6.2.2.4 Sex and age ratios

A ratio is a comparison between two things. Dictionnary.com defines it as the relation between two similar magnitudes with respect to the number of times the first contains the second. Some content errors can be estimated by the use of the sex and age ratios.

3.6.2.2.4.1 Age ratio

The age distribution of the population could be analyzed in terms of age ratios for 5 year age groups (UN, 1983). The age ratio is the ratio of the population in a given age group to one-half the sum of the population of the preceding and following groups, times 100. The formula is for a 5-year age group as follow: ARS = $\frac{{}_5P_a * 100}{1/2[{}_5P_{a-5} + {}_5P_{a+5}]}$; ${}_5P_a$ is the population in the given age group, ${}_5P_{a-5}$ is the population in the preceding age group and ${}_5P_{a+5}$ is the population in the following age group. If there is no extreme variations in the past vital events (death, birth, migration), the age ratios for all age groups should be approximated to 100. Arriaga, (1994) found that the age ratios should be fairly similar across the age categories. However, if there is no abrupt fluctuations in fertility, mortality and migration which are generally age-selective, and if the number of persons in successive age groups gradually decreases through the incidence of mortality, the age ratio for any age group should be closed to 100. The contrary could also happen if there are fluctuations in fertility, mortality, or if migration is significant, age ratios for different groups can deviate considerably from 100 (normative value).

UNIVERSITY of the
WESTERN CAPE

Another possibility is that the age ratio can also deviate from 100 if there is misreporting of age or differential omission in the enumeration of persons belonging to a given age or both. The accuracy of the age data is not the condition that the age ratios cannot deviate from the normative value. The age curve is not linear. It is very difficult to measure the errors in grouped data with any precision, most of the techniques of assessing grouped data cannot generate absolute estimates of census errors by which census data can be corrected. However, the assessment of the data on age for age groups may result by comparing age ratios derived from the census data with expected values.

For some ages, the age ratios may deviate considerably from 100 even where reporting of age is good. An age accuracy index is calculated by taking the average deviation (without regard to sign) from 100 of the age ratios over all ages. The average of the age ratios scores for males and females is treated as a measure of the overall accuracy of the age data across the two sexes and is often referred to as the age accuracy index.

3.6.2.2.4.2 Sex ratio

The sex ratio (S) is an indicator of the population calculating by dividing the absolute figure of males in a given age group by the absolute figure of females in the same age group times 100.

$$S = \frac{Males}{Females} * 100$$

Generally, the ratio of a male to a female birth is assumed to be 105 boys for every 100 girls. The survival probability of a girl is higher than that of boy because of several factors. Females have a survival advantage compared to males because they have two X chromosomes and males one. Even greater level of estrogen in premenopausal period of women almost certainly protects them against the development of heart diseases. Women have a physiological advantage to adapt to changes in the environmental temperature (Waldron, 1976; Austad, 2006). In developing countries, the living conditions of women with social discrimination against the fair sex and the risk of death combined with the complications of pregnancy and delivery expose women at a disadvantage to men. In the same vein, the sex-selective migration of the migration of the working age population affects the ratio of males to females in different age groups. If the effect of migration is removed, then the sex ratio is very high 105 at age 0. The sex ratios are then expected to decline with age, till a ratio of around 100 for persons in their late 20s, and continue to decline around 50 to 60 in the oldest ages. (Poston & Davis, 2009). This situation is the result of lower female mortality as compared to the male mortality.

In Africa, the variation of S is significant from one age to another. However, a sex ratio of more than 100 shows an excess of males over females while a sex ratio of less than 100 shows an excess of females over males. A sex ratio of 100 indicates an equal number of males and females in the population. The information on sex ratio contributes to assessing the quality of data on age. The expected early age ratio is slightly greater than 100. The pattern of sex ratio highlights

UNIVERSITY of the WESTERN CAPE

the sex-age-selectivity in the age distribution and infanticide due to strong preference of sons. The presence of errors of omission, age misreporting and out-migration may be detected by looking at the pattern of sex ratios. The normative age pattern of sex ratios is disturbed by extreme forms of a number of manmade interventions such as war and conflicts which affect the size of the male population and discrimination against females at individual, family and community of which the visual example is the female abortion and female infanticide and which affect the size of female population. However, in the absence of fluctuations and variations in the demographic processes, the accurate age data are rectangular distributed and the age specific sex ratios decline over the life cycle in an equal manner.

Based on the fact that age and sex indexes quantify only the quality of the age and sex information, Arriaga (1994) suggested that, the age and sex data must also be examined graphically.

3.6.2.2.5 The population pyramid

The age and sex distribution of the population can be evaluated in various methods. Some techniques analyze the age data separately for the sex; another cross tabulated the age data by sex. The population pyramid is a basic procedure for assessing the quality of data on age and sex. It can be presented in three ways. Firstly, it makes use of absolute numbers; secondly, it considers the population in each age group as proportion to the total population, male and female combined, while thirdly, it considers the males in each age group as proportion to total males and females in each age group as proportion to total females. The interpretation of population pyramid differs from one way to another; each one has its own understanding. As far as this study is concerned, the researcher would apply the third method of presenting which reflect males in each age group as proportional to total males and females in each age group as proportional to total females. The base of the pyramid is mainly determined by the level of fertility in the population, while the peak is determined by previous level of mortality and fertility. The level of migration by age and sex also affects the shape of the pyramid.

3.6.3 Limitations in the Comparisons

The study on assessment of quality demographic data on age and sex focuses on the declaration on age and sex of respondents (male and female). The study is not able to collect information that identifies areas of residence as rural or urban. It considered the nine provinces as whole South Africa. The analysis focused specifically on population aged 0 to 70. This was initiated in order to avoid the effect of size. More than 95 per cent of the whole population belongs to the age range between 0 and 70. According to the census and survey, the geographic boundaries and dwellings were always changing over the years. Some households and dwellings had no unique identifier (address details). However, a dwelling counted in a given survey may no longer be in existence the next time the following survey takes place thus data accuracy becomes a concern. The study compares 2001, 2004, 2005, and 2007 data sets, and the structural changes in terms of the quality of declaration on age and sex across South Africa are provided. The next chapter dwells on the findings of this research. Comparison across gender, among population groups and between instruments will also constitute a major focus.

UNIVERSITY of the
WESTERN CAPE

CHAPTER FOUR

RESULTS FROM THE DATA ANALYSIS

4.1 INTRODUCTION

The analytical objective of this study was to ascertain the quality of data on age and sex collected from selected censuses and surveys in South Africa. The first part of this data presentation focuses on the description of the observable characteristics and statistics, then, the second one concentrates on the findings on gender and population groups, comparison across provinces to see the improvement across the time and between different instruments over years. The sex ratios and age pyramid are part of these comparisons to highlight these differences.

4.2 Characteristics of Data

UNIVERSITY of the
WESTERN CAPE

Although the data characteristics engage to ascertain the quality of variables such as age and sex (UN, 2008), it has been argued by (Poston, 2005) that age is a more complex demographic characteristic than sex. According to the UN recommendations age is defined as the interval between the date of birth and the date of census expressed in completed years (Poston, 2005). Using this definition, the comparisons of data characteristics on age and sex were made possible between the following instruments 2001 Census and community survey 2007; General household survey 2004 and 2007; Labour force survey 2005 and 2007. All instruments were recent and capturing information on the quality of declaration of age and sex of respondents based on variables such as gender, age, age group, population group, marital status, provinces and relationship to the head of household.

This study intends to ascertain the quality of the data on sex and age for the period of 2001 to 2007. However, several changes which are the consequences of the fluctuations in fertility, mortality, patterns of migration and under enumerations that happened over previous years were observed. In seven years, the population faced changes. In 2007, the total population of South Africa was estimated at 47 850 700. The decline of fertility and mortality rates has strongly contributed to variations in age and sex distribution of the population in South Africa. In order to avoid sampling errors, total enumerations data were used in the analysis of age and sex

distribution at the national, provincial and among population groups. The Statistics SA even claims that the 2001 sample was being designed to be not only nationally representative but also at provincial level representative. It is clear that the 2001 sample was based on the 1996 census. However, the majority of surveys in this study had applied this particular sample, thus it is useful to cross check this representativeness using the age distribution. The sample sizes of households and individuals are large and sampling errors are relatively small for most estimates.

The data evaluation measures the magnitude of error in variables age and sex for the application of the indexes using Whipple's, Myers, age and sex ratios with Combined index of United Nations and population pyramids. These indexes would give the level of magnitude of errors occurred in the data of each instrument, the improvement done overtime in comparing by gender, by population group, by provinces. The check is limited to individual responses in such a way Preston *et al.* (2001) reported that error assessment is important even if correction is not feasible because it shows the degree of confidence that users could be placed in the statistics. Most indexes are explained across the surveys which allow reliable comparisons between GHS 2004 and GHS 2007, LFS 2005 and LFS 2007, GHS 2007 and LFS 2007, GHS 2007 and Community 2007, LFS 2007 and Community 2007; and with Census 2001.

UNIVERSITY of the
WESTERN CAPE

4.3 Method and Techniques

The main objective of this research is to examine and compare the age-sex structure of population and the main methods are cross-tabulations, charts and population pyramids. The evaluation of data quality would focus on such issues as age heaping, over age reporting, for old age and under reporting of children under age 5. The heaping may occur in some ages, as a result of digit preference or rounding generally to numbers ending in 0 or 5. The age heaping would influence the analysis of results. The evaluation of age heaping is completed using Whipple's index and Myer's index.

The main hypothesis in applying these measures is that the actual age structure in a population showed gradual variations. Age data revealed age heaping if there is a concentration of people reported in ages with specific ending digits. Whipple's index and Myer's index are applied to assess the quality of single age data.

4.4 Results of Indexes

4.4.1 Results of Whipple's Index

The application of Whipple's index is done with the assumption that the population is equally distributed by age. The formula used is $[5*\sum P(i)]/\sum P(j)]*100$. Whipple's index is ranged between 100 and 500. The evaluation of quality in responses on age based on Whipple's index used the following standards: first, if Whipple's index is less than 105, the quality is very accurate; second, if the value vary from 105 to 110, it is relatively accurate; if the value of index lay between 110 and 125, the quality of responses on age is fair; if the index range from 125 to 175, the quality of responses on age is bad, if the value is equal to 175 and more, it is very bad.

4.4.1.1 Results of Whipple's Index for General Household Survey at National Level and Provincial Level

4.4.1.1.1Results of Whipple's Index for General Household Surveys at National Level

Table 4.1 shows the results of Whipple's Index at national level for GHS 2004, the index of male is 104.54 against 105.69 for females. However, the male index is lesser than 105 but greater than 100 while the females index is greater than 105 but lesser than 110, this implies that the quality of data at national level for males is accurate whereas for females it is relatively accurate. However, the ending digits 0 and 5 are preferred. From the above indexes in Table 4.1, it is likely that the quality of data is accurate for males than females. The index for both sexes is 105.69 showing that the quality of data distribution is relatively accurate.

The Table 4.1 indicates also the results from GHS 2007. It is observed that the males' index is 106.86 while females had 108.18. However, the two indexes are ranged from 105 to 110, which meant that the quality of data for males and females in GHS 2007 are relatively accurate and the two ending digits 0 and 5 are preferred. The index of both sexes is 107.59 indicating that the quality of its age distribution is relatively accurate. It is probably that male respondents did better declaration of age compared to females.

It was postulated that there is an improvement at national level over years by instruments used and across gender. Comparing GHS 2004 and GHS 2007, it is a bit controversial that, the index of males has increased as well as the female's index, the preference is manifested. The deterioration is observed in all indexes, there is no improvement in the statement of ages. The enumerators could be responsible for this lack of empowerment if they did not master properly what to do on the field work. Over years, for GHS 2007, the quality of data is likely inaccurate compared to GHS 2004 in both sexes.

The Table 4.1 points out the results using the LFS 2005. The males' index was 108.73 which was upper close to 110 the limit of the range. This implied that the quality of age distribution was relatively accurate while some respondents had rounded up their ages. According to females, their distribution was fairly accurate with the index equals to 111.9 indicating that they were attracted in ages with ending digit 0 or 5. Males probably did better declaration of age compared to females. The index of both sexes was 110.47, showing that the quality of its distribution is possibly fairly accurate.

The same Table 4.1 designates the results from LFS 2007 where male and female indexes were very close (105.05 and 105.04 respectively). Their age distribution was relatively accurate which showed that they still have to improve; although, the female respondents probably did better declaration of age than males. Possibly, female respondents had become more conscious in the declaration of age than males. This could also be the result of great effort that the government engaged in the whole process of collecting data that was of good quality. (Starting from the recruitment of personnel, training, techniques and different methodologies applied as approaches of collection, advices provide to enumerators even concerning some particular areas…).

The expectation was that there was an improvement at national level over years by instruments used and across gender. Observing LFS 2005 and LFS 2007, the improvement is progressive. The index of males had decreased from 108.73 to 105.05 while for females has decreased from 111.9 to 105.04 and the index of both sexes 110.47 to 105.04. The set of them still have to bring effort. However, the engagement of female respondents to declare the true age is significant, hence the slight preference continue to occur in their data on age. The possible explanation refers to the large publicity surrounding the operation of LFS 2007. The enumerators had applied all the advices and guides they had received during the training and when going on field work. Over

years the quality of data on age of LFS 2007 is probably relatively accurate compared to LFS 2005.

Definitely, at national level, the best Whipple's index of males had occurred in GHS 2004 (104.54), for females in LFS 2007 with value of 105.04 and the index of both sexes 105.04 in LFS 2007. The highest value of index for males appeared in LFS 2005 (108.73), for females 111.9 in also LFS 2005 and the index of both 110.47 also in LFS 2005.

4.4.1.1.2 Results Whipple's Index for General Household Surveys at Provincial Level

The analysis at provincial level aim is to achieve the following objectives: First, to describe and to analyze the quality of data on age and sex collected from census and surveys had been changed over years across the provinces; Second, to describe and analyze the quality of age declaration is poor at provincial level compared to national level; Third, to see and evaluate the contribution of gender to the poor quality of age declaration.

Table 4.1 shows the results of Whipple's Index at provincial level for GHS 2004, the males' index could be grouped into four categories. The first concerns Western Cape which is equal to 98.97; Free State 96.58 and North West 98.45 which are less than 100 showing that the avoidance in ages with ending digit 0 and 5. Their age-sex distribution is probably very accurate because the deviation to 100 is less than 5 per cent indicating a very slight avoidance that this population observed in numbers with ending digit different from 0 or 5. The second group constitutes Limpopo with Whipple's index equal to 102.82, Mpumalanga 102.84 and Northern Cape 103.63 for male respondents. Males likely did better declaration of age than females in these provinces. The quality of age reporting is probably very accurate among them. The third group is represented by Gauteng with 107.20, and Eastern Cape 109.11. Their age distribution is relatively accurate. The declaration of male respondents highlighted preference in numbers with ending digit 0 or 5. The last is KwaZulu-Natal with Whipple's index 112.38 for male respondents; the quality of data on age is probably fairly accurate. Males rounded their ages or preferred ages with ending digit 0 or 5. It seems that provinces strongly urbanized are with very accurate quality of data on age.

UNIVERSITY of the WESTERN CAPE

The same Table 4.1 provides the results of GHS 2004, for females' index at provincial level. This can be grouped into four categories as follow; the first group is represented by Northern Cape with 96.55, North West with 96.90 and Limpopo with 96.37. These indexes are less than 100 with the deviation less than 5 indicating the quality of female responses can be regarded as very accurate. The female respondents were avoided in ages with ending digit 0 or 5. The quality of their age distribution is likely very accurate. The second group considers Eastern Cape with 107.13, Free State with 107.21 and Gauteng with 109.28 for females Whipple's index. In this group, the quality of age distribution is relatively accurate. The female respondents were attracted in ages ending with digits 0 or 5. The third group is constituted by KwaZulu-Natal with 112.08 and Mpumalanga 113.78. The preference in ages ending with digit 0 or 5 is likely more pronounced among females in this category compared to others. The quality of age distribution is fairly inaccurate and the last group is represented by a single province, Western Cape (104.37) showing quality of data on age accurate, hence it is little closed to upper limit 105.

As far as the index of both sexes is concerned, the Table 4.1 designated the results using GHS 2004 at provincial level. These results were classified into 4 groups as followed; the first category considered Northern Cape (99.95), North West (96.9) and Limpopo (98.92) which the Whipple's index was less than 100, the set of them were avoided to numbers with ending digit 0 or 5. The age distribution of these indexes was very accurate. The second was constituted of Western Cape (101.9) and Free State (102.04) which the quality of data is very accurate. The third represents Eastern Cape (107.98), Gauteng (108.23) and Mpumalanga (107.77) showing the quality of age distribution is relatively accurate. And the last is a single, KwaZulu-Natal (113.17) revealing the quality of its age distribution is fairly accurate.

UNIVERSITY of the
WESTERN CAPE

The expectation was that gender contributed to the poor quality of declaration on age. When comparing the results of males to females in GHS 2004 at provincial level, it reveals that males probably did better declaration of age than females in, Western Cape, Free State, KwaZulu-Natal, North West, Gauteng, Mpumalanga and Limpopo; hence the females are likely declared better compared to males in Northern Cape and Eastern Cape. However, the age reporting of males is probably much better compared to females. The smallest index for males is 98.97 (Western Cape) and the highest is for KwaZulu-Natal (112.38) and the smallest for females is 96.9 (Northern Cape) while the highest is 113.78 (KwaZulu-Natal). Possibly, male respondents

did better declaration on age than female respondents. The results of KwaZulu-Natal could be explained by the issue of education; the population appeared not to have understood properly the question asked or the level of illiteracy seems to be high in that province and the respondents did not remember their dates of birth or might be the training of fieldworkers was not perfectly achieved or they were not also be able to translate properly the questions or they did not follow the received instructions.

The hope was that there is an improvement at provincial level over years by instrument used and across gender. Considering the GHS 2007, Table 4.1 indicates that the results of males at provincial level can be clustered into four groups. Western Cape with 103.62, Northern Cape with 103.86, and Mpumalanga with 104.38 constitutes the first group. Free State (98.5), North West (95.73) and Limpopo (99.37) are represented as the second group. The quality of age distribution is likely very accurate for the first and the second groups and male respondents probably did good declaration of age. The attractiveness or the avoidance in ages with digit ending 0 or 5 was not significant in these provinces concerning males. KwaZulu-Natal 116.03 and Gauteng 113.69 for males' index reveal that the data age distribution is possibly fairly accurate. Male respondents did not properly declare their ages. They tend to round off or are attracted to ages ending with digit 0 or 5. The male index of Eastern Cape was 105.94 showing that the quality of data on age s relatively accurate.

UNIVERSITY of the
WESTERN CAPE

Table 4.1 also shows that the female results of Whipple's index could be grouped into three categories. The first falls between 105 and 110 and comprises Western Cape, Eastern Cape, Northern Cape, Gauteng and Limpopo. The quality of their data was relatively accurate. Females were attracted to numbers with digit ending 0 or 5. They tend to have less consideration for the survey purpose which influences their outcomes. The second represents the Free State (112.62) and KwaZulu-Natal (113.05). The quality of age distribution is likely fairly accurate. The third group is constituted of North West (100.89) and Mpumalanga (104.25) showing that the quality of the data on age is probably very accurate. In these particular provinces female respondents probably did better declaration of ages.

Considering Table 4.1 using the GHS 2007 according to both sexes index, the results could be grouped into four categories; the first considers Western Cape, Eastern Cape, Northern Cape and Free State with results lying between 105 and 110, revealing that the quality of data is relatively

accurate; the second took into account KwaZulu-Natal (114.31) and Gauteng (110. 94), the quality of their data is fair. The third group considered Mpumalanga (104.30) and Limpopo (103.95) showing that the quality of data is very accurate and the last category is North West with 98.62 as Whipple's index which highlights the avoidance in ages with digit ending 0 or 5. The deviation to 100 of North West index is equal to 1.38 per cent which is less than 5 per cent revealing the accuracy of data quality.

Table 4.1 also displays the results of Whipple's index applying the GHS 2007. This reveals that male respondents in Western Cape (103.62), Northern Cape (103.86), Free State (98.5) and Limpopo (99.37) did better declaration of age compared to the female respondents. Based on the results, females had possibly declared their age better than males in Eastern Cape (105.94), KwaZulu-Natal (113.05), North West (100.89), Gauteng (108.03) and Mpumalanga (104.25). The highest index for males is 116.03 (KwaZulu-Natal) and the smallest is 99.37 (Limpopo). The highest index for females is 113.05 (KwaZulu-Natal) and the smallest is 100.89 (North West). The best Whipple's index for males is 99.37 (Limpopo) because it is more close to 100, the perfection; and for females is 100.89 (North West).

UNIVERSITY of the
WESTERN CAPE

4.4.1.2 Comparison Results of Whipple's Index between GHS 2004 and GHS 2007 at provincial level

The hope was that there is an improvement at provincial level over years by instrument used and across gender. As far as the results of GHS 2004 is concerned, the distribution of males' index was close to 100 than for GHS 2007 which means the indexes were generally closer to one another in 2004 than the index of males generated for GHS 2007. With GHS 2004, six out of nine indexes are ranged between 95 and 100, also between 100 and 105; and two indexes are ranged between 105 and 110, then the last ranged between 110 and 125. The same observation can be also done for GHS 2007. Distribution of indexes resulting from GHS 2007 constitutes six out of nine between 95 and 100, then 105 and 110 and the last two between 110 and 125. However, the smallest index of males is 96.58 (Free State) and the highest is 112.33 (KwaZulu-Natal) for GHS 2004 while the smallest females' index is 96.37 (Limpopo) and the highest is 113.78 (KwaZulu-Natal). For the second distribution, the smallest males' index is 95.73 (North

West) and the highest is 116.03 (KwaZulu-Natal) according to GHS 2007 while the smallest females' index is 100.89 (North West) and the highest is 113.05 (KwaZulu-Natal).

A slight improvement appeared in some provinces such as Mpumalanga in which the index of females decreased from 112.08 to 104.25, the age distribution of male was highly accurate and for female the quality was fair has changed to become very accurate. In the North West where the females' index increased little from 96.9 to 100.89, the improvement is progressive; in Eastern Cape males' index decreased from 109.11 to 105.94 hence male respondents still have to bring new effort in terms to decrease its index which would impact the quality of its response.

Referring to the Table 4.1, as far as the Whipple's index of KwaZulu-Natal is concerned, the GHS 2007 compared to GHS 2004 reveals no improvement achieved in the assessment of the quality of data on age and sex. The index of males has increased between the two surveys instead of decreasing. The situation became worse. There is no credibility in their responses but possibly, some demographic factors have influenced the declaration of respondents which could be education, culture, marriage and so on.

For the North West province, the Table 4.1 sets out when comparing the females' index of the two GHS, the improvement is progressive over years; but the contrary appears in the males' case. The females' index had increased and the males were strongly repulsive to ages ending with 0 and 5. It could be the respondents were scared to answer some embarrassing questions or they did not have the interest in the survey or they did not need to be involved in such a contract. The index of males in GHS 2007 reveals the preference in other digit endings different from 0 and 5.

Looking at the results of Limpopo province from Table 4.1, the males' index of the two GHS are close to the perfect 100. The males' index of 2004 is 102.82 possibly giving the declaration on age by respondents to be very accurate. The one of GHS 2007 is 99.37 showing very slight repulsiveness in the declaration of age ending with 0 and 5. The quality of its age distribution is very accurate. Indeed, the females' index in 2004 is 96.37 indicating also slight repulsiveness in the declaration on age and probably very accurate age distribution. The following GHS brings out 106.88 for the female's index showing that the age distribution is possibly relatively accurate

UNIVERSITY of the
WESTERN CAPE

and the preference was pronounced in numbers with digit ending 0 or 5. However, the declaration on age of males was in all probability better than females across the two surveys.

4.4.1.3 Results Whipple's Index for Labour Force Surveys at Provincial Level

Table 4.1 shows the results of LFS 2005 for males. Western Cape, Northern Cape and North West provide good outputs showing that the quality of data on age is likely very accurate. These indexes are less than 105. Males properly declared their ages with no preferences or avoidances. Hence, Free State, Eastern Cape, KwaZulu-Natal and Limpopo results vary between 110 and 125 presenting the quality of response on age is possibly fairly accurate. Their age structure is approximate. Gauteng (109.89) and Mpumalanga (106.84) enhance the preference in the attitude of male respondents who tended to round their ages in numbers with digit ending 0 or 5. The quality of their responses is probably relatively accurate.

Referring to the Table 4.1, female results in LFS 2005 can be grouped into four categories as follows; the first was Eastern Cape, North West, Gauteng and Mpumalanga which results ranged between 110 and 125 showing that the quality of responses on age is fairly accurate. Their age structure is approximate; hence the declaration on age is poor. These female respondents did not properly state their ages. The second constitutes Northern Cape, Free State and Limpopo in which their results varied between 105 to 110 showings that the quality of declaration was relatively accurate, however females preferred to round the ages in numbers ending with 0 or 5. Western Cape (101.03) represented the third category with a good result. Its age structure is possibly very accurate according to the Principe of Whipple's Index. The last category is KwaZulu-Natal (127.39) indicating that the quality of response on age was bad. The preference was probably significant in this group. The phenomena of age heaping were consistent among females. A possible explanation to this case of KwaZulu-Natal is that digit preference is most pronounced among population having a low educational status and also the patterns of digit preference vary from one culture to another.

Regarding the Table 4.1, the index for both sexes using the LFS 2005, Eastern Cape, Free State, Gauteng, Limpopo and KwaZulu-Natal were categorized in the same interval 110 to 125. The quality of their declaration on age is fairly accurate. Northern Cape, North West and

Mpumalanga outcomes are ranged between 105 and 110, indicating that the quality of age structure is probably relatively accurate. The synthetic index of Western Cape (101.32) is perhaps the best compared to other provinces; its age structure is probably very accurate.

Definitely, in the Western Cape, the respondents are likely to properly declare their ages. Feskens *et al.* (2006) suggested that it is more useful to make questionnaire more understandable to reduce the bias in the responses. This could be a possible alternative explanation to this achievement in Western Cape. It could be also explained by the achievement realized in the organization of the whole operation of the LFS 2005. Everyone involved in the procedure provides the best in terms of achieving good results. For instance, the enumerators are well trained and the fieldwork is properly done; consequently, the quality of the data on age is doubtless, highly accurate. The best and smallest Whipple's Index is 101.12 (North West) for males and 101.03 (Western Cape) for females among all provinces. The highest index is 119.54 (KwaZulu-Natal) for males and 127.39 (KwaZulu-Natal) for females avoidance of 0 and 5 is really significant for LFS 2005 for both sexes; hence no province registered an outcome less than 100.

Seeing the Table 4.1, applying the Whipple's index to the LFS 2007 at provincial level, Eastern Cape, Western Cape, North West and Gauteng obtained the very accurate quality of response on age purpose. Male respondents were likely to declare their age properly. The indexes were less than 105. Northern Cape, KwaZulu-Natal and Mpumalanga, according to index of males were between 105 and 110. The quality of data on age was relatively accurate and males' respondents preferred ages with digit ending 0 or 5. Free State males provided (110.31) the highest index of Whipple across provinces. The quality of their declaration on age was probably fairly accurate. Limpopo males with 97.37 indicated slight avoidance in numbers digit ending different from 0 and 5. The deviation to the perfect point was 2.63 which is less than 5 per cent showing that the quality of age structure is likely very accurate.

When considering females, Western Cape, Eastern Cape, Free State, North West, Gauteng and Mpumalanga results are less than 105 which mean the quality of responses is very accurate. Females of Northern Cape and KwaZulu-Natal have fairly accurate indexes ranging between 110 and 125. According to Limpopo with 95.53 of index, indicating the quality of age structure is

probably very accurate hence the deviation is 4.47 less than 5 per cent, meaning the female respondents slightly avoid numbers with digit ending 0 or 5.

Table 4.1 shows the index of both sexes for LFS 2007. These indexes are grouped into four categories as follow; first, Western Cape, Eastern Cape, North West and Gauteng with index less than 105 presented very good data quality which means the age structure is uniformly distributed. The second category is constituted by Free State, KwaZulu-Natal and Mpumalanga with results ranging between 105 and 110. The quality of data on age is likely relatively accurate. Third, Northern Cape with 110.28 indicating the quality of age distribution is probably inaccurate. However, Northern Cape highlights highest index as far as the synthetic index is concerned when applying the LFS 2007. This value indicates that the age structure of the synthetic index is fairly accurate. Limpopo records 96.27 showing that the quality is likely very accurate, but the very slight repulsiveness in numbers with digit ending 0 and 5 is observed.

The LFS 2007 results reveal some achievements across provinces. Population is becoming more familiar with this operation of survey and the consideration can affect its attitudes. Females of Western Cape, Free State and Mpumalanga probably did better declaration of age than males in their provinces; hence males of Eastern Cape, Northern Cape, KwaZulu-Natal, North West, Gauteng and Limpopo did possibly better declaration on age than females of their provinces. The smallest index for males is 97.37 (Limpopo) and 95.53 (Limpopo) for females. The highest Whipple's index for males is 110.31 (Free State) and for females is 111.42 (Northern Cape). The best Whipple's index for males in LFS 2007 is 101.76 (North West) and for females 102.27 (Free State).

4.4.1.4 Comparison between LFS 2005 and LFS 2007 at Provincial Level

Table 4.1 presents the results of the Western Cape, using LFS 2007. The respondents probably declared properly their age and the age distribution was also highly accurate. But the indexes increased with 2 points instead of decreasing to 100 or keeping the same values. This aspect could reflect the negligence of enumerators to search the exact response or the volume of work had been intensified and they have relaxed in their duties.

Taking into account Eastern Cape, the index of males and females in LFS 2005 failed in the same interval (110-125) indicating that the male and female respondents declared their ages with

some preferences, the quality of age distribution is approximate and their declaration was biased. The LFS 2005 in this province experienced fair declaration of age by the respondents. This could be explained by the non-consideration of the importance of the survey by the respondents. However, the LFS 2007 results are very much appreciated. All the mechanisms, procedures and techniques applied could have contributed to improve the quality of the declaration of the respondents. The male's index was 102.25 and the female 102.70. This case suggests that the enumerators or fieldworkers and all personnel involved in the collection, processing, and analyzing were possibly well trained and they applied all the received advices to perform their work.

As far as KwaZulu-Natal is concerned, the Table 4.1 shows that the females' index was 127.39 indicating that the quality of data is rough and index of both sexes were influenced by age preference ending by 0 and 5; but the LFS 2007 presents the best indexes compared to all other Whipple's indexes of this particular province, its value is 108.56 for males and 110.88 for females. The improvement occurred despite the fact that their age distribution is fairly accurate. The trend of the indexes of KwaZulu-Natal reveals some features influencing the declaration of respondents for instance, the problem of education, the good understanding of the purpose of those surveys, the non-ability of some enumerators to attend some areas in particular rural areas. Considering the both instruments, the highest Whipple's index for males is 119.54 (LFS 2005) and using the same instrument, the highest index for females is 127.39 ranged between 125 and 175 showing the bad quality of age distribution. The improvement is progressive in both sexes but females possibly did better than males.

In North West, the males' index of both LFS showed that the quality of data was highly accurate. Males likely did better in the declaration on age than females. North West female's index (111.3) of LFS 2005 indicated that their age distribution was approximate. The quality of female's responses on age is possibly fairly accurate. The respondents were just providing any information to enumerators instead of chasing them out of their homes. However, the LFS 2007 was the best survey with good Whipple's index of males and females (101.76 and 103.15). The quality of responses was highly accurate and the age distribution ascertained. However, the male respondents doubtless did better declaration on age than females.

UNIVERSITY of the
WESTERN CAPE

4.4.1.5 Inter-Instruments Comparison LFS 2005 and GHS 2004

When observing the LFS 2005 and GHS 2004, the female's index range in the same interval (110-124.9) showing that their age distribution was approximate. However the two LFS set out the males' index which range between 105 and 109.9 meaning the age distribution is relatively accurate. Even the preference tends to increase slowly. The contrary is observed with the females' index. It decreased from 110.87 to 103.85; the quality is improved over years for LFS. This could be highlighted by the consciousness of females in their Labour force empowerment participation.

However, if the Whipple's index is less than 100 and smaller than 95, it indicates that the quality of data was highly accurate but, there is a certain age with the digit ending different from 0 and 5 that is slightly preferred among the responses. The LFS 2005 in Limpopo confirmed the attractiveness in the declaration of age by the respondents. The male index (113.79) specified the approximate age distribution and for the female, the age distribution was fairly accurate. The contrary appears in LFS 2007 with the male's Whipple equaling 97.37, indicating the quality of data is highly accurate with slight preference to digit ending different from 0 or 5. A similar case is observed by female respondents in LFS 2007 with 95.53 as value of Whipple's index reveals high quality of data on age hence the deviation from 100 was less than 5.

UNIVERSITY of the
WESTERN CAPE

Table 4.1 Results of Whipple's Index GHS and LFS at National Level

Provinces	GHS 2004			GHS 2007			LFS 2005			LFS 2007		
	Male	Female	Both	Male	Female	Both	Male	Female	Both	Male	Female	Both
Western Cape	98.97	104.37	101.9	103.62	107.71	105.72	101.65	101.03	101.32	104.138	103.31	103.7
Eastern Cape	109.11	107.13	107.98	105.94	105.87	105.9	110.48	113.61	112.28	102.25	102.7	102.51
Northern Cape	103.63	96.55	99.95	103.86	106.12	105.03	103.51	107.2	105.46	109.02	111.424	110.277
Free State	96.58	107.21	102.04	98.5	112.62	106.16	110.01	108.44	111.97	110.31	102.267	105.935
Kwazulu-Natal	112.38	113.783	113.17	116.03	113.05	114.31	119.54	127.39	124.05	108.555	110.88	109.89
North West	98.45	96.9	97.64	95.73	100.89	98.62	101.12	111.3	106.68	101.76	103.15	102.52
Gauteng	107.2	109.28	108.23	113.69	108.03	110.935	109.89	115.37	112.56	102.42	103.87	103.124
Mpumalanga	102.84	112.08	107.77	104.38	104.25	104.3	106.84	110.87	109.03	106.91	103.854	105.248
Limpopo	102.82	96.37	98.92	99.37	106.88	103.95	113.79	108.97	110.9	97.367	95.531	96.267
National	104.54	105.69	105.16	106.86	108.18	107.59	108.73	111.9	110.47	105.05	105.037	105.043

UNIVERSITY of the WESTERN CAPE

4.4.1.6 Inter-Ethnic Variations

Considering the distribution per population groups, the researcher takes into account the disparities in the level of education and historical exposure to statistical operations; Whipple's index emphasizes the evaluation in the quality of declaration of age and sex and underlines that the preferences occurred in particular ages ending with 0 and 5 using different instruments.

4.4.1.6.1 Results Whipple's Index for General Household Surveys per Population Groups

Referring to the Table 4.2, the outcome of male's index for African/Blacks is 101.27 which reveal that the quality of responses on age is very accurate. This index represents the high quality of data showing that the respondents likely declared properly their ages. The age distribution was highly accurate. The highest Whipple's index for males is 111.2 (Indians) and the smallest is 99.59 (Coloureds) which reveals slight avoidance in ages with digit ending 0 and 5. However, the Whipple's index for Coloured males is the best index according to its position; it is more close to 100 the perfect point compared to other population groups. The best and smallest females' index is 104.39 (Coloureds) indicating that the high quality of data on age and the age distribution is also highly accurate. The highest Whipple's index of females is 107.86 (Whites). Coloured, Indian and White females possibly did better declaration on age than males despite the fact that they would have to improve the quality of their responses.

UNIVERSITY of the
WESTERN CAPE

According to both sexes, the best and smallest index is 102.23 (Coloureds), the highest Whipple's index is 108.91 (Whites). African/Blacks and Coloureds scored in the same interval, 100 to 105 indicating the quality of age-sex structure is probably highly accurate. The indexes of Whites and Indians/Asian for both sexes are ranged between 105 and 110 showing preference in the attitudes of respondents which reveals that the quality of responses on ages is relatively accurate.

Definitely, only African/Blacks males did likely better declaration on age than females. This seems to diverge from the expectation; their lower level of education could have affected the quality of declaration on age. In other groups females declared probably better their ages compared to males despite the fact that they would have to improve the quality of their responses.

Seeing the Table 4.2, the males' index for African/Blacks and Whites in GHS 2007 is ranged between 105 and 110 showing that the quality of responses on age is relatively accurate; but males possibly did better declaration on age than females. The index of Coloured males is 103.54; the quality of age distribution is highly accurate indicating that males may declare properly their ages. The highest value of index for males is 117.37 (Indians), its quality is fair.

Using the GHS 2007, the Whipple's index of females African/Blacks and Whites vary between 105 and 110 indicating that the quality of responses on ages is relatively accurate. Then, Coloureds and Indian females obtain respectively 111.27 and 111.7 as value of Whipple's index. These indicators reveal that the quality of responses on age is fairly accurate and the distribution of age structure is approximate.

The index of both sexes applied for GHS 2007 can be grouped into two categories as follow; the first group constitutes African/Blacks, Coloureds and Whites which vary between 105 and 110 indicating that the quality of data on age and sex structure is relatively accurate. The second concerns Indians/Asians with 114.47 as synthetic Whipple's index. The quality of its data on age distribution is fairly accurate.

UNIVERSITY of the
WESTERN CAPE

4.4.1.6.2 Cross-Ethnic Variations between GHS 2004 and GHS 2007

Observing both GHS, the cross ethnic variations appear in different population groups. The increase of Whipple's index according to African/blacks highlights the evidence of preference in the attitudes of both respondents (males and females). This variation of the indexes in 2007 could be explained by the weakness of the enumerators or the approach did not convince the respondents. The issue of low educational status among this population group should be also considered. The change operates in terms to improve the recognition of marriage which is influencing the declaration on age in some couples especially those belong to African/blacks, knowing that is this population group which is affected by the marriage pattern. Considering the lack of the registration system in these areas, the new (marriage) law aims to clarify some features which imply some discrepancies. Some historical, legal and social reasons for the difficulties on the perception of marriage probably affects the assessment of age when using the logical imputation method in case respondents are not able to remember correctly or to report their exact age in the absence for their counterparts. Also, if the data presents a case where a 12

year old is married, it means there is a problem or the information is not consistent which will affect the quality of declaration of age. This problem can arise for misinterpretation of concept especially when the couple is living together as partners, the responses provided to the enumerators are sometimes not correct (section 3.6.2.1, page 51).

Considering the Coloureds population group, the quality of responses on age for males has been improved in such a way that the trend of preference has been changed; instead of slight avoidance to digit ending 0 or 5, Coloured males are attracted to these digit endings. The index increases from 99.59 to 103.54 in GHS 2007 which shows a decline in the quality of responses on age is likely highly accurate. A possible explanation to this situation may be the influence of the culture, or possibly the improvement in the level of education among this population group. Looking at the females' index in GHS 2007, it increases from 104.39 to 111.27 and the preference is most pronounced. The distortion is visible in the age distribution. The reality is that female respondents did not find any interest in the survey, or did not perceive the importance of the survey, or probably, they did not receive enough information surrounding the survey.

Taking into account the population of Indians/Asians, the increase observed males and females index could be seen as a barrier to the whole process in terms of improvement. Storms & Loosveldt (2004) explained that different cultural backgrounds present challenges because of cultural relevance which also affect the response rate. However, a possible explanation to this issue of preference is that the influence of culture on the declaration on age differed from one culture to another. Although, there was traditional believe in some Asian communities as mentioned (Minister of Viet Nam, 2011) in the Monograph in Viet Nam, that for instance, years starting with "Quy" are good and lucky; Quy year means good or lucky year in the lunar calendar year ending with digit three. For example, there are 1973, 1983, 1993, 2003, 2013 and so on. Therefore, parents choose having children in these particular years. In the situation which respondents cannot properly remember their date of birth, those who were born in these years ending with digit three correspond to "Quy" years will tend to have ages ending with the digit 6, 9. According to GHS, the females' index in 2004 is 105.26; the quality of the distribution is relatively accurate. The same index move to 111.7 in 2007, the quality of data is fair instead of changing in positive way, the situation becomes worse. Males' index in GHS 2007 experienced very slight increase of 0.5. It is also possibly to assume that this population group could not have

confidence in the government policies or probably, they are scared to face difficulties in their investments in the future because of increased insecurity and no safety life.

The White population is also affected by this issue of preference in the declaration on age. However, females' index in GHS 2004 and 2007 are ranged in the same interval between 105 and 110 which denote the attractiveness in the quality of response on age. The quality of their distribution is relatively accurate. The contrast is that this population group is well educated and it supposed to master the issue of these surveys better than other population group. McCarty *et al.* (2007) demonstrated that some political issues have a significant role on the informant and they can influence the response rate. The White population tends to diverge from the expectation according to the high educational status which could have influenced positively their declaration on age. The distortion observed in their age distribution also confirmed the fact that in some area, access was impossible to collect data of good quality. However, the males' index had decreased from 110.03 to 108.10, a slight improvement occurred but this population group still have to do more.

Definitely, the African/Blacks males doubtless declared better their ages than females, the contrary occurs with the population group of Indians/Asians. The attitudes of Whites are the same, males and females recorded the same value of index.

UNIVERSITY of the
WESTERN CAPE

4.4.1.6.3 Results Whipple's Index of Labour Force Surveys per Population Groups, Link with the Participation in the Labour Force

The male population group of Coloureds probably highlights good declaration of age by its respondents with 100.37 as Whipple's index. The quality of data structure is highly accurate. This situation could reflect when the instruction given to enumerators concerning the rounding up of age by respondents is applied very well. The population group of Indian and Whites provided indexes ranging between 110 and 125 which reveal preference in ages with digit ending 0 or 5. The quality of its data on age structure is fairly accurate, the distribution is approximate. The African males' group resulted 109.46 as Whipple's index showing preference in attitudes of respondents, the quality of responses on declaration on age is relatively accurate.

According to LFS 2005, every female population group has its particularities. African/Blacks, Whipple's index equals 113.96 revealing the quality of responses on age is fairly accurate and its

data distribution is approximate. The females are attracted to numbers with digit ending 0 or 5. Coloured females registered 102.8 on Whipple's index which is likely the best compared to the whole population groups of South Africa when using the LFS 2005; the quality of its responses on age was highly accurate. As far as Indian females are concerned, its index is equal to 127.02 indicating the attractiveness in numbers with digit ending 0 or 5; its age-sex structure was affected by some distortions, thus its quality of responses on age was bad. Looking at the White females' index, its value represents 106.31 showing the slight preference in numbers with digit ending 0 and 5; the quality of responses on age is relatively accurate.

The index of both sexes LFS 2005 can be grouped into three categories; the first group constitutes African/Blacks and Indians/Asian females which values of Whipple's index range between 110 and 125, the quality of age-sex distribution on age is fairly accurate. The synthetic index of Whites is 109.28 indicating the presence of preference in the attitudes of respondents; hence the quality of responses on age is relatively accurate. The last group is Coloureds with 101.69 as synthetic Whipple's index showing that the quality of its distribution is highly accurate.

Definitely, considering the population groups using LFS 2005, Whipple's index reveals that the tendency of male respondents likely did better declaration on age compared to female respondents in the subgroups of African/blacks, Coloureds and Indians while White females probably declared better their ages compared to males.

Considering the LFS 2007, the Table 4.2 highlighted that the Whipple's index of African/Black, Coloured and Indian/Asian male respondents are ranged in the same interval which is 105 to 110 showing the slight preference in the attitudes of respondents; the quality of age-sex structure is relatively accurate. However, the index of Whites is 97.66 indicating quality of age-sex distribution is likely very accurate, hence it is less than 100 which implies that male respondents are repulsive with numbers digit ending 0 and 5.

Observing the Table 4.2, African/Black and Coloured females recorded the index range between 100 and 105 indicating that the quality of data on age-sex distribution is highly accurate, hence the good quality of responses is possibly observed. According to Whites, the index is 111.36 revealing the preference in the attitude of respondents; the quality of age-sex structure is fairly

accurate. Observing the female's Indians/Asians, its result is 128.81 showing that the quality of responses on age is likely bad; the age-sex structure is possibly approximate.

With the same instrument, the synthetic Whipple's index of African/Blacks and Indians/Asians vary between 105 and 110 indicating the quality of responses on age is relatively accurate; the respondents are attracted to ages with digit ending 0 or 5. As far as Coloureds and Whites population groups are concerned, its synthetic index vary between 100 and 105, meaning the quality of responses on age is probably highly accurate and its age-sex structure is uniformly distributed.

4.4.1.6.4 Cross-Ethnic Variations between LFS 2005 and LFS 2007

The African/Blacks male's index of LFS 2005 and LFS 2007 are ranged in the same interval (105 to 110). Referring to the standards of Whipple's index, all its distributions are relatively accurate. The respondents tend to round their ages up. Despite the fact that the index remains in the same interval with LFS 2007, the slight improvement is realized for males with the small decrease in the index [(109.46 (LFS 2007) greater than 108.72 (LFS 2005)]. The female's index of LFS 2005 (113.96) reveals that their age distribution is approximate. Using LFS 2007, female' index is 104.598 implying the quality of age distribution is very accurate. There is a great change in their attitudes which can be seen in the decrease of their index. Possibly, the campaign surrounding the operations of collection of data, the process and field work was all achieved.

The population group of Coloureds highlights possibly the good declaration of age by respondents during the LFS 2005; all indexes indicate that the quality of data is probably highly accurate. The males' index was 100.37 and females' 102.80. The female confirmed the same Whipple's index which the value was 102.80 in LFS 2007, but the male increased their value to 105.29 which show that the quality of data is relatively accurate. In regard to population groups, using LFS 2007, Coloured females' index is doubtless the best 102.81; the smaller Whipple's index compared to the Whites males and the Indian females has the highest of 128. 81. The Indian females did probably bad declaration of age and their age distribution was distorted indicating the more preference to digit endings 0 and 5 was observed.

The index of White males positively decreased from 112.41 (LFS 2005) to 97.66 (LFS 2007) indicating a great improvement in the quality of its data which is becoming very accurate. However, the index of LFS 2007 reveals a slight avoidance in ages with digit ending 0 or 5. The White females' responses reveal preferences to digit ending 0 or 5. A possibly explanation is that the Labour Force participation of this subgroup is much representative than others. Most of the investments and the economy of the country are in their control. The contrary situation is observed by white females with the increase of their index which change from 106.31 to 111.36. However, the result of white females is far different from the expectation knowing their position in the Labour force participation and with their high educational status. the Coloured population group has the best result with the both index equal to 103.94 in LFS 2007, followed by African/Blacks which present 108.72 for males and 104.60 for females.

Observing the Table 4.2, Indian/Asian population group, present the attractiveness of the respondents in particular ages with digit endings 0 or 5 according to the three indexes of males obtained from LFS 2005, GHS 2004 and GHS 2007. The quality of data is fair and their indexes are ranged between 110 and 125. The distortion is established in their age distribution. The females' index of LFS 2005 and 2007 follow the same trend, but the quality of their data is possibly bad with their Whipple's index ranged between 125 and 175. The expression of the female's preference was more (well) pronounced in ages with digit endings 0 and 5.

UNIVERSITY of the
WESTERN CAPE

The decrease in the males' index according to the LFS revealed that the great improvement is carried on and its effects could be appreciated through the attitude of males in terms of preference in ages with digit endings 0 and 5. The quality of males' data moved from fairly to relatively accurate.

As far as White population group is concerned, males' index increased over years. These are the results of GHS 2004 and LFS 2005 (110.03, 112.41). The preference is increasing. The quality of data is fairly accurate. Instead of reducing the gap with all the strategic built in terms of assessing the quality of data, the indexes reveal that their age distribution was abnormal. Even the LFS 2007, the males' index is 97.66 which highlight avoidance in ages with digit ending 0 or 5. The attitude of this population group is probably emphasized by the political issues linked with apartheid.

In few words, Table 4.1 shows overall variations of the Whipple's index at national and provincial level in South Africa. This implies an improvement in the quality of age data. The Table 4.2 further shows overall variation of the Whipple's index among population groups in South Africa. Although higher index for females than males was observed among African/Blacks and Coloureds using GHS 2004 and GHS 2007; among Indian/Asians and Whites using LFS 2007, also among African/Blacks and Indian/Asians when applying LFS 2005. In the same vein, higher index for males compared to females was observed among Indians/Asians and Whites using GHS 2004 and GHS 2007; among Coloureds and Whites using LFS 2005; also among African/Blacks and Coloureds using LFS 2007. Digit preference is most pronounced among subgroup having low education status; low level of Labour force participation and the marital status also influence the declaration on age of respondents. It points out that the pattern of digit preference varies from one culture to another.

Table 4.2 Results Whipple's index per population groups GHS and LFS

Population groups	GHS 2004			GHS 2007			LFS 2005			LFS 2007		
	Male	Female	Both	Male	Female	Both	Male	Female	Both	Male	Female	Both
African/Blacks	101.27	105.67	103.67	106.91	107.53	107.25	109.46	113.96	111.96	108.717	104.599	106.378
Coloureds	99.59	104.39	102.23	103.54	111.27	107.67	100.37	102.8	101.69	105.288	102.807	103.94
Indian/ Asians	111.2	105.26	108.14	117.37	111.7	114.47	121.64	127.02	124.41	109.735	128.814	119.481
Whites	110.03	107.86	108.91	108.1	108.18	108.14	112.41	106.31	109.28	97.661	111.355	104.73
National	104.54	105.69	105.16	106.86	108.18	107.59	108.73	111.9	110.47	105.05	105.037	105.043

UNIVERSITY of the
WESTERN CAPE

4.4.1.7 Results Whipple's Index for Community Survey 2007 and Census 2001

Overview

The 2001 Census was the second all-inclusive operation of data collection that Statistics S.A. conducted under the new democratic government. The government made a decision to move away from the 5-year to 10-year Censuses. This decision created a gap in data between 2001 and the following Census which was scheduled for 2011 (Stats, 2007). To ensure consistency, a decision was made to carry out a Community Survey in 2007. However, the objectives of the 2007 community survey and the census 2001 were different in terms of geographical scope. Given the difference in terms of target and coverage, it may not be possible to compare the two results even if the variables age and sex remain the same. Thus, the Community survey was representative at provincial, metropolitan, municipal and district levels, therefore, the results can be extrapolated to the entire country for making the inferences.

4.4.1.7.1 Results Whipple's Index for Census at National Level

The Table 4.3 shows also that the results of Whipple's index at national level for Census 2001. The male's index was equal to 96.92, for females 97.23 and index of both sexes or synthetic index was 97.09. They are all indicating that the quality of each data is probably very accurate hence the deviation from 100 which is the perfect is less than 5 per cent. However, females did better declaration of responses on age compared to males. Referring to the standard of Whipple's index, the responses on age in the population census in South Africa can be regarded as very accurate. The fact that the value of each of these indexes are less than 100 indicate that there was very slight avoidance in ages with digit ending different from 0 or 5.

UNIVERSITY of the
WESTERN CAPE

4.4.1.7.2 Results Whipple's Index for Community Survey at National Level

Table 4.3 indicates similarly the results of Whipple's index from Community Survey 2007. The male respondents registered 96.81 against 95.49 for females and the index for both sexes was 96.12. These indexes are all belonging in the same range with its deviation less than 5 per cent. They are showing that the quality of each age distribution is very accurate in regard to the standards of the Whipple's index. However, there is slight avoidance occurred in ages with digit ending 0 or 5, in other words there is a certain age which digit ending is different from 0 or 5 is

preferred among males and females. Although, males likely did better declaration than females compared to the responses.

4.4.1.7.3 Comparison Results of Whipple's Index between Community Survey 2007 and Census 2001 at National Level

The expectation was to evaluate if there is an improvement at national level over years and across gender between Census 2001 and Community Survey 2007.

Table 4.3 shows that at national level, both instruments encountered best quality of the Whipple's index. According to census, males' index is higher than for females showing that age was more accurately reported among females than males. The deviation to 100 for males is higher than for females. The contrary is observed in Community survey with females' index higher than for males revealing that males did better declaration of response on age than females. The quality of responses of all these indexes at national level was also very accurate despite the fact that the increase has been observed in the avoidance. The overall results of census were better compared to Community survey. However, the index has been rising with Community survey in 2007, indicating deterioration in the quality of age data. There is no improvement encountered. A possible explanation is that these two instruments are totally different, Census operation covered the entire country while community survey just sampling and apply the inference at the end. Census probably captures better information than survey.

UNIVERSITY of the
WESTERN CAPE

4.4.1.8 Results Whipple's Index for Community Survey 2007 and Census 2001 at Provincial Level

4.4.1.8.1 Results Whipple's Index for Community Survey 2007 at Provincial Level

The Table 4.3 displays overall results of Whipple's index using the Community survey even at provincial level indicates that the slight heaping in numbers with digit ending different from 0 or 5. However, the following provinces Western Cape, Eastern Cape, Northern Cape, KwaZulu-Natal, Gauteng and Mpumalanga have encountered good results of its indexes for males, females and synthetic Whipple's index. The deviation to 100 is less than 5 per cent for each index indicating that the quality of responses on declaration of age is very accurate.

As far as Limpopo, North West and Free State are concerned, especially female respondents recorded higher index than males in all the three provinces, indicating that age was more accurately reported among males than females. However, the deviation to 100 for females was higher than for males showing that the avoidance was more pronounced among females than males across the three provinces. In general, the patterns of digit preference vary from one culture to another and the digit preference is most pronounced among population having low educational status. For instance, in Western Cape, females' index is 97.54, the deviation to 100 is less than 5; the quality of this data referring to Whipple's method is very accurate while in Limpopo, the females' index is 92.20, the deviation to 100 is more than 5 showing the quality of responses is relatively accurate. Both indexes indicate avoidance in numbers with digit ending different from 0 or 5, but the avoidance is more pronounced in Limpopo than in Western Cape among females.

Observing results of some provinces such as Limpopo, North West and KwaZulu-Natal where population has in general low educational status and where some influences of culture are more pronounced, the value of its index shows that the quality of data is relatively accurate. However, males did better declaration than females compared to all provinces, only in two provinces (Northern Cape and KwaZulu-Natal) where females provide better responses on age than males.

UNIVERSITY of the
WESTERN CAPE

4.4.1.8.2 Results of Whipple's Index for Census 2001 at Provincial Level

Observing the Table 4.3, the avoidance is pronounced in both sexes especially in eight provinces except Western Cape. However, male respondents did better declaration on age than their female counterparts. The quality of the data is very accurate. Mpumalanga and Limpopo results indicate relatively accurate quality of data on age. Their deviation from 100 is greater than 5 (range between 5 and 9.99). Western Cape, Free State and Limpopo present similar comments. Males reported better their ages compared to females in those provinces, but Western Cape and Free State were very accurate while Limpopo was relatively accurate. In the remaining provinces females did better declaration on age compared to males. A possible reason to this gap is that Limpopo and Mpumalanga are provinces with predominance of rural areas. In those particular provinces the level of education is low and cultural factors influence females' attitude in the society. However, Western Cape registered the smallest Whipple's index (100.53) for males and

Limpopo (93.06) realized the highest for males while Gauteng (99.69) presented the smallest value of Whipple's index for females and Limpopo (92.13) indicated the highest Whipple's index.

4.4.1.8.3 Comparison Results of Whipple's Index Census 2001 and Community Survey 2007

Regarding both results, male respondents of Western Cape, Free State and Limpopo did better declaration on age compared to females. As far as Community survey was concerned, Eastern Cape, North West, Gauteng and Mpumalanga presented the same trend which was contrary when applying Census 2001. However, the improvement is observed in Eastern Cape, Free State, North West, Mpumalanga and Limpopo among males while in Western Cape, Northern Cape, Mpumalanga and Limpopo, the decrease of the deviation from 100 which is the perfect point indicated amelioration in the quality of data. A possible reason could be the revision of questionnaire over years for instance, from 1996 to 2007, the question on age has been refined and the two important questions merging into one try to capture the exact age of the respondent: "What is (the person's) date of birth" and age in completed years.

Table 4.3 Results Whipple's index Census 2001 and Community survey 2007

	C S 2007			Census 2001		
Provinces	Male	Female	Both	Male	Female	Both
Western Cape	98.31	97.54	97.92	100.53	102.49	101.57
Eastern Cape	96.59	95.4	95.92	95.77	96.75	96.31
Northern Cape	96.07	99.68	97.86	98.57	98.25	98.40
Free State	98.09	94.44	96.26	95.69	95.35	95.51
Kwazulu-Natal	95.63	96.22	95.95	98.83	98.72	98.77
North West	96.26	92.34	94.34	95.51	95.75	95.64
Gauteng	97.3	95.43	96.38	98.95	99.61	99.30
Mpumalanga	97.22	95.48	96.32	94.32	94.74	94.55
Limpopo	95.61	92.2	93.66	93.06	92.13	92.55
National	96.81	95.49	96.12	96.92	97.23	97.09

UNIVERSITY of the WESTERN CAPE

4.4.1.9 Results Whipple's Index for Census 2001 and Community Survey 2007 per Ethnic Groups

4.4.1.9.1 Results Whipple's Index for Census 2001 per Ethnic Groups

Table 4.4 presents an overall best quality of the Whipple's index. This implies an improvement in the quality of the data. Females had recorded a higher index than males among all the four population groups. The age was accurately reported among males than females. Specifically in order, this trend is observed by Whites; then, Coloureds; followed by Indians and Africans/Blacks. However, Whites did better declaration on age compared to other population groups. The highest index is the one which the deviation to 100 is greatest than others. According to these population groups, Africans/Blacks record the highest index for males (95.64), for females (95.58) and the synthetic (95.61). The White population group obtains the smallest index for males (100.27), for females (102.44) and the synthetic (101.59) compared to other groups.

4.4.1.9.2 Results of Whipple's Index for Community Survey 2007 per Ethnic Groups

UNIVERSITY of the
WESTERN CAPE

Considering the Table 4.4 of results per ethnic groups, both sexes indicated very accurate data quality except black females which Whipple's index range between 90 and 95 revealing that the quality of its data was relatively accurate. The female respondents reported their age ending with a certain digit different from 0 or 5. However, male respondents of Black and Indian population did better declaration on age compared to females while the reporting on age was better among female respondents of Coloured and White population group compared to their counterparts. Knowing that Black population group is majority in rural areas, they were exposed to many patterns. Education remains a great concern among the Black population which is facing different problems such as poverty, illiteracy among females and cultural factors. The smallest Whipple's index for males is 100.64 (White) while the highest is 95.76 (Black). In addition, Black females registered the highest Whipple's index (94.13) while Coloured females obtained the smallest Whipple's index (100.05).

4.4.1.9.3 Comparison Results of Whipple's Index Census 2001 and Community Survey 2007 per Ethnic Groups

Referring to both results, Coloured population group improved the quality of their responses over years. In the same vein, Indian and White females observed a decrease of their Whipple's index showing that the quality of its distribution was accurate. The reporting of age among these females was probably accurate. Whipple's index of Indian and White male respondents was modified indicating deterioration of the quality despites the fact that the quality of reporting of age remained probably very accurate. However, Black population registered the highest Whipple's index over years for both sexes; White males highlighted the smallest Whipple's index over years, hence Coloured females realized a decrease of their Whipple's index showing progressive improvement of the quality of its data on age.

Table 4.4 Results Whipple's Index of Census 2001 per Ethnic group

Population groups	Census 2001			Community Survey 2007		
	Male	Female	Both	Male	Female	Both
African/Blacks	95.64	95.58	95.61	95.76	94.13	94.90
Coloureds	101.18	102.62	101.95	98.93	100.05	99.50
Indian/ Asians	101.4	103	102.23	98.35	97.13	97.77
Whites	100.27	102.44	101.59	100.64	99.71	100.18
National	96.92	97.23	97.09	96.81	95.49	96.12

UNIVERSITY of the WESTERN CAPE

4.4.2 Results of Myers Index

The evaluation of data quality on age using techniques of Myer's index and age specific index can show whether there is any other digit ending preferred by respondents in the population census or surveys data. The value of Myer's blended index can range from zero to 90. This value is higher; the more preference there is for certato digits (age heaping). The age specific index is equal to $100*(T_u/T)$. The value of age specific index is ranged from 0 to 9. If the value is 0, there

is neither preference nor avoidance of any age ending with these numbers. If any value is higher than 0, positive sign, the preference exists at the age ending with that number. If the value is less than 0, negative sign, there is an avoidance of age response at that particular number.

4.4.2.1 Results of Myer's Index for General Household Surveys at National Level

4.4.2.1.1 Results Myer's Index GHS 2004 and GHS 2007 at National Level

The application of Myers techniques to evaluate the quality of data in this study brought the following results at the national level, for GHS 2004; the both indexes are closed (6.78 for males and 6.96 for females). The two indexes are closer to 0 than 90. They are very low. However, Myers also detects the digit preference whether is attractive or repulsive to ages with digit ending starting from 0 to 9. In the case of GHS 2004, the digits 3, 1, 7 and 9 were repulsive for male in one hand, and, the digits 5, 8, 6, 0, 4 and 2 were attractive for male. Accordingly, females, 7, 1, 3, 6, 5 and 9 were repulsive and 2, 8, 4 and 0 attractive. Definitely, males and females were possibly repulsive with numbers ending with 1, 3, 7 and 9 for the GHS 2004; and the set of them were attracted for the numbers ending with 0, 2, 4 and 8.

UNIVERSITY of the
WESTERN CAPE

Observing the table of Myers using GHS 2007 at national level, the index of male is 4.18 and females is3.68. They are very low. The declaration of age for females was better than for males. The column $100*(T_u/T)$ indicates the preferences. Males were attracted to numbers ending with digit 2, 5, 0 and 7 but repulsive to digits 1, 4, 3, 9, 6 and 8. Females were repulsive for ages ending with 1, 6, 8, 9 and 4 and attracted by 3, 7, 2, 0 and 5. However, both sexes were attracted to 0, 2, 5 and 7; repulsive to 1, 4, 8and 9.

Table 4.11 Distribution of digit preferences at national level for G H S per gender

	GHS 2004		GHS 2007	
Sex	Repulsive	Attractive	repulsive	Attractive
Male	3 1 7 9	5 8 6 0 4 2	1 4 3 9 6 8	2 5 0 7
Female	7 1 3 6 5 9	2 8 4 0	1 6 8 9 4	3 7 2 0 5

4.4.2.1.2 Comparison Results Myer's Index GHS 2004 and GHS 2007 at National Level

The results of GHS 2007 compared to GHS 2004 show a substantial decrease indicating a progressive improvement for both sexes but the males' index was more close to 0 than females showing that males in 2007 probably did better in the declaration of age than females. The difference in index for males was 4.83, for females 2.21; the simple comparison set out the interest of male respondents to improve the quality of their age-sex structure. The results of GHS 2007 revealed that many corrections had taken place in the whole operations of the survey which explained the improvement occurred; even in the preference of respondents. As similarities in the attitude, the respondents were attracted to ages ending with digit 0 and 2 in the two GHS, they were avoided ages digit ending with 1 and 9 in the two surveys.

4.4.2.2 Results of Myer's Index for Labour Force Surveys at National Level

4.4.2.2.1 Results Myer's Index LFS 2005 and LFS 2007 at National Level

The Table 4.5 indicates outcomes obtained in LFS 2005 Myer's index 4.50 for males and 5.49 for females. The indexes were low and very close to 0 than 90. The female respondents possibly did good declaration on age compared to males. The $100T_u/T$ showed that males were attracted to digits 3, 2, 4, 0 and 5; repulsive to digits1, 7, 6, 9 and 8. The females were repulsive to digit 1, 7, 6, 3, 8 and 4; attractive to digits 9, 2, 5 and 0. Hence, both sexes were attracted to 0, 2 and 5; avoided by 1, 6, 7 and 8.

Considering the table of Myers LFS 2007 at national level, the males Myers index is 3.18 and the females 3.87; the two are very low compared to 90. The declaration of age of males was probably better than for females. Males were attracted to ages ending with digits 8, 4, 7, 5, 6, 0 and 2; they were repulsive with numbers ending with 1, 3 and 9. The females were attracted to ages ending with digits 7, 4, 5, 2 and 0; repulsive with 1, 9, 6, 8 and 3. Both sexes were attracted to 0, 2, 4, 5 and 7; repulsive to 1, 3 and 9.

The results of LFS 2007 are likely lower than those of LFS 2005 (3.18< 4.50 and 3.87< 5.49) showing the great change. The decrease in index for males is 1.32, for females is 1.70 showing

that females did better improvement than males. Despites the fact that the improvement is probably most pronounced in female respondents than males, males likely did better declaration on age than females (4.50 < 5.57; 3.18< 3.87). Taking into account the preferences in both surveys, it revealed that the respondents were avoided only in numbers with digit ending 1 and the preference were pronounced in ages with terminal digit 0, 2 and 5.

Table 4.12 Distribution of digit preferences at national level for Labour Force Surveys per gender

	LFS 2005		LFS 2007	
Sex	Repulsive	Attractive	repulsive	Attractive
Male	1 7 6 9 8	3 2 4 0 5	1 3 9	8 4 7 5 6 0 2
Female	1 7 6 3 8 4	9 2 5 0	1 9 6 8 3	7 4 5 2 0

4.4.2.2.2 Comparison Results Myer's Index for GHS and LFS at National Level

Definitely, the results of Myer's index at national level are likely good, there are close to 0. The detection of preference set out some inconsistencies which were not significant. The Tables of distribution of Myer's index using GHS 2004, LFS 2005 GHS 2007 and LFS 2007 at national level indicate that the age that had the highest frequency were ages ending in 4 and 5. The least frequency ages were those ending in 7 and 1. An explanation to this failure is linked with the question asked for the year of birth many respondents tend to report being born in the years with terminal digit 0 as for instance, 1940, 1950, 1960, 1970 and so on; generally this occurs with the respondents who do not precisely remember the date of birth and even the year they were born close to the year ending in 9 or 1.

UNIVERSITY of the WESTERN CAPE

Table 4.5 Results Myer's Index GHS and LFS at National and Provincial Level

Provinces	GHS 2004		GHS 2007		LFS 2005		LFS 2007	
	Male	Female	Male	female	Male	Female	Male	Female
Western Cape	5.08	4.76	3.51	6.22	7.39	4.76	5.68	5.97
Eastern Cape	5.75	8.45	5.34	5.38	6.12	6.53	5.14	6.78
Northern Cape	7.82	8.51	6.47	5.72	4.29	6.99	6.94	6.43
Free State	8.18	7.59	4.58	4.55	4.76	6.51	6.08	5.76
KwaZulu-Natal	9.31	9.99	4.56	5.28	5.62	7.39	3.92	6.86
North West	5.72	8.06	5.38	4.42	6.00	6.59	4.17	4.54
Gauteng	7.96	8.73	8.89	5.26	7.29	7.91	4.21	4.98
Mpumalanga	4.66	8.62	8.51	6.42	4.86	5.79	5.94	5.10
Limpopo	6.27	8.02	5.28	4.49	5.66	4.37	5.93	5.30
National	8.51	6.39	3.68	4.18	4.50	5.57	3.18	3.87

4.4.2.3 Results of Myer's Index for Census 2001 and Community Survey 2007

Overview

Despite the difference in research designs and coverage, the Community Survey was found to be comparable in many aspects with other Statistics Surveys, Censuses and other external sources. The population is an approximation to 2001 numbers and it is not new data. The question asked relative to the age and sex are the same. The reader should understand that the figures are within a certain interval of confidence (Stats, 2007).

UNIVERSITY of the
WESTERN CAPE

4.4.2.3.1 Results of Myer's Index for Census 2001 at National Level

Statistics in Table 4.6 indicate that the values of Myer's index in the population Census in South Africa is very low and has improved. Myer's index registered the same value in the population, only 2.77 at national level. A high Myer's Index implies poor age reporting whereas a low Myer's Index indicates good age reporting. The maximum value of Myer's Index is 90 and the minimum value is 0. In this case of South Africa, the index is on the lower side (less than 10), which implies that the quality of age reporting is very accurate. Myer's confirms that the age heaping at ages ending with digit 0 or 5 in the population census of South Africa is insignificant. The $100*(T_u/T)$ shows that both males and females were attracted to digits 1, 9 and 8; repulsive

to digits 0, 5 and 3. Especially at national level, males and females did well reporting of age, both sexes scored the same value. The terminal digit of the age with the highest frequency was 1 and the least 4.

4.4.2.3.2 Results of Myer's Index for Community Survey at National Level

Table 4.6 highlights the results of Myer's index using Community Survey 2.08 for males and for females 2.66 showing that males did better reporting on age than females. Community survey using Myer's index presents very good results at national level and referring to the whole set. Males presented a higher index than females implying better age reporting among females than males. However, females were attracted to digit ending 4, 6 and 7 while avoided to digit 5, 0, 9 and 3. Males were fascinated in terminal digit 7, 6 and 4 while repulsive to digit 5, 9, 3 and 0. The preference for these digits among males may be attributed to the greater tendency to overestimate the age, whilst for females; it might be due to underestimation of their age.

Table 4.13 Distribution of digit preferences at national level for 2001 Census and Community Survey 2007 per gender

UNIVERSITY of the
WESTERN CAPE

Sex	Community 2007		Census 2001	
	Repulsive	Attractive	repulsive	Attractive
Male	5 9 3 0 8	7 6 4 1 2	4 0 2 3 5 7	1 9 8 6
Female	5 0 9 3 1	4 6 7 2 8	4 0 2 5 7 3 6	1 9 8

4.4.2.3.2 Comparison results Myer's Index between Census 2001 and Community Survey 2007 at National Level

It was assumed that there is an improvement at national level over years, across gender and between Census and Community survey. The result of Community survey 2007 compared to Census 2001 revealed a slight decrease of indexes (males and females). This decrease of indexes is showing that the preference was not having any significance plausible in the context of this population. The improvement was progressive in both sexes. The similarities observed are that males and females were attracted and avoided in numbers with the same digit ending in census 2001; is not strange because they have obtained the same result, but with Community survey, the

same scenario occurred with different results of Myer's index. However, males and females in both Census and Community survey were repulsive to ages ending with 0, 3 and 5; despite this slight avoidance, the quality of reporting age was very accurate in the two instruments.

Table 4.6 Results Myer's Index Census 2001 and Community 2001

	C S 2007		Census 2001	
Provinces	Male	Female	Male	Female
Western Cape	1.77	1.91	2.40	2.00
Eastern Cape	3.58	4.75	4.12	4.43
Northern Cape	3.58	2.28	2.75	2.55
Free State	2.49	2.89	2.76	3.48
KwaZulu-Natal	2.6	2.37	2.81	2.71
North West	2.37	3.68	3.53	3.21
Gauteng	1.92	2.3	2.16	1.99
Mpumalanga	2.29	3.39	3.67	3.22
Limpopo	2.83	3.69	3.57	4.45
National	2.08	2.66	2.77	2.77

UNIVERSITY of the
WESTERN CAPE

4.4.2.4 Variations of the Myer's Index across the Ethnic Groups

Using the repartition per population groups, Myer's index emphasizes the evaluation in the quality of declaration on age and sex and underlined the preferences occurred in particular ages with digit ending 0 and 5 using different instruments.

4.4.2.4.1 Results of Myer's Index for General Household Surveys per Population Groups

4.4.2.4.1.1 Results Myer's Index GHS 2004

The calculation of Myer's index using GHS 2004 per population group provides the following results. The Myer's index for African/Blacks according to males is 5.70, and females 9.29. Males are repulsive to numbers ending with 3, 1 and 7; and attractive to those ending with 9, 6, 5, 8, 0, 4 and 2. In contrary females are repulsive to numbers ending with 7, 1, 3, 5, 6, 8 and 9; and attractive to those ending with digit 2, 0 and 4. Both sexes were attracted to age ending with 0, 2, and 4 and avoided to ages ending with 1, 3 and 7. The highest frequency for age specific indicator in the case of African/Blacks is age ending with 4 and the lowest frequency is age with

digit ending 3; followed by age ending with 7. The males possibly declared better their ages than the females and the both Myers indexes are close to 0 than 180. The quality of data was doubtless accurate.

As far as Coloureds group is concerned, the index for males was 4.17, and for females 6.07. Males likely did better declaration of age than females. Males were attracted to 9, 8, 3, 6 and 2; repulsive to 7, 1, 4, 5 and 0. Females were repulsive to 7,6,1,8 and 9; attractive to 5, 2, 4, 0 and 3.Both were avoided by 1 and 7; attracted to 2 and 3. Based on Myer's results, the quality of age distribution of Coloureds group was probably accurate. The highest frequency was age ending with 3 and the smallest frequency was age ending with 7.

Regarding the Indian/Asian population group, the index of Myers using GHS 2004 for males was equal to 8.09; and the index for females was 9.50. Obviously, males were repulsive to numbers ending with 1, 3, 7, 6, 5 and 8; attracted to 9, 4, 2 and 0. Females were attracted to numbers ending with 6, 2, 9, 8, 7 and 0; avoided numbers ending with 1, 5, 4 and 3. The set of them were repulsive to 1, 3 and 5; attracted to 0, 2 and 9. The terminal digit of the age with highest frequency was 0 and the least was 1. The declaration of males was probably better than the females.

UNIVERSITY of the
WESTERN CAPE

Looking at the Table of White population group, the index of Myers using GHS 2004 for females is equal to 7.38; for males 17.49. Males were attracted to 5, 7, 2, 0 and 6; avoided numbers ending with 9, 1, 8, 3 and 4. The female avoided massively numbers ending with 7, 1, 6, 4, 3, 9, 2 and 5; they were also attracted only to numbers ending with 8 and 0. However, both sexes avoided numbers ending with 1, 3, 4 and 9; attracted only to 0. Both indexes were low, thus, the quality of data was accurate. Female respondents possibly did better declaration than male respondents.

4.4.2.4.1.2 Cross-Ethnic Variations Myer's Index for GHS 2004

Considering the GHS 2004 when comparing the four population groups, Coloureds doubtless provided the best results; the smallest Myer's index for males is 4.17 (Coloureds) and the smallest Myer's index for females is 6.07 (Coloureds). However, the index of males (Coloureds) is likely the best compared to all indexes of population groups and at national level using GHS

2004. The highest value of male's index is 17.49 (Whites) and for females is (9.50) Indians/Asians. According to GHS 2004, males were attracted to ages with digit ending 2 and repulsive to digit ending with 1. Hence, females were attracted to ages with terminal digit 0 and avoided digit ending with 1. The quality of data on age was likely very accurate.

4.4.2.4.1.3 Results Myer's Index GHS 2007 Per Ethnic Groups

The application of GHS 2007 per population group using the techniques of Myer's provides the following results and highlights the improvement done when comparing the results of the GHS 2004 and GHS 2007.

Looking at the table of Myer's index of the African/Blacks the GHS 2007 indicates a value for males is 3.66 and that for females is 3.93. Together the quality of its data was likely accurate. The results were showing very slight preference for both sexes. Unfortunately males and females were attracted to the same numbers ending with 0, 2, 5 and 7; repulsive also to the same ages with terminal digit 1, 6, 3, 4, 9 and 8. The terminal digit for age specific indicator with highest frequency according to this population group is 0 and the lowest was 1. However, the male respondents probably did better declaration on age than females.

UNIVERSITY of the
WESTERN CAPE

Referring to the table of Myer's index of Coloureds, the males' index is 2.77, and 6.26 for females. The declaration of age for males was in all probabilities appreciated and compared to the females was likely the best. However the males were attracted to numbers ending with 8, 4, 6, 7, 2, 5, 3 and 0; repulsive to those ending with 1 and 9. The females were few for numbers ending with 1, 9, 8 and 4; markedly represented for those ending with 7, 6, 3, 2, 0 and 5. Both sexes were impressed to numbers ending with 0, 2, 3, 5, 6 and 7; repulsive with numbers ending with digit 1 and 9. The highest frequency for the age specific indicator was 5 and the lowest was 1. The quality of male's age and sex distribution is doubtless accurate and for females was likely accurate.

Looking at the Indian/Asian group, the Myer's index of males is 8.33; and 10.70 for females. Within, the Indian population group, males were likely to declare their age better than females. The males avoided the numbers ending with 7, 4, 9, 8, 3 and 1; attracted to those ending with 6, 2, 0 and 5. The females were attracted to digits 9, 3, 5, 2 and 0; repulsive to digits 1, 4, 6, 8 and

7. Both sexes avoided the ages ending with 1, 4, 7 and 8; attracted to ages ending with digits 0, 2 and 5. The highest frequency for age specific indicator is 0 and the lowest is 1. The quality of its data on age is probably accurate.

Regarding the table of Myer's index of Whites population group, the index of males is equal to 7.27, and then female's index is 5.48. These indexes are closed to than 100; but when looking at the preferences there were some particularities which occurred. The males were avoided to ages ending with 4, 3, 9, 1, 8, 5 and 2; attracted to 6, 7 and 0. The females were avoided to ages digit endings with 9, 1 and 4; attracted to those ending with 7, 3, 2, 8, 5, 6 and 0. The contrast was that both sexes were repulsive to numbers ending with 1, 4 and 9; attracted to 0, 6 and 7. The highest frequency for the age specific indicator is 0 and the lowest is 9. The females possibly did better declaration of age than males. The two indexes are low which mean the quality of data is likely accurate.

Table 4.14 Distribution of digit preferences per population groups for General Household Surveys

		GHS 2004		GHS 2007	
		Repulsive	Attractive	repulsive	Attractive
Blacks	Male	3 1 7	9 6 5 8 0 4 2	1 6 3 4 9 8	0 2 5 7
	Female	7 1 3 5 6 8 9	2 0 4	1 6 3 4 9 8	0 2 5 7
Coloured	Male	7 1 4 5 0	9 8 3 6 2	1 9	8 4 6 7 2 5 3 0
	Female	7 6 1 8 9	5 2 4 0 3	1 9 8 4	7 6 3 2 0 5
Indians	Male	1 3 7 6 5 8	9 4 2 0	7 4 9 8 3 1	6 2 0 5
	Female	1 5 4 3	6 2 9 8 7 0	1 4 6 8 7	9 3 5 2 0
Whites	Male	9 1 8 3 4	5 7 2 0 6	4 3 9 1 8 5 2	6 7 0
	Female	7 1 6 4 3 9 2 5	8 0	9 1 4	7 3 2 8 5 6 0

UNIVERSITY of the WESTERN CAPE

4.4.2.4.1.4 Cross-Ethnic Variations Myer's Index GHS 2007

The application of Myer's index method to the GHS 2007 allows checking the cross-ethnic variations among the population groups in South Africa. Considering the results of GHS 2007, it reveals that the African/Blacks indexes are close to national indexes and the gap between national index and index of males was 0.02 (3.68-3.66 =0.02), and with females was 0.25 (4.18-3.93). The best male's index compared to all indexes of population groups and at national level

was probably 2.77 (Coloureds) and for the females was possibly 3.93 (African/Blacks). The highest value of male's index was 8.33 (Indians/Asians), for females was 10.70 (Indians/Asians). In general, males were attracted to ages with digit ending only 0 and avoided to numbers with terminal digit 1 and 9. Hence, females were attracted to ages ending with 0 and 5, repulsive to numbers with digit ending 1 and 4. Observing the result of GHS 2007, it indicates that males probably declared properly their age compared to females in the population groups of African/Blacks, Coloureds and Indians. Indeed, White females likely did better declaration on age than males.

4.4.2.4.1.5 Inter-Ethnic Variations Myer's Index between GHS 2004 and GHS 2007

Considering the GHS 2004 and GHS 2007, the results obtained highlight different interpretations. At provincial level, males were likely declared better the age than females. This may be confirmed by the majority groups as follow African/Blacks, Coloureds and Indians despite the fact that Indian population group tend to increase the preference in their declaration on age over years. The decrease observed in the Whites population did not change the direction of its attitudes; White females were likely declared better the ages than males. The similarities in preference is occurred with females and males in GHS 2007 were attracted to ages digit ending with 0 and both avoided to numbers with terminal digit 1. According to GHS 2004, both sexes did not have any similarity in preference but they do have the same avoidance to ages with digit ending 1.

UNIVERSITY of the
WESTERN CAPE

When comparing the results of GHS 2004 and GHS 2007, the population of South Africa could be divided into two sub-groups, one side constitutes African/Blacks and Coloureds and another side, Whites and Indians. Hence the index of first group is decreasing which mean the improvement is carried on, the second group indexes are increasing which show that the phenomena of age heaping is not yet achieved in this second group.

4.4.2.4.2 Results of Myer's Index for Labour Force Surveys per Population Groups

4.4.2.4.2.1 Results of Myer's Index LFS 2005 per Population Groups

The application of Myer's index using the LFS 2005 enhances to assess the accuracy of data involved. According to African/Blacks population group, the Myer's index of males is equal to 10.50, and for females is 5.68. They are low which mean the quality of data was probably accurate. Males avoided numbers ending with 1, 6, 9 and 7; were attracted to 3, 2, 8, 4, 0 and 5. The females were attracted to numbers ending with 2, 5 and 0; repulsive to 1, 7, 6, 3, 8, 9 and 4. Both sexes were attracted to 0, 2 and 5; repulsive to 1, 6, 7 and 9.The female's respondents possibly did better declaration of age than the males. The terminal digit with the highest frequency for African/Blacks was 0 and the least was 1.

As far as the Coloureds group is concerned when applying the Myer's techniques, the index for males is 5.68, for females 4.64. Based on these results, the females probably did better declaration of age than the males. However, the quality of age distribution of Coloureds was accurate. Males were attracted to 8, 9, 3, 6, 2, 0 and 4; repulsive to 1, 5 and 7. The females were avoided to 1, 3, 8 and 7; attracted to 0, 6, 5, 4, 9 and 2. However, both sexes, were together repulsive to 1 and 7; attracted to 0, 2, 4, 6 and 9. The highest frequency for the age specific indicator in the case of Coloureds group was 2 and the lowest was 1.

UNIVERSITY of the
WESTERN CAPE

With Myer's index Table of Indian/Asian group, the index by sex is 11.00 for males and 12.70 for females. However, these indexes of Myers are close to 0 than 90 the full range of Myer's which means the quality of age distribution of Indian/Asian population group possibly is relatively accurate. The values of indexes are greater than 10, the deviation. Indeed, males were attracted to 4, 8, 2, 5 and 0 but repulsive to 1, 9, 6, 7 and 3. The females were fascinated with numbers ending with 9, 1, 2, 5 and 0, repulsive to 8, 6, 3, 7 and 4. The both sex together were attracted to 0, 2 and 5; repulsive to 3, 6 and 7. The lowest frequency for the age specific indicator according to Indian/Asian group is 8 and the highest was 0. Males probably did better declaration on age than females. The preference for these digits among males may be attributed to the greater tendency to overestimate the age, whilst for females; it may be due to underestimation of their age.

Considering the table of Myers index of Whites, when applying the LFS 2005, the Myers index of males is equal to 6.45 males, for females is 6.44. However, female respondents probably did better declaration on age than males. The astonishment is that the preference with numbers in this particular population group was the same. The males and females were avoided to numbers ending with 1, 3, 8, 4 and 7; attracted to 2, 9, 6, 0 and 5. The set of them likely did good declaration of age. The two Myers indexes were low and close to 0, consequently, the quality of data was in all probabilities very accurate. The terminal digit with the highest frequency for this particular group was 5 and 1 the lowest.

4.4.2.4.2.2 Cross-Ethnic Variations Myer's Index using LFS 2005

Considering LFS 2005, Coloured group likely encountered the best results with 5.68 for males' index and 4.64 for females compared to the other population groups. Coloureds recorded the smallest value of Myers index with LFS 2005. Females possibly declared better their ages than males. In the same vein, African/Black and White females probably did better declaration than males. The contrary was observed by the population of Indians with males' index likely greater compared to females. The highest value of Myer's index for males is 11.00 (Indians) and is 12.70 (Indians) for females. Particularly, the index for males obtains at national level is likely less compared to the smallest occurred across the population groups.

UNIVERSITY of the
WESTERN CAPE

4.4.2.4.2.3 Results Myer's Index for LFS 2007 per Population Groups

The Table 4.7 shows that for African/Blacks, the Myer's index of males is 3.23 while for females is 3.96 indicating the quality of responses on age probably was very accurate. The sub-group of Coloureds presents 6.20 for males, while 5.66 for females. Both indexes are close to 0 than 90 which means the quality of age and sex distribution was possibly very accurate. GHS 2007 was particular for Indian females who changed the trend despite the fact that its index possibly remains the highest compared to other population groups. Although males' index is 11.12 while females' index is 10.51 showing the quality of its data likely was relatively accurate, hence the value of each index is slightly greater than 10, the deviation. The smallest index for females is 3.96 (African/Blacks) which is probably bigger than the one obtained at national level. African/Blacks males with 3.23 of Myer's index possibly did better declaration on age than

females. At the stage of population groups, African/Black males have probably the smallest index compared to other population groups; but it is likely bigger than the index found at national level which is 3.18. The value of highest index is doubtless 11.12 for Indians/Asians. Coloureds, Indians and Whites females likely did better declaration on age than males.

Table 4.15 Distribution of digit preferences per population groups for Labour Force Surveys

		LFS 2005		LFS 2007	
		Repulsive	Attractive	repulsive	Attractive
Blacks	Male	1 6 9 7	3 2 8 4 0 5	1 3 4	5 2 7 0 6 8 9
	Female	1 7 6 3 8 9 4	2 5 0	1 9 6 3 8	7 4 5 2 0
Coloured	Male	1 5 7	8 9 3 6 2 0 4	1 3 9 7	6 2 0 4 8
	Female	1 3 8 7	0 6 5 4 9 2	9 1 7 8	4 0 3 2 5 6
Indians	Male	1 9 6 7 3	4 8 2 5 0	1 8 4 7 5	0 3 9 6 2
	Female	8 6 3 7 4	9 1 2 5 0	4 9 6 8 7	0 2 3 1 5
Whites	Male	1 3 8 4 7	2 9 6 0 5	1 5 6	4 0 9 8 7 2 3
	Female	1 3 8 4 7	2 9 6 0 5	2 3 4 1 8	0 9 6 5 7

UNIVERSITY of the
WESTERN CAPE

4.4.2.4.2.4 Cross-Ethnic Variations Myer's Index between GHS and LFS.

When looking at the results of Myer's among population groups, there were some similarities. For instance, African population group has the same terminal digit for the highest frequency 0 and the lowest 1 for the GHS 2007, GHS 2004 and LFS 2005. This phenomena of preference factor indicates that the Indian respondents were possible attracted to the age ending with 0. It was exactly the same for African group when using GHS 2007 and LFS 2005. For the Whites population group, the smallest frequency was 9 and the highest 0.

Comparing female's and male's index, the results showed that according to the GHS 2004, GHS 2007, LFS 2005 and LFS 2007, male respondents probably did better declaration of age than female respondents except in the case of Whites group in GHS 2007, the case of coloureds and Whites groups in LFS 2005. The African group realized improvement in their data over years. This occurred with the decrease of their indexes; in GHS 2004, males' index was 5.70, in 2007 it was 3.66. According to females the value is 9.29 in GHS 2004 and in 2007 it is 3.93; the improvement is progressive over years. The same curve was observed with LFS 2005 and 2007 for males and females. Coloureds males and both sexes of Whites have registered the same trend

according to GHS; its indexes have decreased in GHS 2007. Different curve was achieved by the Indian population group. Its indexes have increased in GHS 2007, but when applying LFS, the males' index increased slightly from 11.00 to 11.12 and females' index reacted in contrary, it decreased in LFS 2007. One reason of that was the problem of measurement of instrument, LFS and GHS do not have the same scope, the collection of information differed from one instrument to another and the techniques of approaches were also different.

Table 4.7 Results Myer's Index per Population Groups

Population Groups	GHS 2004		GHS 2007		LFS 2005		LFS 2007	
	Male	Female	Male	Female	Male	Female	Male	Female
African/Blacks	5.70	9.29	3.66	3.93	10.50	5.68	3.23	3.96
Coloureds	4.17	6.07	2.77	6.26	5.68	4.64	6.20	5.66
Indian/Asians	8.09	9.50	8.33	10.70	11.00	12.70	11.12	10.51
Whites	17.49	7.38	7.26	5.48	6.45	6.44	5.06	4.56
National	8.51	6.39	3.68	4.18	4.50	5.57	3.18	3.87

UNIVERSITY of the
WESTERN CAPE

4.4.2.4.3 Results of Myer's Index for Census 2001 and Community Survey 2007 per Population Groups

4.4.2.4.3.1 Results of Myer's Index for Census 2001 per Population Groups

The Table 4.8 presents in general but higher value of Myer's index among females than males, thus showing more accurate age reporting among males than females. Despite the fact that its quality of age data were very accurate in the African/Blacks, its indexes (3.27 and 3.33) were highest than the one in all the three other population groups, implying better age reporting in all population groups. African/Black males were attracted to digits 1, 9 and 8 while repulsive to numbers with digit ending 0, 5, 3 and 2. The females were attracted to digits 1, 9 and 8 while repulsive to digits 4, 0, 5 and 7. The digit ending of the age with highest frequency was 1 and the smallest one was 4

According to Coloureds, the index of males was 2.26 and for females 2.32. However, age was more accurately reported among males than females. Males preferred to digit ending 1, 9, 0 and 8 while repulsive to digit 4, 3, 2 and 5. Females were fascinated to digit 0, 9, 8 and 1 while avoided to digit 5, 7, 6 and 3. The value of Myer's index is considered very small compared to

the full range of this index which varied from 0 to 90. The quality of data was highly accurate. As far as the Coloured population group is concerned, the age with digit ending 1 was the age corresponding to the highest frequency and the least one was 4.

As far as Indians are concerned, the females' index is 2.24 and 2.15 for males. The females' index is greater than for males showing that the age reporting was better for males compared to females. Males were motivated to digit ending 8, 0 and 9 while avoided to digit 3, 6, 5 and 1. Females were attracted to digit ending with 0, 6, 5 and 7 while repulsive to digit 2, 3 and 4. However, the quality of age reporting was accurate within the two distributions. The terminal digit of the age with the highest frequency is 2 while with the least frequency is 0.

The Table 4.4 presents White males index using Census 2001 equal to 1.44 and for females is 1.6. It is less for females indicating that the quality of responses on age for males is better compared to females. However, the quality of age and sex distribution is very accurate in regard to standards of Myer's index. Males were attracted to digit ending with 1, 0 and 7 while avoided digits 3, 5 and 4. Females were attracted to age digit ending 0, 1 and 9 while repulsive to digit 2, 3 and 4. The digit ending of the age with highest frequency is 1 and the smallest one is 4.

UNIVERSITY of the
WESTERN CAPE

4.4.2.4.3.2 Cross-Ethnic Variations Myer's Index for Census 2001

Taking into account the Census 2001, when comparing the four population groups, Whites provides the best results; the smallest Myer's index for males 1.04 and the smallest Myer's for the females 1.60. Whatever, the males index of Whites is the best compared to all indexes of population groups and at national level using Census 2001. The highest value of males index is 3.27 (Africans/Blacks) while for females is 3.33 (Africans/Blacks). Males avoided census 2001 ages with digit ending 3 and 5.

4.4.2.4.3.3 Results of Myer's Index for Community Survey 2007 per Ethnic Groups

Observing the Table 4.8, the overall results of Myer's index for Community Survey 2007 per ethnic groups is fewer than 5. Indeed, the values of these indexes revealed that the quality of data on age is likely very accurate. The reporting of age is probably very accurate among both sexes

(respondents). However, Black respondents scored highest value of Myer's (2.33) for males, (3.26) for females while Coloured and Indian male respondents register the smallest Myer's index (1.97). In the same vein, White female respondents score smallest in Myer's index (1.76). Black, Coloured and Indian female respondents obtain higher Myer's index compared to male counterparts which means the reporting of age was probably better among male respondents than female respondents. White population present different option with female respondents who did better report of their age than males despite the fact that the quality of their index had deteriorated, the value of its index has decreased.

Table 4.16 Distribution of digit preferences per population groups for 2001 Census and Community Survey 2007

		Census 2001		Community 2007	
		Repulsive	Attractive	repulsive	Attractive
Blacks	Male	4 0 2 5 3 7	1 9 8 6	5 9 3 0 8 2	7 4 6 1
	Female	4 0 5 7 2 3 6	1 9 8	5 9 0 3	4 7 6 1 2 8
Coloured	Male	4 3 2 5 7 6	1 9 0 8	9 3 0 5	4 1 6 7 2 8
	Female	4 7 3 2 6 5	0 9 8 1	9 8 2 1	6 7 3 5 0 4
Indians	Male	3 6 1 5 4 7 2	8 0 9	9 8 1 0 7	6 5 2 4 3
	Female	2 3 4 1	0 6 5 7 9 8	8 0 5 7	2 3 4 6 9 1
Whites	Male	3 5 4 2	1 0 7 8 9 6	3 4 1 8 0 9	6 7 2 5
	Female	2 3 4 6 5	0 1 9 8 7	1 0 9 2 3 7	6 4 5 8

UNIVERSITY of the
WESTERN CAPE

4.4.2.4.3.4 Inter-Variations Ethnic Groups for Census 2001 and Community survey 2007

Comparing both instruments, the indexes of Black, Coloured respondents and Indian male respondents has decreased showing probably a progressive improvement over years while indexes of Indian female and White population has increased indicating deterioration in the quality of declaration on age. This means deterioration of data quality was prevalent in Indians and White population groups. The age and sex distribution of these population groups remain likely accurate despite the variation observed. Black respondents registered highest index for both sexes. Coloured and Indian male respondents score together the smallest Myer's index (1.97) while White females obtained the smallest Myer's index (1.76) which is lesser than the one obtained at national level.

Table 4.8 Results Myer's Index for Census 2001 and Community Survey 2007 per Population Group

	C S 2007		Census 2001	
Population Groups	Male	Female	Male	Female
African/Blacks	2.33	3.26	3.27	3.33
Coloureds	1.97	1.99	2.26	2.32
Indian/Asians	1.97	2.30	2.15	2.24
Whites	1.98	1.76	1.44	1.6
National	2.08	2.66	2.77	2.77

4.4.2.7 Results Myer's Index at Provincial Level

The Table 4.5 also presents an overall marginal increase and decrease observed in the Myer's index. It displays to see if there was an improvement done in the quality of age data. It further shows a higher index for males compared to females in all the instruments indicating that age is more accurately reported among males than females, or the inverse. The trend is observed at the provincial level of the whole South Africa; if there, the index has been rising over years with different surveys, indicating deterioration in the quality of age data.

UNIVERSITY of the
WESTERN CAPE

4.4.2.7.1 Results of Myer's Index for General Household Surveys per Province

4.4.2.7.1.1 Results of Myer's Index for GHS 2004 at Provincial Level

Looking at the results of GHS 2004, all the indexes were low, close to 0 than 90 which meant that the quality of data was accurate despite the fact that the slight preference or avoidance appears in the age-sex structure of the population. The males of Eastern Cape, Northern Cape, KwaZulu-Natal, North West, Gauteng, Mpumalanga and Limpopo probably did better declaration on age than females; 7 out of 9 provinces had good declaration. The smallest index for males is 4.66 for Mpumalanga and the highest was 9.31 for KwaZulu-Natal, compared to other provinces. This index for Mpumalanga is probably smaller than the one found at national

level. Males are likely more conscious during the operation of GHS 2004 than females which can be discovered through their attitudes. Mpumalanga is a dominantly rural area, yet it exhibits the lowest index. Again, this seems to indicate that age is properly directed in this situation where there is a high level of cooperation.

Considering the females, only 2 out of 9 provinces likely did better declaration than males. However, Western Cape and Free State females provide probably better responses on age than males. The highest female's index is 9.99 (KwaZulu-Natal) and the smallest is 4.76 (Western Cape). The female's index found in Western Cape is possibly the best compared to other provinces and even at national level.

4.4.2.7.1.2 Results of Myer's Index for GH.S 2007 per Province

According to GHS 2007, females come out with likely better declaration on age than males. This is possible confirmed in 6 provinces as follow Northern Cape, Free State, North West, Gauteng, Mpumalanga and Limpopo. The highest index for females is 6.42 (Mpumalanga) and the smallest is 4.42 (North West). This particular index for North West is probably bigger than 4.18 found at national level which confirms the hypothesis that: "The quality of declaration of data on age and sex is likely poor at provincial compared to national level".

Taking into account the indexes of males, just three provinces had succeeded to declare probably better data on age than females. However, males of Western Cape, Eastern Cape and KwaZulu-Natal came out with possibly better results than their female counterpart. The highest value of male's index is 8.89 (Gauteng) and the smallest is (3.51) Western Cape which is likely smaller than the one find at national level. This situation is the case that the hypothesis is rejected; the quality of declaration of data on age and sex is poor at provincial compared to national level.

4.4.2.7.1.3 Comparison Results Myer's Index between GHS 2004 and GHS 2007

Observing the results of the two surveys, the improvement occurred at different stages. Taking the GHS 2004, males were dominant with possibly better results compared to females, the contrary appears with domination of females, and this can be linked to the empowerment or to

the participation of females in the Labour Force which was brought to light in their attitudes. The policies built by the government to improve the quality of health of its population increased the integration of females in the Labour sector.

The decrease of male's index is observed in 7 out of 9 provinces. This issue concerned Western Cape, Eastern Cape, Northern Cape, Free State, KwaZulu-Natal, North West and Limpopo. This decrease explained the progressive improvement which is achieved by the related population. Eastern Cape and North West had a very slight improvement in the quality of declaration of data on age-sex. The highest male's index for both surveys is 9.31 (KwaZulu-Natal) in 2004 and the smallest is 3.51 in 2007 which is likely smaller than the one find at national for both surveys.

The decrease of female's index was observed in 8 out of 9 provinces as followed: Eastern Cape, Northern Cape, Free State, KwaZulu-Natal, North West, Gauteng, Mpumalanga and Limpopo. Feskens *et al*. (2006) suggests that it was more useful to make questionnaire more understandable to reduce the bias in the responses. This situation finds its answer in the revision of the questionnaire over years. The population has become more involved and concerned by this purpose; knowing the importance of the survey, the change of trend can be explained by the application of all advices the fieldworkers received during their training and this approach convinced their respondents. The highest value of females' index observed in the two surveys is possibly 9.99 (KwaZulu-Natal) and the smallest is 4.42 (North West) which is likely bigger than 4.18 found at national level.

4.4.2.7.2 Results of Myer's Index for Labour Force Surveys per Province

4.4.2.7.2.1 Results of Myer's Index for LFS 2005 at Provincial Level

As far as the LFS 2005 is concerned, the whole results of Myers index at provincial and national level are good showing that the quality of declaration on age-sex is in all probabilities accurate; hence the index of males observed less value in seven provinces. Its indexes are likely close to 0 than 90. This has been observed in Eastern Cape, Northern Cape, Free State, KwaZulu-Natal, North West, Gauteng and Mpumalanga. Referring to these provinces, males probably did better quality of declaration on age-sex at provincial level. The highest index for males is 7.39

(Western Cape) and the smallest is 4.29 (Northern Cape). This index of Northern Cape is also smaller than 4.50 found at national level.

Considering females, only in two provinces where females likely did better declaration on age and sex. Western Cape and Limpopo presented better results of females. The index 4.37 (Limpopo) is the best compared to the eight provinces and even to 5.57 found at national level. Females tend to round off ages even those who are not able to remember with precise date of birth when stating their age preferred for instance 1940, 1950, 1960, 1970, etc.

4.4.2.7.2.2 Results of Myer's Index for LFS 2007 per Provinces

Looking at the Table 4.4 of Myer's results, in Western Cape, Eastern Cape, KwaZulu-Natal, North West and Gauteng, male respondents possibly declare better their age than females. The quality of their responses is likely accurate. The highest value of the index is 6.94 (Northern Cape) and the smallest is 3.92 (KwaZulu-Natal) which is probably bigger than (3.18) observed at national stage.

Considering the LFS 2007, in Northern Cape, Free State, Mpumalanga and Limpopo, females' index is less than for males. Females probably did better responses on the declaration of age and sex. The smallest index for females is 4.54 (North West) compared to other provinces. It is bigger than (3.87) observed at national level.

UNIVERSITY of the
WESTERN CAPE

4.4.2.7.2.3 Comparison Results Myer's Index using LFS at Provincial Level

Regarding results of both surveys, some changes are observed in LFS 2007. In fact, the decrease of males' index occurred in Western Cape, Eastern Cape, KwaZulu-Natal, North West and Gauteng. The improvement is progressive in these areas. The seriousness of the fieldwork and advices received during collection operation are possible explanations for this purpose. The increase observed in Northern Cape, Free State, Mpumalanga and Limpopo can be explained by the predominance of rural areas in these provinces which can be also be associated with the issue or influence of low level of educational status, the consideration of marital status of respondents. The highest value of males' index for both surveys is 7.39 (Western Cape) observed in LFS

2005, and the smallest is 3.92 (KwaZulu-Natal) observed in LFS 2007 which is higher than 3.18 for National level.

The females' index decreases in Northern Cape, Free State, KwaZulu-Natal, North West, Gauteng and Mpumalanga. The slight improvement is observed in Northern Cape, Kwazulu-Natal and Mpumalanga while a substantial improvement occurred in Free State, North West and Gauteng. The females had probably expressed better the quality of responses on age-sex compared to males in these provinces. The highest index for females is 7.91 (Gauteng) and the smallest is 4.37 (Limpopo) which is also higher than 3.87 observed at National level). The preference for these digits among males may be attributed to the greater tendency to overestimate the age, whilst for females; it may be due to underestimation of their age.

4.4.2.7.3 Results of Myer's Index for Census 2001 and Community Survey 2007 at Provincial level

4.4.2.7.3.1 Results of Myer's Index Census 2001 at Provincial Level

Considering Table 4.6, the overall Myer's index is less than 5 indicating that the reporting of age was properly done by the respondents. However, these findings revealed that the pattern of preference or avoidance at ages ending in any digit from 0 to 9 when applying Census 2001 was probably insignificant. All values of Myer's index for both sexes at provincial level were very low and close to 0 than 90 showing that the quality of data on age was probably very accurate despite the presence of slight preference or avoidance observed in the age-sex distribution of the Census 2001 population. Eastern Cape, Free State and Limpopo out of nine provinces showed higher index for females than males implying males did better reporting of age compared to females. The exact contrary appears in other provinces where females probably declared better their ages than males. Both sexes in Gauteng obtained the smallest value of Myer's index while male respondents of Eastern Cape scored the higher Myer's index (4.12) and female respondents of Limpopo registered also highest Myer's index (4.45)

4.4.2.7.3.2 Results of Myer's Index for Community Survey 2007 at Provincial Level

Considering Table 4.6, Myer's index is less than 10 in general which reveals that the quality of age reporting was good. These indexes confirm that the heaping at ages ending in any digit from

0 to 9 using the 2007 Community survey was likely insignificant. All the indexes were low, close to 0 than 90 which meant the quality of data was accurate despite the fact that the slight preference or avoidance appeared in the age-sex structure of the population.

Northern Cape and KwaZulu-Natal out of nine provinces indicate higher index for males than females. This reveals that females reported better their age than males. Northern Cape males' index is higher than the one of KwaZulu-Natal males whilst the females' index of the latter resulted higher than the one of Northern Cape. Indeed, in Northern Cape, males mostly preferred ages with digit ending 4, 7, 6 and 1 while avoiding digit 2, 0 and 5; females were repulsive to digit ending 1, 9, 0 and 2 while attracted to digit 3, 6, 7 and 8. In KwaZulu-Natal, females were attracted to numbers with digit ending with 4, 7 and 6 while avoided digit ending 5, 9, 3 and 8; males preferred numbers with digit ending 4, 7, 6 and 1 while repulsive to numbers with digit ending 5, 9, 0 and 3. The terminal digit with highest frequency is 4 and the digit ending register smallest frequency is 5

The contrary was observed in other provinces where males did better declaration on age than females. As far as Western Cape is concerned, males obtained the smallest value of index compared to the set of indexes in Table 4.6. They were attracted to ages with terminal digit 1, 6, 0 and mostly repulsive to terminal 9 and 5. Females were also mostly attracted to numbers with digit ending 6, 4 and 7 while avoided age digit ending with 9, 5, 8 and 3. The digit ending of age with highest frequency was 6 and the least one is 9.

According to Eastern Cape, males show the highest index compared to the full range of males' index. Females' index is the highest compared to the full set of data in Table 4.6. It is also greater than for males indicating that males did age reporting better than females. The latter were mostly attracted to ages with digit ending 6, 4 and 7, while repulsive to terminal digit 5, 9, 3 and 0. Males mostly preferred in numbers ending with 6, 7 and 4 while avoided digit ending with 3, 5, 1 and 9.

Looking at the results, Free State male's present better index than females which mean the quality of age reporting is more accurate among males compared to females. However, males were mostly attracted to ages with digit ending with 4, 6 and 7 while repulsive to numbers ending with 5, 1, 9 and 0. Females were most attracted to ages ending with also 4, 6 and 7 while avoided numbers ending with 0, 5 and 9.

Observing Table 4.6, North West reveals that the male's index is less compared to females which mean the quality of age reporting is more accurate among males than females. Females were interested in reporting ages ending with 4, 7, 8 and 6 while avoided terminal digit 5, 9 and 0. Males were most attracted to ages ending with 7, 4, 2 and 1 while avoided 5, 9, 0 and 3.

Taking into account Gauteng province, females' index was higher compared to males indicating the quality of age reporting is more accurate among males than females. Males were mostly attracted to numbers with terminal digit 4, 6, 2 and 3 while avoided numbers ending with 8, 5, 9 and 0. Females were fascinated to report ages with terminal digit 4, 6, 2 and 7 while repulsive mostly to numbers ending with 5, 9 and 3.

According to Mpumalanga, females' index is higher compared to males showing that the quality of age reporting is more accurate among males than females. Males were strongly attracted to numbers ending with 7, 6 and 4 while avoided numbers mostly ending with 3, 5 and 8. Females were also strongly attracted to ages with terminal digit 7, 4 and 6, while avoided numbers ending with 0, 5 and 3.

Regarding Table 4.6, Limpopo male's index is less compared to female indicating that the quality of age reporting is less accurate among females than males. However, males were fascinated most in reporting ages with digit ending 7, 8 and 6 while avoided strongly in numbers with terminal digit 0, 3 and 5. In the same vein females were obstinate in ages ending with 7, 6 and 4 while repulsive to numbers ending with 5, 9 and 0.

4.4.2.7.3.3 Comparison Results of Myer's Index for Census 2001 and Community Survey 2007 at Provincial Level

When looking at the findings obtained at provincial level, the data informed that male population has improved in general the quality of its data. Except male of Northern Cape, the overall index has decreased over years in eight provinces. As far as females were concerned, four provinces out of nine observed deterioration in the quality of its data. There are Eastern Cape, North West, Gauteng and Mpumalanga. Except Gauteng, Black population is mainly living in these provinces with rural predominance. Four provinces observed improvement of data quality in both sexes over years. These provinces are Western Cape, Free State, KwaZulu-Natal and Limpopo. The preference or avoidance in certain digits ending is likely very slight; one can conclude this

pattern is probably insignificant across the South African population. However, Western Cape, both sexes registered the smallest Myer's index for both sexes while Eastern Cape resigned in the highest Myer's index for both sexes. Male respondents of Northern Cape also scored the same as those of the Eastern Cape.

4.4.2.8 Comparison Results Myer's Index for Community Survey 2007 between National and Provincial Level.

The highest value of males' index is 3.58 (Eastern Cape and Northern Cape) and the smallest is 1.77 (Western Cape). The highest value of females' index is 4.75 (Eastern Cape) and the smallest is 1.91 (Western Cape). However, Eastern Cape encountered the highest values of index according to males and females on one hand and Western Cape registered the smallest and best values of index compared to the full range of results at provincial level. The age reporting in the Western Cape was better improved compared to the one at national level.

The difference between index at national and highest index for males for instance was equal to 1.5 (3.58 – 2.08 = 1.5) which is even less compared to the smallest males' index (1.77); this implied that these indexes were reliable; they were close one to another. The gap between them is small in general. Looking at these values of index, it reveals that the difference between the two highest indexes is 1.17 (4.75 – 3.58 = 1.17) which is also less compared to the smallest index of male or female. The quality of age reporting is probably more accurate among males than females across provinces.

4.4.3 Results of Combined Index of United Nations (CIUN)

The evaluation of the quality of age and sex data using the method of United Nations is applied in this study in terms to ascertain the fluctuations occurred in the 5-year age group distribution. The best of this index is that it reflects the variation in quantity of omissions per age group, the wrong declaration of age and preferences for ages ending with any digit. It explains better the degree of exactness of statistics per age than the previous. It provides the precision per age group instead of single age. The standard for quality of sex and 5-year age group distribution are as follow; the CIUN less than 20 means the quality of sex and 5-year age group distribution is accurate. When it lies between 20 and 40 the quality of the distribution is considered inaccurate.

UNIVERSITY of the
WESTERN CAPE

In the CIUN case, it is more than 40, it is very inaccurate. The recommendation is to pay attention before doing editing of data or making projections because it is useful especially in the case of CIUN which is higher than 40. The CIUN is evaluated based on the assumption that normal age and sex distribution which implies the sex ratio decrease gradually from one age group to another. The population using 5-years age group decreases as age increases and its value will be approximately equal to the mean value of the both 5-years age groups previous and after it. The index also shows which sex is contributing in better declaration of age.

4.4.3.1 Results of CIUN for General Household Surveys at National level

4.4.3.1.1 Results of CIUN for GHS 2004 at National stage

At national level, the result of CIUN obtained from GHS 2004 is 28.37; the Table set out that the distribution of the age and sex is probably inaccurate. It also provides the distribution of sex ratio per age group. For the GHS 2004, the age specific indicator of male is 6.15; for female is 8.62 and the indicator of variation of sex ratio is 4.53. The index of females is higher than for males showing that the more accurate age reporting is likely observed among males than females.

UNIVERSITY of the
WESTERN CAPE

4.4.3.1.2 Results of CIUN for GHS 2007 at National level

Considering GHS 2007, the CIUN is 20.78; the age specific indicator for the male is 6.83; 3.81 for the female and 3.38 for the indicator of variation of sex ratio. Definitely the quality of sex and age group distribution of GHS 2007 is possibly inaccurate. However, the age reporting was likely better among females than males, hence the males observed greater index of age than females.

4.4.3.1.3 Comparison Results of CIUN between GHS 2004 and GHS 2007 at National Level

Looking at Table 4.10, the results of GHS 2004 and 2007 show that, the slight improvement is really carried on. However, the slight decrease observed could not bring a visible change in the quality of data on age and sex distribution. This situation could be attributed to the effect of migration during the end period of apartheid. Historically, the impact of apartheid is significant on the population of South Africa. The sex and age indexes experience abnormal variations for

the age groups over 50. There are two dimensions to observe in this particular agreement; firstly, during apartheid period, some people were moved in search of peace land (Zimbabwe, Mozambique, Lesotho, Botswana and Swaziland) and secondly, others were faced death because of atrocity treatments of colonialism.

4.4.3.2 Results of CIUN for Labour Force Surveys at National Level

4.4.3.2.1 Results of CIUN for LFS 2005 at National Level

According to the LFS 2005, the CIUN was 19.74, the age specific indicator for the males was 4.58, for females 4.46 and the indicator of variation of sex ratio was 3.57. The quality of sex and 5 year age group distribution of LFS 2005 is likely relatively accurate based on the criteria of United Nations (19.74 was less than 20); hence it is close to upper limit 20. This could be the effect of the training, the instructions and awareness addressed to the enumerators. The age specific for males is higher than for females showing that the age reporting is probably better for females compared to males.

4.4.3.2.2 Results of CIUN for LFS 2007 at National Level

UNIVERSITY of the
WESTERN CAPE

Considering the LFS 2007, the CIUN is 27.05, the age specific indicator for males is 6.86 and for females is 3.90 and the indicator of variation of sex ratio is 5.43. The quality of 5-years age and sex distribution is probably inaccurate. The age reporting of females' respondents is more accurate compared to males' respondents.

4.4.3.2.3 Comparison Results of CIUN between LFS 2004 and LFS 2007 at National Level

Looking at both results, LFS 2005 is probably better compared to LFS 2007. Instead of decrease, the CIUN has increased considerably in 2007 and this abnormal variation in the age and sex indicators affected the quality of age and sex distribution of LFS 2007 consequently, the quality of age sex reporting is deteriorating, CIUN move from 19.74 to 27.05. The CIUN is developed based on the assumption that normal age and sex structures which mean the sex ratio gradually decreases from one age group to another and at the highest age group, the value of this index ranges from 95 to 100. Population size corresponding to 5-year age groups decreases as age

increases. The male's index of age has increased consistently from 4.58 to 6.86 and the female's index of age has decreased slightly. The sex ratio has also increased. This could be explained by the attitude of male respondents who tend to overestimate their ages which bring them to shift in older ages and also the females respondents who tend to under estimate their ages which push them to shift in younger ages.

4.4.3.3 Results of CIUN for Census 2001 and Community Survey 2007

Overview

In the same vein, as the researcher has mentioned at the beginning of the sections 4.4.1.7 and 4.4.2.3, Community Survey 2007 was representative at government levels and the results can be extrapolated to the entire country for inferences. The variations of the indexes during the 6 year period had a government purpose of building policies into the 2001 censes in line with improving the living conditions of its people.

4.4.3.3.1 Results of CIUN for Community survey 2007 at National Level

Observing the Table 4.5, the CIUN obtained from Census data is 18.32. Based on the standards of this index, the distribution of the age and sex is probably accurate. Whatever, the specific age indicator of males are 4.08, for females 3.53 and the indicator of variation in sex ratios is 3.57 for the whole population census. Females did better reporting on age compared to male respondents.

4.4.3.3.2 Results of CIUN for Census 2001 at National Level

Considering the census 2001, the CIUN show 19.30 indicating that the age and sex structure is possibly relatively accurate; hence the value of its index is more close to upper limit (20). The age specific indicator of males is 5.23, for females 4.08 and the indicator of variation in the sex ratios is equal to 3.33 at national level. The females did likely better declaration of age compared to males.

UNIVERSITY of the
WESTERN CAPE

4.4.3.3.3 Comparison Results of CIUN between Census 2001 and Community Survey 2007

Taking into account the results of Census and Community survey, the slight improvement arises in statistics related at national level. The small experimental decrease of statistics could not be able to produce a visible change in the quality of data on age and sex distribution. However, the age specific indicator for males decreases by 0.55; for females by 1.15 (5.23 – 4.08 = 1.15), the variation in sex ratios increases only by 0.24 (3.57 – 3.33 = 0.24) and the CIUN decreases by 0.98 (19.30 – 18.32 = 0.98). The pattern of migration could influence these fluctuations of statistics, whatever; the quality of distribution is accurate with each of these instruments. Both instruments indicated that the reporting on age was better for females than males.

4.4.3.4 Results of CIUN for GHS at Provincial Level

4.4.3.4.1 Results of CIUN for GHS 2004 per Province

Seeing that in GHS 2004 results, some provinces presented similarities, it has been grouped at provincial level into two categories. The first group includes Western Cape, KwaZulu-Natal and Gauteng with range of CIUN is between 20 and 40. This shows that the quality of sex and 5 year age group distribution is likely inaccurate. The second group is constituted with the following provinces: Eastern Cape, Northern Cape, Free State, North West, Mpumalanga and Limpopo with CIUN are likely greater than 40. This indicates that the quality of sex and 5-year age group distribution is likely greater inaccurate.

Furthermore, the quality of responses on the declaration on age also grouped the population in two categories; the first group constitutes of Western Cape, Northern Cape, Free State, KwaZulu -Natal and Mpumalanga where males probably did better responses on age compared to females. The second took into account Eastern Cape, North West, Gauteng and Limpopo where males recorded higher index of age specific than females which implies that females likely did more accurate age reporting than males. The highest index is 80.15 (Mpumalanga) and the smallest is 34.39 (Western Cape). However, the possibly explanation to this case of Mpumalanga is that this province has a predominance of rural areas with its factors influencing the quality of responses; or may be the low level of educational status which also influences the quality of response or, the

UNIVERSITY of the WESTERN CAPE

pattern of out-migration which is observed for instance among younger males in search of better conditions of life (new opportunities for employment or fulfill their education).

4.4.3.4.2 Results of CIUN for GHS 2007 at Provincial Level

Considering GHS 2007, the results at provincial level are probably the same as those found in the previous survey. However, Western Cape, KwaZulu-Natal, North West and Limpopo constitute the first category with its results range between 20 and 40 indicating that the quality of 5-years age and sex distribution is likely inaccurate among these provinces. The second category takes into account Eastern Cape, Northern Cape, Free State, Gauteng and Mpumalanga with its results possible greater than 40 reveal that the quality of age and sex structure is possible very inaccurate. The Table 4.6 showed that in seven out of nine provinces, males obtain higher index of specific age than females which implies in these provinces except Western Cape and Northern Cape, females probably registered better quality of age reporting compared to males while in Western Cape and Northern Cape males in all probabilities did accurate reporting on age than female respondents. The highest index of CIUN is 52.74 (Mpumalanga) while the smallest is 36.16 (Western Cape).

UNIVERSITY of the
WESTERN CAPE

4.4.3.4.3 Comparison Results of CIUN between GHS 2004 and GHS 2007 at Provincial Level

Looking at the table of results, the provinces could be divided in two parts; the first set concerns provinces which its CIUN has increased. There are Western Cape, KwaZulu-Natal and Gauteng. The rest of provinces constitute those which its index has decreased. However, the value of CIUN is likely high among provinces. The CIUN of Western Cape, KwaZulu-Natal, North West and Limpopo are ranged in the same interval 20 and 40 which mean the quality of sex and 5-year age group distribution is possibly inaccurate. As far as Eastern Cape, Northern Cape, Free State, Gauteng and Mpumalanga are concerned, its indexes are ranged over 40 indicating the quality of responses on declaration on age and sex is likely very inaccurate compared to national index.

4.4.3.5 Results of CIUN for both LFS at Provincial Level

4.4.3.5.1 Results of CIUN for LFS 2005 at Provincial Level

Using the LFS 2005, the results of CIUN in provinces are not probably of good quality in general. Western Cape, Northern Cape, KwaZulu-Natal, North West, Gauteng and Limpopo realize CIUN range between 20 and 40 indicating that the quality of data on age and sex is likely poor which mean the quality of age distribution is possibly inaccurate among these provinces. Considering Eastern Cape, Free State and Mpumalanga, its results of CIUN are above 40 showing that the quality of sex and 5 year age group distribution is likely very inaccurate. Regarding the Table 4.6, in Northern Cape, North West and Gauteng, females have higher average age ratio deviation than males which mean the age reporting is better for males compared to females in these provinces. Looking at the results, Western Cape, Eastern Cape, Free State, KwaZulu-Natal, Mpumalanga and Limpopo present males with higher average age ratio deviation than females; this implies that the age reporting is better for females than for males in these provinces. The highest index of CIUN is 58.15 (Eastern Cape) and the smallest was 29.87 which is even higher than 19.74, at national level.

UNIVERSITY of the
WESTERN CAPE

4.4.3.5.2 Results of CIUN for LFS 2007 at Provincial stage

As far as LFS 2007 is concerned, the CIUN in provinces are possibly not accurate. Only Free State and Gauteng present index range between 20 and 40 which mean the quality of sex and 5-year age group distribution is probably inaccurate according to the standards established by the United Nations. Thus Western Cape, Eastern Cape, Northern Cape, KwaZulu-Natal, North West, Mpumalanga and Limpopo register CIUN above 40 showing that the quality of age and sex distribution is likely very inaccurate. Despite the fact that all indexes are probably inaccurate at provincial level, Gauteng finds at least 27.60 which is less than 30, but greater than 20. However, this index is likely greater than 27.05, the one find at national level. The Table 4.6 shows that in Western Cape and Free State female respondents registered higher average age ratio deviation than male respondents implying that the age reporting is probably much better for males compared to females. Indeed, Eastern Cape, Northern Cape, KwaZulu-Natal, North West, Gauteng, Mpumalanga and Limpopo registered the inverse, where females have lower average

age ratio deviation than males indicating the age reporting is worse for males than females. The highest index of CIUN is 58.30 (North west) and the smallest is 27.60 (Gauteng).

4.4.3.5.3 Comparison Results CIUN between LFS 2005 and LFS 2007

The CIUN of these instruments find at the provincial level are probably not accurate. The results are above 20. The quality is likely poor over years. However, Eastern Cape, Free State and Gauteng have observed the slight decrease in the value of its indexes that could not even affect any change in the trend. According to Western Cape, Northern Cape, KwaZulu-Natal, North West, Mpumalanga and Limpopo, its indexes have increased showing the quality of 5-years age and sex distribution is probably very inaccurate compared to CIUN obtained at national level.

4.4.3.6 Results of CIUN for Census 2001 and Community Survey 2007 at Provincial Level

4.4.3.6.1 Results of CIUN for Census 2001 at Provincial Level

Considering Census 2001, the findings present three out of nine provinces are probably accurate with their CIUN less than 20. These provinces are Western Cape, Gauteng and Free State. The quality of the data is likely better compared to the one obtained at national level (20.84). However the quality of their 5-year age distribution is possibly accurate. In fact, Eastern Cape, Northern Cape, KwaZulu-Natal, North West, Mpumalanga and Limpopo indicate indexes range between 20 and 40 revealing the quality of each sex and 5-year age group distribution is doubtless inaccurate. Gauteng presents the smallest CIUN (15.05) which was less than the one obtained at national level. Hence, Limpopo indicated the highest value of CIUN (30.13). As far as the table of variation of age specific indicator was concerned, one could see six out of nine provinces indicate male respondents did better declaration of age compared to females. However, in Western Cape, Northern Cape and KwaZulu-Natal, female respondents did probably better declaration of age than males. The quality of reporting on age was consistent among males. The contrary was observed in Western Cape, Northern Cape and KwaZulu-Natal where female respondents realized better declaration on age than male counterparts.

UNIVERSITY of the
WESTERN CAPE

4.4.3.6.2 Results of CIUN for Community survey 2007 at Provincial Level

As far as Community survey is concerned, the results at provincial level can be clustered into two categories. The first group considers Eastern Cape, Northern Cape, Free State, Kwazulu-Natal, North West, Gauteng, Mpumalanga and Limpopo which the results are ranged between 20 and 40 showing the quality of 5-years age and sex distribution is probably inaccurate. The second group is a single, Western Cape with the result less than 20 indicating that the quality of age and sex distribution is possibly accurate; however, the quality of age reporting in this group is likely better compared to the first group and even compared to national level, hence the values of CIUN are probably higher among provinces compared to the one at national level. The highest index is 33.58 (Limpopo) and the smallest is 17.93 (Western Cape) which is probably the best compared to national index.

4.4.3.6.3 Comparison Results CIUN of Census 2001 and Community Survey 2007 at Provincial Level

UNIVERSITY of the
WESTERN CAPE

Referring to both instruments, the indexes of Western Cape, KwaZulu-Natal and Mpumalanga have decreased showing there is probably a progressive improvement over years. Only Western Cape has maintained the accuracy of its data which is good quality compared to what observed in other provinces. KwaZulu-Natal and Mpumalanga age structure are probably inaccurate despite the improvement observed. However, male respondents of KwaZulu-Natal, Eastern Cape, Northern Cape, Free State, North West, Gauteng and Limpopo observed deterioration in the quality of reporting of their ages. As far as female respondents are concerned, only Northern Cape, Free State and Gauteng observed similar patterns. These females likely deteriorate the quality of declaration on age over years. The improvement occurs in the quality of reporting by female respondents in Western Cape, Eastern Cape, KwaZulu-Natal, North West, Mpumalanga and Limpopo. However, the smallest index has increased ($15.05 < 17.93$) and also the highest ($30.13 < 33.58$).

Table 4.9 Results of CIUN for GHS, LFS, Census and Community Survey at National and Provincial Level

Provinces	GHS 2004	GHS 2007	LFS 2005	LFS 2007	C S 2007	Census 2001
Western Cape	34.39	36.16	31.00	41.34	17.93	18.19
Eastern Cape	44.85	40.77	58.15	43.72	27.08	26.14
Northern Cape	66.50	47.92	35.70	57.30	24.15	20.45.
Free State	63.51	50.07	52.96	38.31	26.66	18.49
KwaZulu-Natal	38.93	39.35	33.93	40.04	24.76	26.05
North West	44.93	38.63	31.35	58.28	31.83	20.61
Gauteng	37.86	45.94	29.88	27.60	22.39	15.05
Mpumalanga	80.15	52.74	52.04	58.14	22.83	26.32
Limpopo	44.32	37.21	39.17	48.73	33.58	30.13
National	28.37	24.38	19.74	27.05	18.32	20.84

4.4.3.7 Results of CIUN for GHS per Ethnic Group

4.4.3.7.1 Results of CIUN for GHS 2004 per Population Group

Looking at the Table 4.10 and 4.12, the GHS 2004 presents African/Blacks with 32.36 which is range between 20 and 40; referring to the standards of CIUN, the quality of its data on sex and 5-year age group distribution is likely inaccurate compared to others. The index of age ratio deviation for males is 6.77 while for females is 9.22 implying that the age reporting is likely better for males than for females. The index of variation in sex ratios is 5.46. As far as Coloureds, Indians/Asians and Whites are concerned, its indexes are above 40 indicating the quality of data on sex and 5-years structure is possibly very inaccurate compared to CIUN at national level. The highest index is 63.34 (Indians/Asians) and the smallest is 32.36 (Africans/Blacks) which is likely high than 28.37, the national index. The females of subgroups Coloureds and Whites registered higher average age ratio deviation than males showing that the age reporting is better for males compared to females. Taking into account the subgroup of Indians/Asians, males registered for the index of age ratio deviation 14.80 while females obtained 7.33; this reveals that the female respondents did better in reporting age than male respondents. The highest index of CIUN among population groups is 63.34 (Indians/Asians) while the smallest is 32.36 for African/Blacks.

UNIVERSITY of the WESTERN CAPE

4.4.3.7.2 Results of CIUN for GHS 2007 per Population Group

Taking into account the GHS 2007, African/Blacks and Coloureds index of United Nations are ranged between 20 and 40 showing that the quality of responses is likely poor. Its sex and 5-year age group distribution are possibly inaccurate. In addition, Indians and Whites get above 40 as results showing that the quality of distribution on age and sex is probably very inaccurate. This situation is a confirmation of the same results. There seems to be a strong relationship between race and age declaration. This comes as a paradox that the population group with a long exposure to birth certificate is used with the poor level of age declaration. The highest index is 75.06 (Indians/Asians) and the smallest is 24.18 (Africans/Blacks) which is smaller compared to national index. According to the results found in appendices Table (4.12), males of Africans/Blacks, Indians/Asians and Whites subgroups have registered higher average age ratio deviation than females which mean the females of these population subgroups realized possibly better age reporting compared to males. Coloureds, Indians and whites males improved their quality of declaration. Indeed, female respondents of Coloureds population group have scored higher average age ratio deviation than males indicating that the age reporting is better for male respondents than for female respondents in the same subgroup.

UNIVERSITY of the
WESTERN CAPE

4.4.3.7.3 Inter- Ethnic Variations GHS 2004 and GHS 2007

Observing the two GHS, the indexes of African/Blacks and Coloureds have decreased in 2007; but the slight decrease realized did not allow African/Blacks to change the range in other words the quality of its data was not likely improved; but Coloureds group moves from range 40 to above, to previous range (20 to 40) in GHS 2007. The improvement is not really sensible because the quality is still likely inaccurate. According to Indians and Whites, the position is worse; the quality is doubtless very inaccurate. The high index observed in 2004 has increased as well in GHS 2007. The possibly explanation is probably the non-commitment (involvement) of this population group in the survey programs or also may be the education aspect must be considered. As far as Whites population is concerned, there was no evidence if all their households have been really visited with the problems of access faced by the fieldworkers, or possibly, the political issues continue to influence their declaration or may be some addresses did not be found and the field workers only tried to capture wrong information.

4.4.3.8 Results of CIUN for both LFS per Ethnic group

4.4.3.8.1 Results of CIUN for LFS 2005 per Ethnic Group

Regarding the results of LFS 2005 per population groups, African/Blacks and Coloureds indexes are in the same interval 20 to 40 indicating the quality of sex and 5-year age group distribution is probably considered incorrect compared to national index. In the same vein, Indians and Whites indexes are above 40 showing the quality of responses on age and sex data is possibly very inaccurate compared to national index. The highest index is 77.14 (Indians/Asians) and the smallest is 21.35 (Africans/Blacks) which is higher than national index. The Table 4.12 indicates that the females of subgroup Africans/Blacks and Indians/Asians have higher average age ratio deviation than males showing that the quality of age reporting is possibly better for males than for females. The males of subgroup Coloureds and Whites have observed higher average age ratio deviation than females implying that the quality of age reporting is better for females compared to males.

4.4.3.8.2 Results of CIUN for LFS 2007 per Ethnic Group

UNIVERSITY of the
WESTERN CAPE

Observing the results of LFS 2007, Africans/Blacks population group presents its outcomes range between 20 and 40 showing the quality of sex and 5-year age group distribution is likely inaccurate. Some distortions occurred in its structure. The subgroup Coloureds, Indians/Asians and Whites have encountered indexes range above 40 indicating that the quality of responses on declaration of age and sex is probably very inaccurate. Their population structure by sex and five-year age groups is considered very incorrect. The highest value of CIUN is 69.83 (Indians/Asians) while the smallest is 30.00 which is even higher than the national index. Referring on Table 4.10, female respondents of Africans/Blacks, Indians/Asians and Whites ethnic groups achieve higher index of age ratio deviation than male respondents showing than males which mean age reporting of male respondents probably did better compared to female respondents. It was important to know that a substantial deviation was explained in terms of age misreporting. The Coloured males realized the contrary; they scored more in deviation of age than females implying the age reporting is better for females than Coloureds male.

4.4.3.8.3 Cross-Ethnic Variations LFS 2005 and LFS 2007

According to both LFS, the increase of index has been observed by the Africans/Blacks, Coloureds and Whites population groups. The increase achieved by the Africans/Blacks does not affect the trend of its index or change the interval it belongs to which showing that the quality of responses on declaration on age and sex likely remains inaccurate. Coloureds index shifted in other interval above 40 showing the situation becomes worse, the preference is probably more pronounced; the quality of its sex and 5-year age group distribution is possibly very incorrect. The index of Indians/Asians decreased slightly and it keeps the same interval.

4.4.3.9 Results of CIUN for Census 2001 and Community Survey 2007per Ethnic Group

4.4.3.9.1 Results of CIUN for Census 2001 per Population Group

Considering the 2001 Census, African/Blacks (23.20) and Indians (22.24) index of United Nations are ranged between 20 and 40 showing that the quality of age and sex distribution is likely inaccurate. The age reporting is probably poor among these population groups except Whites. However, Indian females (3.25) did the best reporting on age compared to other ethnic groups. The variation in the reporting of Coloured females (3.86) followed, and then whites females (4.08). The highest index is 23.20 (African/Blacks) and the smallest is 18.72 (Coloureds) which is likely smaller compared to national index (20.84). However the White population indicates 19.30, the value of CIUN which is closed to the upper limit. This reveals that the quality of age and sex structure of its data is likely relative accurate. Considering males, Coloureds (4.18) realized the best reporting on age compared to other males.

UNIVERSITY of the
WESTERN CAPE

4.4.3.9.2 Results of CIUN for Community Survey 2007 per Ethnic Groups

Observing the results of Community Survey 2007 per population groups, only Coloureds got the smallest value of CIUN (19.23) less than 20 revealing the quality of data is likely accurate. This is greater than the one obtained at national level. The CIUN of Black/Africans and Whites are ranged between 20 and 40 indicating the quality of each data is not probably accurate. Indians score the highest CIUN (41.26) showing the quality of its data is inaccurate. However, female

respondents of Black/Africans, Coloureds and Whites did likely better reporting on age data compared to male counterparts. Hence, Indians males declared probably better their ages than female respondents.

4.4.3.9.3 Inter-Ethnic Variations of CIUN for Census 2001 and Community Survey 2007

Looking at the results of both instruments, only Black/African population has decreased the value of Whipple's index. This indicates a slight improvement in the quality of the declaration on age. Nevertheless, the quality of its data remains probably inaccurate. Coloureds, Indians and Whites encountered an increase of their index. However, regardless of the slight deterioration observed by Coloureds, their index was still less than 20 showing that the quality of data on age was probably accurate. The indexes of Black/Africans and Whites ranged between 20 and 40 revealing the inaccuracy of quality of its data on age. In the same vein, Indians registered a great deterioration of the quality of responses hence, they shifted to range 40 and above which indicates the quality of declaration is doubtless inaccurate. Considering the variation of age specific indicator, the reporting on age has changed the trend. Over years, the reporting on age is improved among Black female respondents while male counterparts deteriorated in the quality of their declaration. However, the contrary occurred among Indian, Coloured and White respondents. Over years, males improved slightly the quality of their responses on age while the preference patterns are most experienced among female respondents.

UNIVERSITY of the
WESTERN CAPE

Table 4.10 Results of CIUN per Population Groups for GHS, LFS, Census and Community Survey.

Population Groups	GHS 2004	GHS 2007	LFS 2005	LFS 2007	C S 2007	Census 2001
African/Blacks	32.361	24.176	21.354	30	22.8	23.2
Coloureds	40.307	29.544	31.85	45.899	19.23	18.72
Indian/Asians	63.339	75.063	77.14	69.831	41.26	22.24
Whites	45.631	74.594	40.637	59.924	22.43	19.3
National	28.368	24.383	19.744	27.051	18.32	20.84

4.4.4 Results of Age Ratios Analysis

United Nations (1952) used age ratios per 5-year age group to detect age misreporting in the population where fertility has not fluctuated greatly during the past and where international migration has not been significant. However, Arriaga (1994) established that the age ratios for all groups should be comparatively close across the age groups. The normative is that the age ratios for all categories should be approximate to 100 if there were no extreme fluctuations in the past events. The formula is AR = (100*5Pn) / [0.5*(5Pn-5+5Pn+5)]; where 5Pn is the pop in the given age group, 5Pn-5 is the population in the preceding age group, 5Pn+5 is the population in the following age group. Based on the fact that the age index and sex index quantify only the quality of the age and sex information, Arriaga (1994) suggested that, the age and sex data must also be examined graphically. The researcher has applied for the assessment the age ratios per year.

4.4.4.1 Description of Age Ratio Graphs at National Level for GHS 2004

UNIVERSITY of the
WESTERN CAPE

As shown in Figure 4.1 of age ratios for GHS 2004, it is easily seen and matches where some events had influenced or created imbalanced in the age ratios as well. However, Age ratios of males (C_{im}) and for females (C_{if}) frequently deviate from the normative value of 100. This is the case starting from age 0 to 29; both age ratios are relatively narrowed. C_{im} vary from high of 136.26 at age 32 to a low of 70.30 at the age 61 years for males while for females C_{if} vary from 135.12 at the age 42 to 67.48 at the age 61 years.

The high age ratio 136.26 at age 32 for males and 135.12 at age 42 for females are invariably associated with exceptionally low age ratios observed in previous and precedent age ratios precisely 82.08 (at age 31) and 71.34 (at age 33) for males; then 71.80 (at age 41) and 85.18 (at age 43) for females. This pattern of age ratio can possibly be due to the weakness of enumerator bias in addition to misreporting in the age information they may also be the result of unforeseen fluctuation in the levels of fertility, mortality or the change in migration which is always sex selective. The high age ratios specially are occurred in the working age which is positive to migrate.

Taking into account for instance an age ratio of 89.08 for males in the age 20 means that, there are 10.93 per cent likely less than males in that particular age 20, as compared to the average number of males in age 19 and 21 years. In the same vein, an age ratio of 71.80 for females in the age 41 indicates that, there are probably 28.20 per cent less than females in age 41, as compared to the average number of females in age 40 and 42.

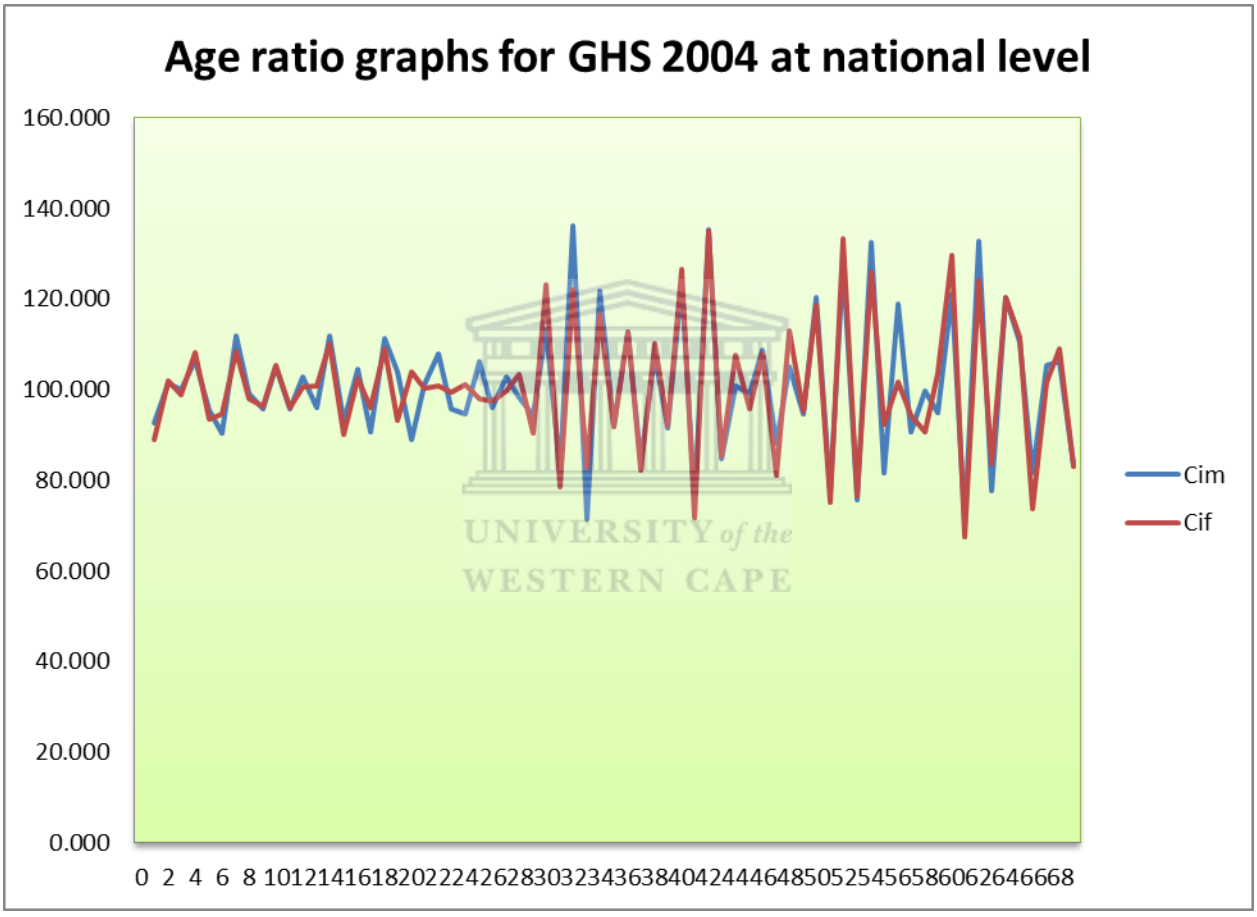
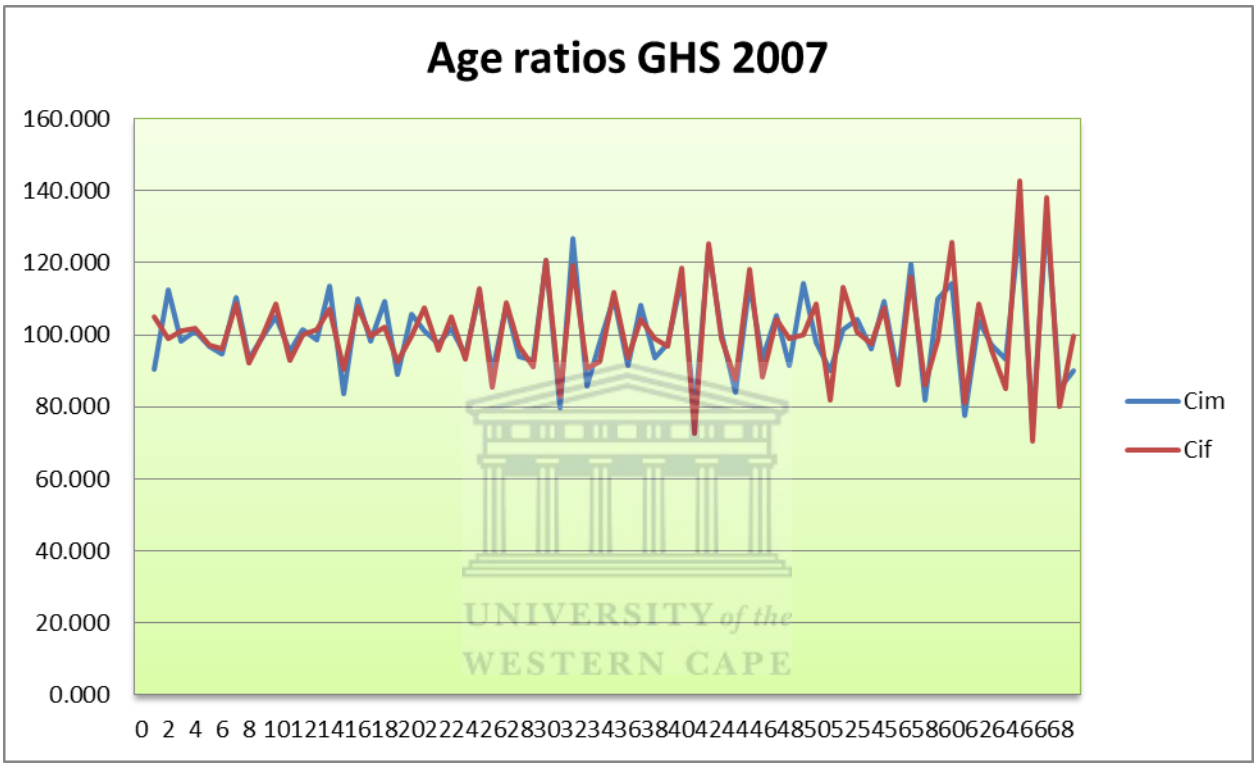


Figure 4.1 Age Ratio Graphs for GHS 2004

4.4.4.2 Description of Age Ratio Graphs for GHS 2007 at National Level

Regarding figures of age ratios for GHS 2007, both curves are probably followed the same trend as those of GHS 2004. C_{im} and C_{if} are moderately deviated from the normative point of 100. The high age ratio for males is 132.76 at age 67while the high for females is 138.27 at age 67. The low age ratio is 74.45 for males while 70.36 for females at the same age 66 years.

Considering an age ratio of 120.60 for males in the age 30 implies that, there are 20.60 per cent likely more males in that age 30, as compared to the average number of males in ages 29 and 31 years. In the same vein, an age ratio of 112.79 for females in the age 25 shows that, there are 12.79 per cent likely more females in this particular age 25, as compared to the average numbers of females in ages 24 and 26 years.



4.2 Figure Age Ratio Graphs at national level GHS 2007 by age

4.4.4.3 Comparison Trends Age Ratio Graphs of GHS 2004 and GHS 2007

The age ratio graphs depart from the general pattern shows strong fluctuations and disruption. There are strong suggestions of irregularities in the reporting of age among the respondents. The peaks are mostly observed for ages ending with 0 or 5 from age 25. Compared to GHS 2004, variations in age ratios in GHS 2007 are relatively narrow. The age ratio 132.76 has been found to be the highest for males and 138.27 for females at the same age 67 years. However, there is no clarification relative to greater fluctuations in age ratios in GHS 2004 are due to unforeseen changes in the levels of fertility, mortality and patterns of migration or may be are due to age

misreporting or differential omissions of respondents of a specific age the time of enumeration or due to the enumerator weaknesses or the combined effect of the three factors.

4.4.4.4 Description of age ratios graphs at national level for LFS 2005

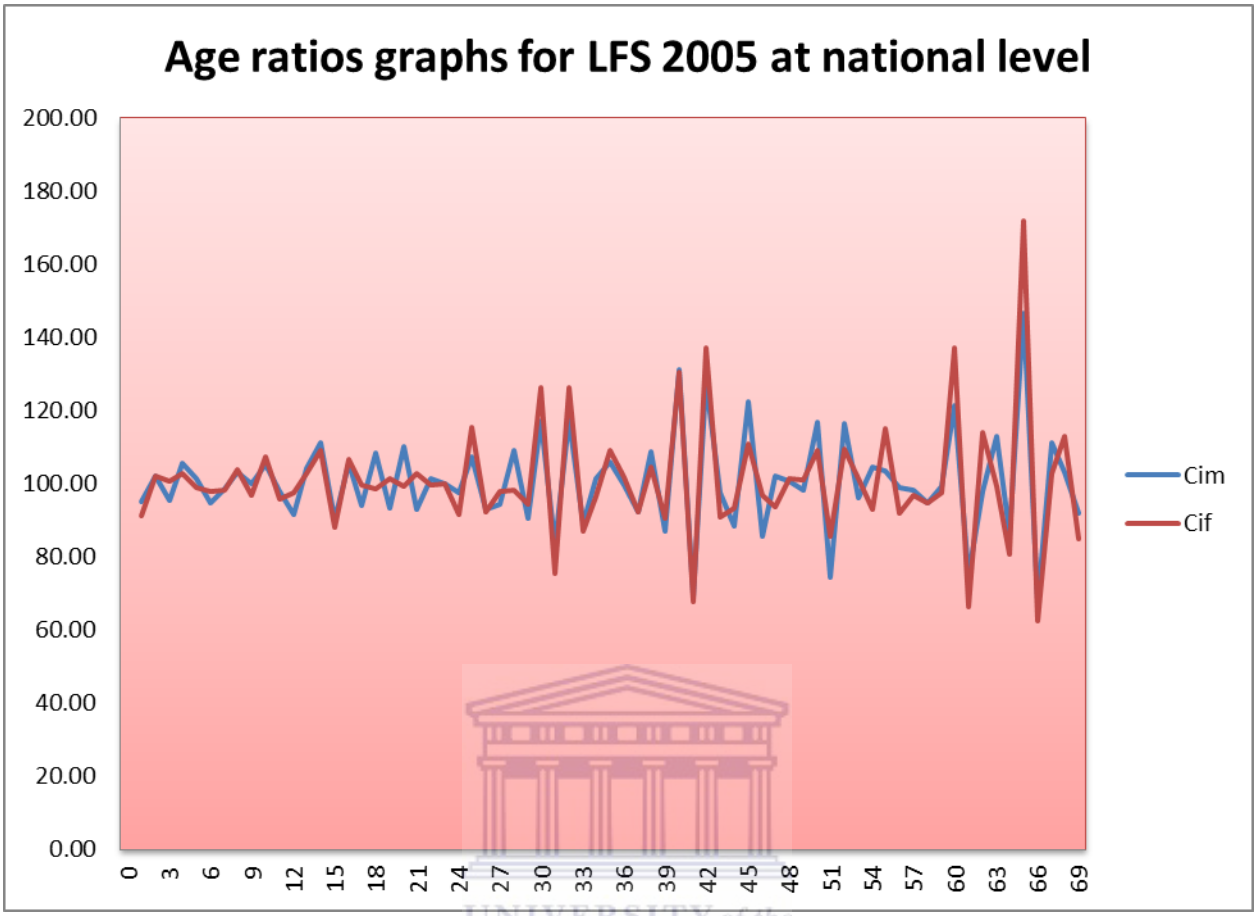
As far as LFS 2005 is concerned, the age ratio curves highlight two parts with the first moving from age 1 to more less age 29 and the second part from age 30 to the end. The first part, both curves relatively narrow from the normative but, the second have some big variations.

In LFS 2005, age ratios for males vary from a high of 146.67 at age 65 to a low of 67.23 at age 66 while for females, these age ratios vary from 171.85 at the same age to 62.48 at also the same age. As the researcher had mentioned at the beginning of this particular section, the curves of age ratios are moderately narrow from the normative. This age ratio (146.67) at age 65 reveals inconsistencies in the data structure which mean that the quality is doubtless inaccurate.

The point 131.16 of C_{im} in the age 40 years indicates that there are possibly 31.16 per cent more males in that age 40 years, as compared to the average number of males in age 39 and 41. In the same way, an age ratio of 75.42 for females in the age 31 years reveals that, there are probably 24.58 per cent less females in age 31 years, as compared to the average number of males in age 30 and 32 years.

UNIVERSITY of the
WESTERN CAPE

For instance, high age ratios at age 30 are invariably associated with exceptionally low age ratios at age 29 and 31. This pattern of age ratios may arise due to misreporting in the age information. It can also emerge due to unexpected change in the levels of mortality (epidemics, war, famine…), or for stream migration which is also sex selective.



UNIVERSITY of the
WESTERN CAPE

Figure 4.3 Age Ratios Graphs for LFS 2005

4.4.4.5 Description of Age Ratios Graphs at National Level for LFS 2007 (Chart 8)

According to LFS 2007, the age ratios curves follow the same trend despite the little difference in the reliability of results. Observing the results, the highest age ratio for males is 142.86 at age 67 while for females is 136.46 at age also 67. The lowest is 71.19 for males at age 41 while the lowest for females is 70.32 at 51 years. The gap between the highest and the lowest value is consistent close to the double of smallest age ratio for both sexes.

The variations of age ratios are moderately narrow from the normative 100. Considering an age ratio of 134.17 for males at age 65 in LFS 2007, it shows that there are 34.15 per cent more males at age 65, as compared to the average number of males at age 64 and 66 years. Consistently, high age ratios at age 65 have been estimated for the both surveys (LFS 2005 and

LFS 2007), among males and females. Additionally, high age ratio at the age 65 is automatically linked with low age ratios at age 64 and 66 years. A possible explanation to this pattern of age ratio is that it could be the misreporting in the age information. They could also be the effect of sudden changes in the levels of fertility, or mortality, or migration stream.

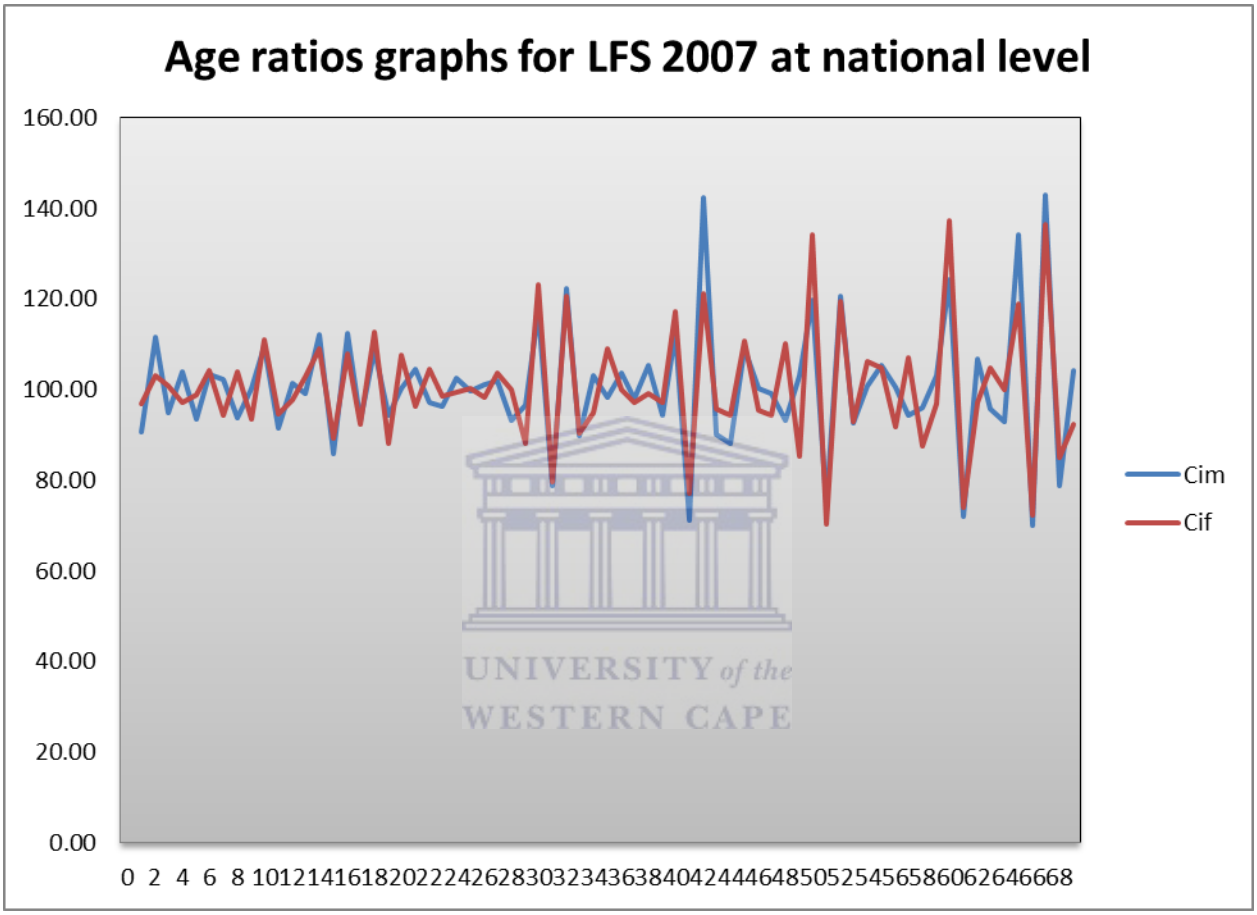


Figure 4.4 Graphs Age Ratios per year LFS 2007 at National Level

4.4.4.6 Comparison of Trends of Age Ratios Graphs between LFS 2005 and LFS 2007

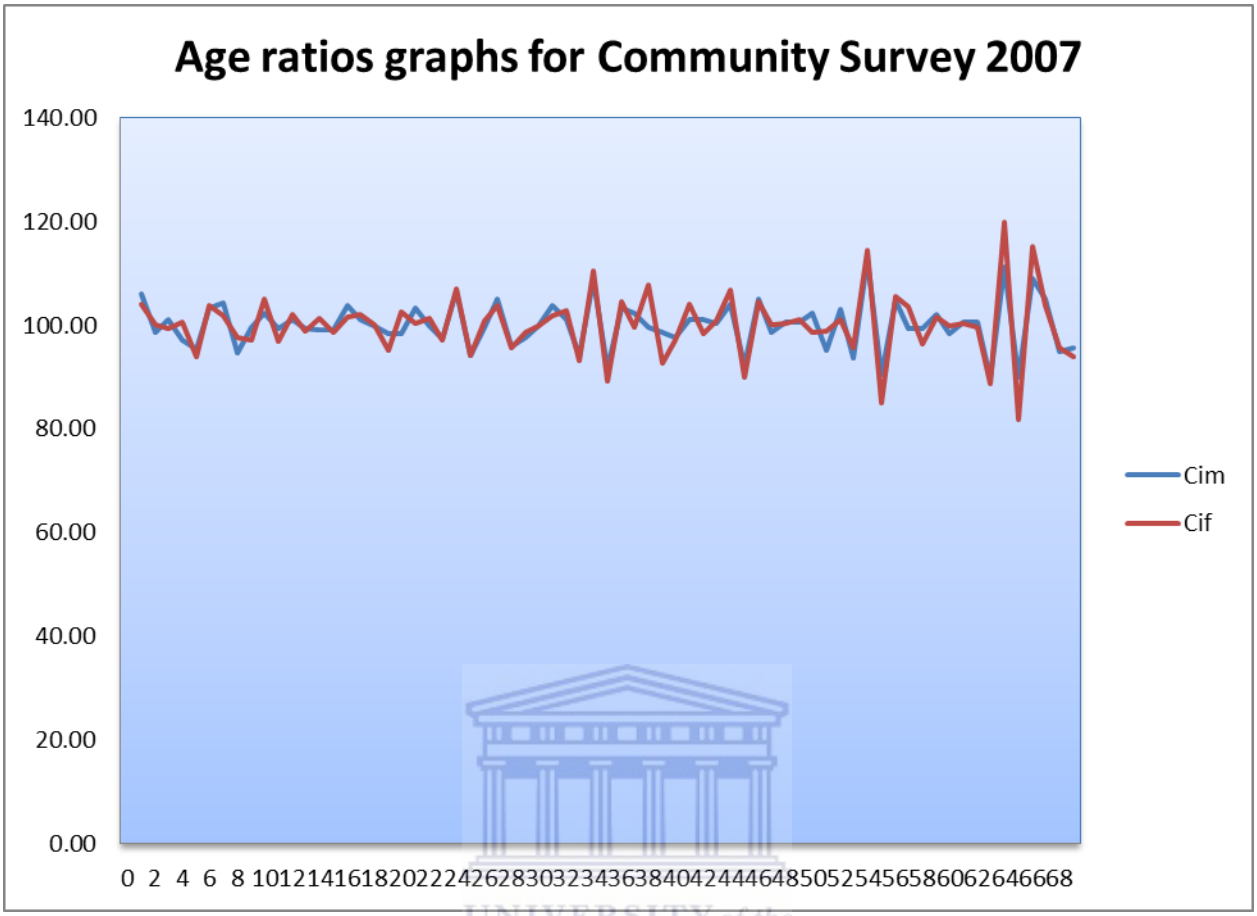
Compared to LFS 2005, variations in age ratios in LFS 2007 are relatively narrow. According to the total population, the highest sex ratios for males in LFS 2007 has been estimated to be 142.86 at age 67 whereas in females, the age ratio has been found to be highest at also the same age. There is no evidence that the variations in age ratios in LFS 2005 are due to unforeseen changes in the levels of fertility, mortality or are arisen to the differential omission of persons of a specific age at the time of enumeration. Shryock & Siegel (1976) argue that the age ratios should

not be taken as valid indicators of error for particular age. Hence the age ratio score is the mean absolute deviation of age ratios for different age groups from 100. And they are calculated separately for males and females. However, both curves of age ratio LFS are presenting the poor quality of reporting on age by respondents.

4.4.4.7 Description of Age Ratios Graphs for Community Survey 2007

Checking the Table 4.13, the age ratios per year of Community survey 2007 indicate some age misreporting errors. Age ratios of females (C_{if}) and age ratios of males (C_{im}) always deviate from the normative value 100. C_{im} vary between 89.80 at age 63 which is the smallest frequency and 112.60 at age 54 which is the highest frequency while C_{if} are ranged between 81.86 (smallest frequency) at age 65 and 119.85 (highest frequency) at age 64. The peaks on the curves indicate the most preferred age in reporting while the concaves show the under reported ages. The high age ratio 114.53 at age 54 for females, is invariably associated with exceptionally low age ratios observed in previous and precedent age ratios precisely 95.65 (at age 53) and 85.06 (at age 55). This pattern of age ratio can possibly be due to misreporting age information, or may result from unexpected change in the level of fertility, mortality and migration. However, the high age ratios are usually observed in the working age which is positive to migrate. Considering an age ratio of 108.64 for males in the age 34 years implies that there are 8.64 per cent probably more than males in that particular age 34, as compared to the average number of males in age 33 and 35 years.

UNIVERSITY of the
WESTERN CAPE



UNIVERSITY of the
WESTERN CAPE

Figure 4.5 Graphs Age Ratios by Age for Community Survey 2007

4.4.4.8 Description of Age Ratios Graphs for Census 2001

Looking at the figures of Census 2001, the curves are relatively narrowed. They always deviate from the perfect value 100. The smallest value of C_{im} is 85.29 at age 50 while the smallest value of C_{if} is 86.23 at age 60. The highest value of C_{im} is 121.38 at age 61 while the highest value of C_{if} is 130.23 at age 61. According to the curves of age ratios, some peaks and trough arise on the graph to set out some inconsistencies. For instance, 111.94 (at age 31 years) and 113.09 (at age 51 years) are the peaks of C_{im} while 88.35 (at age 30 years) and 89.83 (at age 34 years) represent some troughs of C_{if}. The high age ratio 108.78 at age 29 for males, is invariably associated with exceptionally low age ratios observed in previous and precedent age ratios precisely 96.18 (at age 28) and 88.82 (at age 30). This pattern arises in the case of misreporting in the age information, or may be due to unforeseen variations in level of migration, mortality and fertility. An age ratio of 88.35 for females in the age 30 years indicates that there are 11.18 per cent

possibly less than females in that particular age 30, as compared to the average number of females in age 29 31 years.

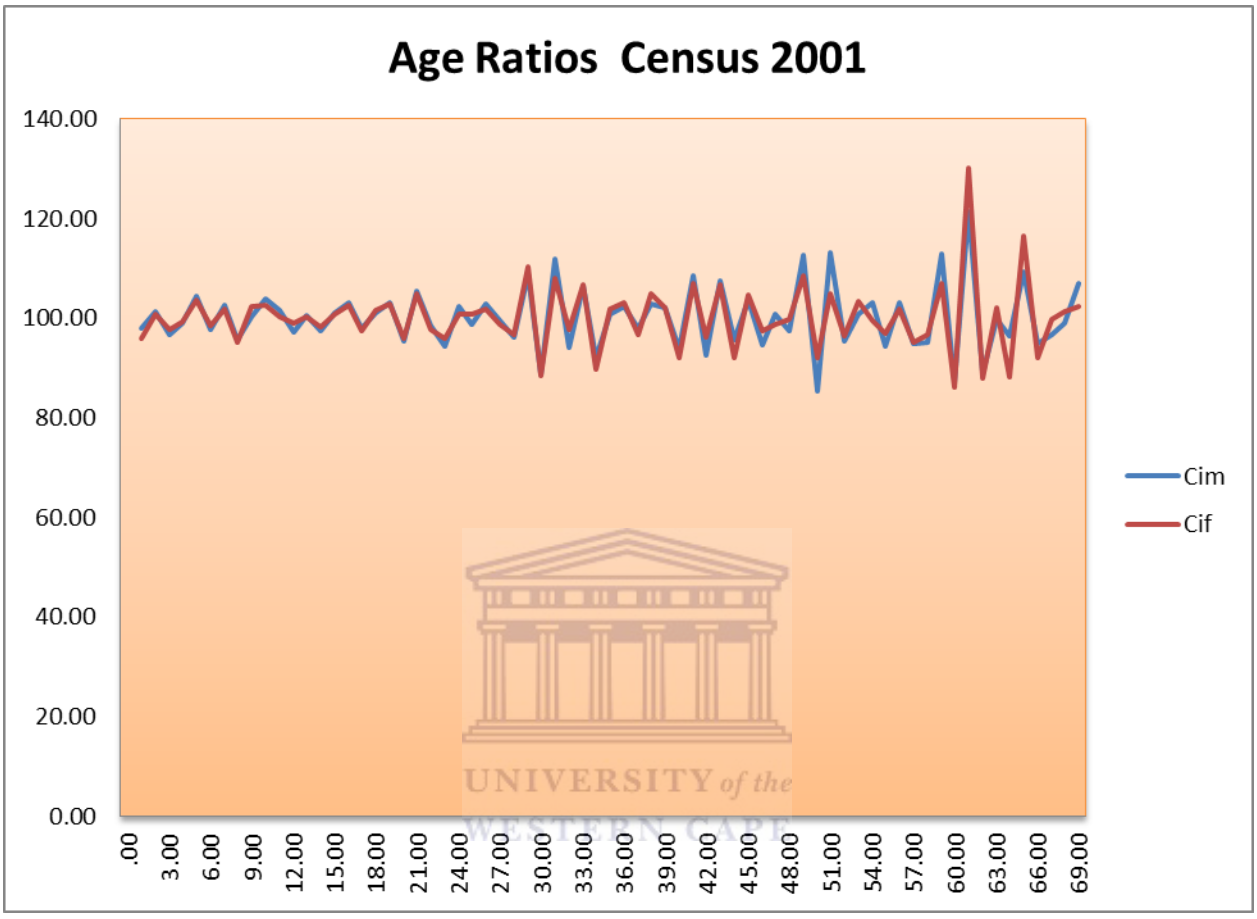


Figure 4.6 Graphs Age Ratios by Age for Census 2001

4.4.5 Results Sex Ratios Analysis

Over last decades, sub-Saharan countries had faced an increase of sex ratio at birth above its normal level of 105-106 boys for every 100 girls. A sex ratio score can be regarded as in the high range where given sex data are suspected as notably affected by errors other than actual fluctuations in population trends. It is equal to the male population in a given age or age group divided by the female population in the same age or age group times 100.

SR = $100*(5P_m/5P_f)$ or SR = $100*(P_{mx}/P_{fx})$

Sex ratio in South Africa has increased slightly, from 2001 to 2007. According to Census, the sex ratio is below 100 showing slightly more females than males in South Africa. This indicates the preponderance of females over males across the country in 2001 and the opposite in 2004, 2005, and in 2007.

The pattern of sex ratio cannot be only attributed to errors in the data but are also influenced by sex selective migration too. In this situation, selective migration from South Africa to neighbouring countries could have been in favour of males. This could have reduced the sex ratios in the country and increased sex ratios in the receptive countries. However the migration drift being observed in South Africa in the recent past could have reduced, the overall sex ratios have increased in the country in the absence of significant fluctuation in births, deaths and migrations. The sex ratios are expected to be high at infant ages because the sex ratio at birth is positive to males. After early childhood, the ratios are expected to decline continuously to reach very low levels of the highest ages when females' mortality rate is much better than the males' mortality rate.

According to the whishes expressed by Arriaga in the previous section, the sex data must be examined graphically in this study.

UNIVERSITY of the
WESTERN CAPE

4.4.5.1 Interpretation of Sex Ratios Curves for GHS 2004 and GHS 2007 by Age at National Level

The Figure 4.6 shows the curves of sex ratios by age for the whole population in South Africa according to GHS 2004 and GHS 2007. Usually, the sex ratio at birth is 104-106. Sex ratios decrease with age as in normal populations; males at every age have higher mortality rates than females (Shryock & Siegel, 1976).

Men tend to overestimate their ages. This practice could have shifted men into older ages while women tend to underestimate their age which could have shifted them into younger ages, hence, causing errors in age and sex data. A lower sex ratio in the ages under 15 to some extent may be considered as under enumeration of children since sex ratio is supposed to be high at such ages. In GHS 2004, the sex ratios at age 0 and 13 have observed this pattern of under enumeration whereas in GHS 2007, a lower sex ratio were observed in Ages 0, 1, 3, 4, 5, 6, 9, 10 and `13.

Chaurasia (2005) found that age ratios are primarily measures of age misreporting. However, the sex ratio at birth in GHS 2004 is less than 100, Data in Table 4.14 indicates that the sex ratio among population from age 1 to age 5 are likely regular. The values are ranged between 100 and 105 decreasing as age increases. The sex ratio among population aged fewer than 15 are relatively moderate but, at age 15, the sex ratio has the highest value 108.89.

Generally, the variations observed in the sex ratios are used to assess the quality of data. The fluctuation in mainly middle age groups depend on the level of specific sex migration in a population, special when net migration rates are large and vary by sex. If the deviation is large from 100, there is large possibility to get errors in the data.

The omission of young adult can increase if the factor of migration is increasing in a country as happened in Tribal population of Central India. In the Central India, the sex selective migration of the working age population affected the ratio of males to females in different age groups (Chaurasia, 2005). This pattern is found in the GHS 2004 with the working age population having the sex ratio likely less than 100. Observing the table, it reveals an unusual shortage of males particularly at age 20, 23, 24, 28, 29 and 30. The chart on sex ratios shows that there is concavity in these ages. This should be under reporting of males in these ages. This situation is common in all population enumeration. The sex ratios of these ages decline suddenly compared with the previous ages from 15 to 19, especially compared with the sex ratios of the same generation in the previous surveys.

It is expected that the sex ratios among the population aged 60 years and older for instance in all surveys or censuses will be low. Poston & Davis (2009) confirmed that the lower female mortality as compared to the male mortality contributes to decline the sex ratio levels around ages 50 to 70. This pattern is observed in the results of GHS 2004 with the sex ratios positive to males at age 52, 55, 59, 60, 63 and more.

Looking at age 69, the sex ratio is 58.88 indicating that probably for 59 males there are 100 females at the same age 69.

Sex ratios of population are high mostly because of their low living standards and lack of education which lead similar mortality rates between males and females. Chaurasia, (2005) argues that living conditions of women associate with social discrimination against the fair sex

place women at a disadvantage to men in most of the developing countries. The sex ratio is probably positive to females at age 7, 8, 14, 15, 16, 18 and 19. It is probably positive to males at the age 20, 23, 24, 28, 29, 30, 34 and more. All these facts indicate that the quality of age and sex distribution of GHS 2004 is not accurate.

As far as GHS 2007 is concerned, the sex ratio at birth is less than 100. It is likely higher than at age 1. The increase observed at age 2 indicates possibly decrease in the rate of fertility and also probably increase in the rate of mortality. From population aged 3 to 20, the variation of sex ratios shows a similar pattern to those for the population in 2004. At age 18, the value of sex ratio is highest 107.40.

Few men are gone out of South Africa due to Labour migration. Therefore, the Figure 4.7 is not likely to reflect this population patterns. The observed pattern is due probably to geographically mobile young men who were within South Africa being under enumerated relative to females in the same ages in the survey. The concavity appears in the sex ratios at age between 21 and 31 indicating the shortage of males. At age 61, the sex ratio is 57.27 there are 5 males for 3 Females. All these features reveal that the quality of the sex distribution is not accurate.

UNIVERSITY of the

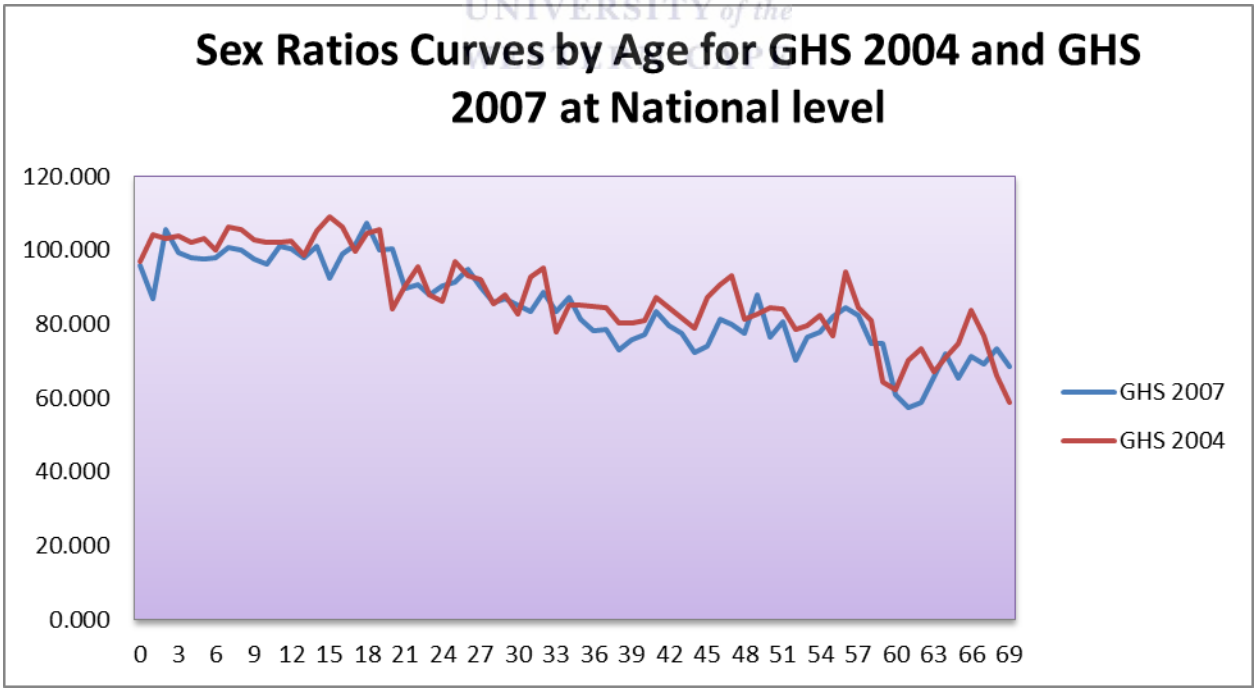


Figure 4.7 Sex Ratios Curves for GHS by Age at National level

4.4.5.2 Comparison of Trends Sex Ratios GHS 2004 and GHS 2007

Comparing both curves of sex ratios; the slight increase is observed at birth which is still less than 100, the normative. The improvement of the sex ratio at age 0 is still far from the expectation. This situation may reflect the problem of registration of vital events in many regions with predominance of rural areas. Starting from age 1 to 6, the trend has changed and the sex ratios are consistent and narrowing the normative 100. The high age ratios occurred at age 7, 8, 14 to 19 reveal the problem of preference in sons. But the absence of young males influences the results at age 17, 20 to 39 which indicate low sex ratios (less than 100). The fluctuations observed in the structure of sex ratios GHS show that the quality of data distribution of GHS is not accurate.

For the old ages, the observation done in GHS 2007 show that for instance at age 52, 53, 54, 58, 59 and more the sex ratios are positive for males because every male can probably have at least one female. There are more females than males.

4.4.5.3 Interpretation of Sex Ratios Curves by Age at National level for LFS

UNIVERSITY of the
WESTERN CAPE

The analysis of age specific sex ratio for LFS 2005 reveals that for the youngest ages 0, 1, 2, 4 and 5 the sex ratios tend to be range between 97.09 and 107.03. LFS 2005 has observed 104.26 at birth; this is due to biological factors which result in an excess of male births. The abrupt change occurs at age 1 where the sex ratio increases to 107.03 which are abnormal. This amount is the highest sex ratio for this instrument. A possible explanation could be that some parents have preference for male children over female children, or may be errors due to selective under enumeration of given sex as well as errors due to enumerator bias.

A lower sex ratio in the age 3, 6 and 12 may suggest under enumeration of children since sex ratio is supposed to be high at such ages. The high male mortality is possibly explanation to this purpose.

The propensity by women to under estimate their ages could have shifted them into younger ages hence, causing errors in age and sex data. At age 9, 14, 15 and 18, the sex ratios are high; probably there is probably errors in the data or, may be the fertility rate has increased.

Looking at the table of results, at age 21, 22, 25, 26, 27, 29, 31 and 32, there is an unusual shortage of males. This means that if there are no fluctuations in fertility and mortality may be it is the result of big migration which is often sex-selective. The graph shows at these ages the concavity.

According to LFS 2005, the sex ratios are invariably positive for females at age 14, 15, 18 and more. For instance, at age 1, there are 100 females for 107 males. In the same vein, at age 14, there are 100 females for 106 males; every female has a chance to have at least one male.

The sex ratios were also positive for males at age 58, 59, 60, 62, 65 and more. For example, at age 58, there were 77 males for 100 females; 62 males for 100 females at age 65.

As far as the LFS 2007 is concerned, the sex ratios among the youngest aged 0 to 5 are flexible. They are ranged between 95.19 and 105.66. At age 0, it is 97.69 less than 100 showing the decrease in both rates fertility and mortality, for ages 2 and 4 the sex ratio is increased showing excess of males in the population of these ages; this may be possibly errors in the enumeration reports or may be the issue of preference for sons over girls or may be errors due to weakness of enumerators. The contrary is observed at ages 3 and 5, where there are excess of females in that population, this may be due to the biological factors (boys are likely vulnerable than girls).

UNIVERSITY of the
WESTERN CAPE

Though at age 0, 1, 3, 5, 8, 11 and 13, the sex ratios are less than 100 showing under enumeration of children since sex ratio is supposed to be high at such ages. The affinity of women tends to round down ages could have transferred them into younger ages and in the same vein, males tend likely to overestimate their age which could have shifted them in older ages.

Data in Table 4.14 indicates that the sex ratios at ages 7, 9, 16, 17 and 19 are relatively high and especially at age 19 is the highest sex ratio. There are probably errors in the data, or may be the fertility rate has increased.

Observing the graph, it reveals a concavity in the working age especially at ages 22, 23, 25, 26, 27, 28, 29, 30, 31, 32 and 33; which is an unusual shortage of males in these ages. This situation could be explained by these males who migrated in neighboring countries in search of better life or better employment opportunity which resulted to stream migration, knowing that migration is

almost sex-selective; or there is fluctuation in fertility and mortality rates, or probably errors in the reporting of data.

Looking at data in table of preparation, the sex ratios are regularly positive for females at age 4, 7, 9, 16, 17, 19 and more. For example, at age 19 there are 108 males for 100 females at the same age; at age 9, there are 105 males for 100 females.

The sex ratios are also frequently positive for males at ages 22, 27, 28, 30, 31, 41, 44 and more. For instance, at age 41 there are 100 females for 73 males; at age 30, there are 100 females for 82 males.

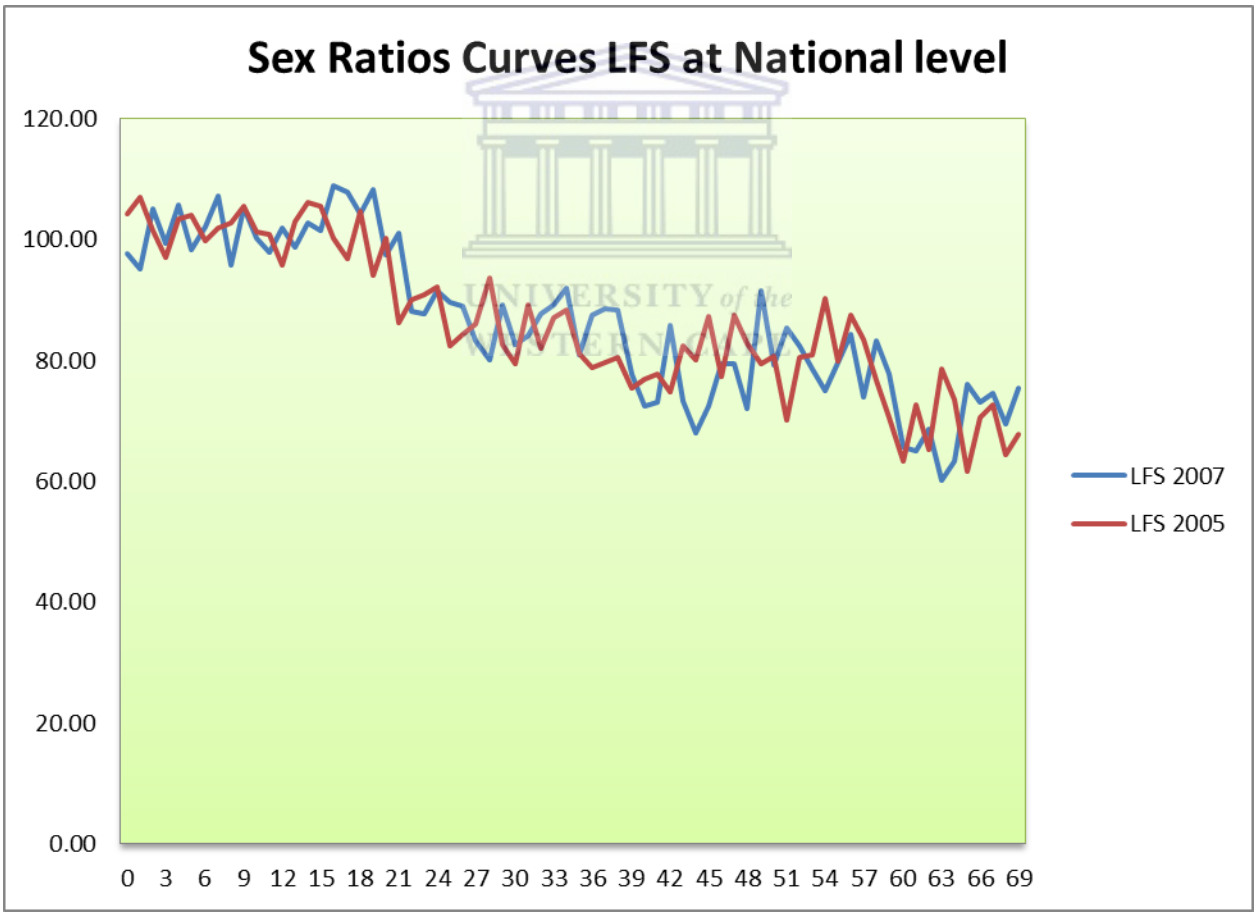


Figure 4.8 Sex Ratios Curves of LFS by Age at National level

4.4.5.4 Comparison Sex Ratios LFS 2005 and LFS 2007

Observing both records of LFS 2005 and LFS 2007, the sex ratio at birth has decreased from 104.26 to 97.69 (6.57 points) and also at age 1, the sex ratio presents the same patterns, it deteriorates from 107.03 to 95.19 (11.84 points). However, the two data follow the same trend in some ages and differ in other ages. This is the result of fluctuations observed mostly in fertility, mortality rate and migration stream of the reference population. Normally, after the sex ratio at birth, the curve of sex ratio must be decrease as the age increasing. But, the LFS 2007 presents some features different to LFS 2005 as seen in the graph.

According to the concavity in the working age, the two curves have some intercession points as the figure described. These points represent age 22, 25, 26, 27, 29, 31 and 32. The highest sex ratio observed at age 1 with value 107.03 for LFS 2005 is probably fewer compared to the highest sex ratio observed at age 16 with value 108.97 for LFS 2007.

Looking at the table of results, the sex ratio is regularly positive for females at age 4 using the two instruments but the LFS 2005 looks much better than the LFS 2007 when both probabilities are compared (105 is greater than 103 males). In the same vein, sex ratio is regularly positive for males at age 22 using the two instruments but the 2005 LFS looks much better than the 2007 LFS when both probabilities are compared (88 is less than 90 males); 90 is more close to 100 than 88. Despite the fact that both instruments present many features in their sex ratios distribution, the quality of its data structure LFS 2005 is better compared to LFS 2007.

4.4.5.5 Interpretation Sex Ratios Curves by Age Community survey 2007 and Census 2001 at National level

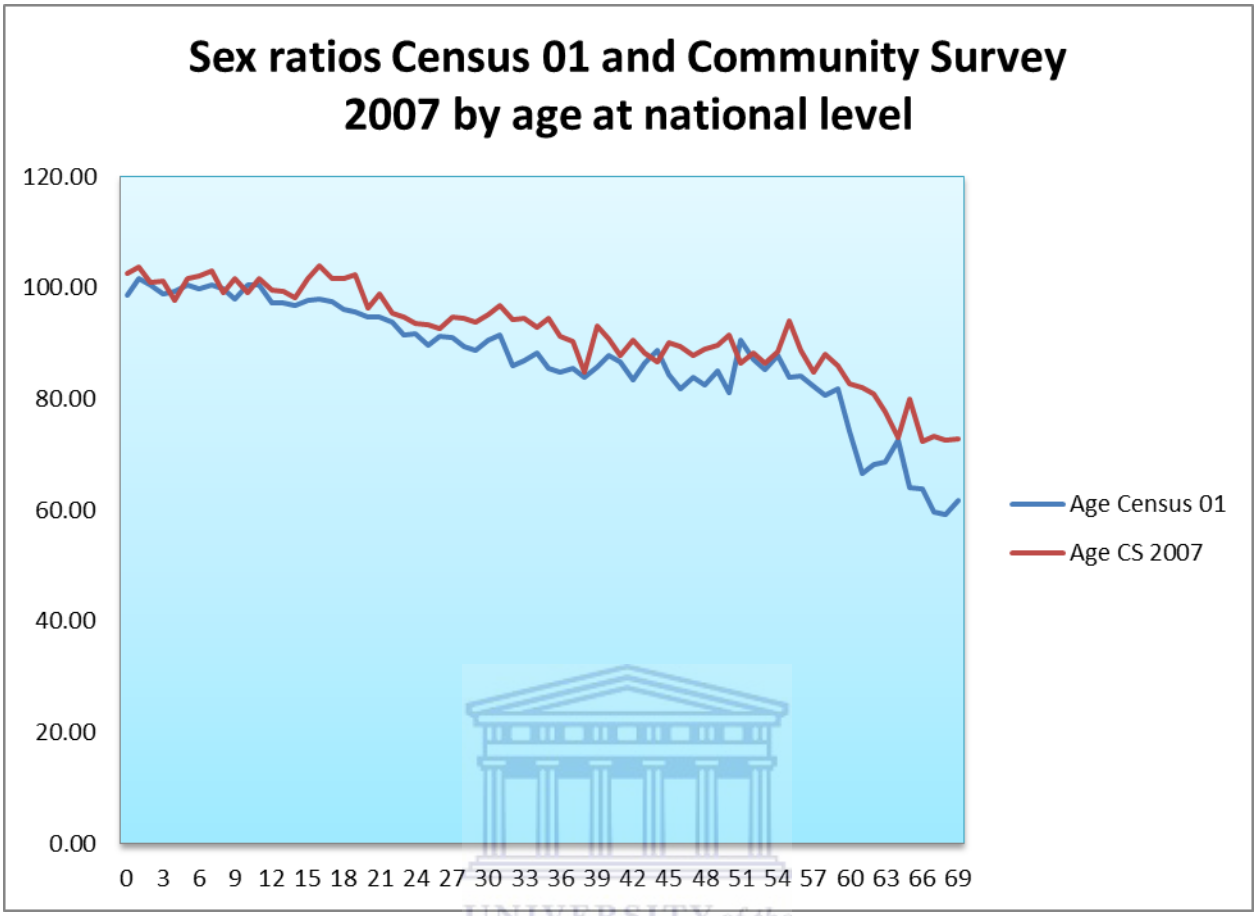
The omissions, age misreporting and out-migration might be detected by looking at the pattern of sex ratios. An analysis of age specific sex ratios for Census 2001 reveals a deficit of males in the age 0, 3, 4, 6, 8, 9 and 12 to 14 years whereas that for Community survey 2007, reveals a deficit of males in only ages 4, 8, 10 and 12 to 14. It also shows for 2001 a deficit of females in the ages 1, 2, 5, 7, 10 and 11 while for 2007, a deficit of females represents 0, 1, 2, 3, 5, 6, 7, 9 and 11. There are many possible factors that may explain this situation.

Men tend to overestimate their ages. This practice could have shifted men into older ages while women tend to underestimate their age which could have shifted them into younger ages, hence, causing errors in age and sex data. A lower sex ratio in the ages under 15 may suggest under enumeration of children since sex ratio is supposed to be high at such ages. The sex ratios in 2001 census at age 0, 3, 4, 6, 8, 9, 12, 13 and 14 observed this pattern of under enumeration whereas in Community survey a lower sex ratio were observed in Ages 4, 8, 10, 12, 13 and 14. The pattern of sex ratio seems to be consistent with the 2001, 2007 because the cohorts which had lower sex ratios 6 years before, still have lower sex ratios 6 years later.

Looking at the curves of age specific sex ratios, some features are discovered: there are peaks showing high points in the graphic and concaves indicating low points of the graphic. Considering the curve of Census 2001, for example the points 101.56 (at age 1), 100.60 (at age 7) and 100.54 (10) are peaks while the points 97.23 (12), 91.42 (23) and 83.32 (42) represent troughs. In the same order, the curve of Community survey is highlighting for instance the points 103.84 (at age 1), 102.97 (at age 7) and 103.92 (at age 16) as peaks while 97.73 (at age 4), 99.06 (at age 8) and 84.8 (at age 38) are concaves.

The highest sex ratio for census 2001 is 101.56 (at age 1) and the lowest is 59.31 (at age 68) while the highest sex ratio for Community survey 2007 is 103.92 (at age 16) and the lowest is 72.36 (at age 66). Applying the census 2001, the age specific sex ratios are relatively positive for males at age 1, 2, 5, 7, 10 and 11 although they are comparatively positive for females at ages 0, 3, 4, 6, 8, 10, 12 to 69. Using the Community survey 2007, the age specific sex ratios are moderately positive for males at ages 0, 1, 2, 3, 5, 6, 7, 9, 11, 15 to 20 while they are relatively positive for females at ages 4, 8, 10, 12, 13, 14, 20 to 69.

Observing the concavity in the working age, the two curves present some intercession points as the figure are showing.



UNIVERSITY of the
WESTERN CAPE

Figure 4.9 Sex Ratios Curves by Age for Community Survey 2007 and Census 2001

4.4.6 Population Pyramids

The best method used to produce population pyramid was cross tabulation between age and sex. Indeed, it highlighted some changes in the age and sex structure of the population. It can be observed from each figure as they are prepared in Table 4.2.3.79 of appendices.

4.4.6.1 Interpretation Age-Sex Structure of Population by Age using GHS 2004 at National level

Figure 4.10 in appendices is illustrated the age-sex structure by single age of the total population of South African enumerated in GHS 2004. This graph is typical of a developing country where birth rates are high but conditions are harsh and life expectancy is short. It has a broad base

shape indicative of a relatively greater proportion of younger age and a steadily decreasing proportion of older people. For males aged Under 5 there are slightly greater than females. The age 14 has the largest frequency of males and also for females. Population under age of 5 consists of 8.89 percent of the total population while a below the age of 15 years constitutes 30.75 percent of the total population. The pattern of age-sex distribution in the country changes across ages as Figure 4.10 is presenting. This age and sex structure is showing greater influence of the younger age on the overall population of the country. The proportion of elderly (aged 65 and above) is only 5.78 percent while the proportion of the population aged 15 and 65 or the working age represents 63.47 percent of the total population.

Looking at the Table 4.17, the dependency ratio of young females is 44.99 per cent, at the same time as 52.38 per cent for males and the total dependency ratio for young is 48.46 per cent. The elderly dependency ratio for males represents 7.09 per cent, 10.89 per cent for females and 9.11 per cent as far as the old dependency ratio is concerned. The total dependency for males is 59.47, and for females is 55.88 per cent and the proportion is 57.57 per cent.

The cohort age 0 is likely wider than the cohort age 1 and the band age 2 is probable larger than the band age 1 because of the steady increase in the number of births in the past 3 years. However, this did not mean increasing fertility rates but rather an increase in the difference of people in reproductive age. Because of rapidly declining fertility in previous years, is considerably narrower than the band for the age 4. The highest proportion of the population is the band at the age 14 years. From the band 19 through the band 69, the pyramid narrows swiftly and steadily, suggesting the high fertility rate before the age 42. The graph for South Africa continues to narrow in the old age but not as rapid as in the younger ages. In regard to all these features in the distribution of age and sex data on age, the quality of data structure GHS 2004 is inaccurate.

4.4.6.2 Interpretation Age-Sex Structure by Age using GHS 2007 at National level

According to GHS 2007, the Graph 4.11 in appendices reflects mainly characteristics of developing countries because the health standards of living are low and many children die at an early age. The sex ratio at birth is 96.023 less than 100. For males Under 5, there are slightly

lesser than females at the same age. The population under 5 consist of 9.8 per cent of the total population while a below the age of 15 years constitutes 31.94 per cent of the total population. The dependency ratio of elderly (aged 65 and above) is only 9.46 per cent while the dependency ratio of under 15 represents 51.37 per cent. The total dependency ratio which is the sum of both previous corresponds at 60.83 per cent while the working age represents only 39.17 per cent. Consequently, the population of South Africa was young in general. The age 14 has the highest frequency of males (129) and also for females (127). The increase observed in the dependency ratio could be explained by the scarcity of jobs; the creation of employment is not followed the same trend with the growth of the population.

The 0 cohort is larger than the 1 cohort and the 2 cohort is also wider than the 1 cohort. This could be errors arose in the reporting due to the weakness of enumerators, or may be the increase observed in the number of births, or probably it could be an increase of mortality rate observed at age 1.

The population of males aged from 21 to 35 is revealing some missing in those ages. This abrupt fluctuation could be seen as a pattern of migration which is affecting the structure of the population of South Africa. The contrary phenomenon appears at age 45, 49, 50 till 60 where the migrants return home to prepare the retirement period. The top of pyramid indicates old people where there are more females than males due to female longer life expectancy. These discrepancies in the structure of the population of GHS 2007 revealed that the quality of data on age and sex structure of this population is not accurate.

Comparing the two GHS, the population aged 0 to 14 has increased remarkably, from 14 939 in 2004 to 17 540 in 2007. Within the three years from 2004 to 2007, the number of children aged 0 to 14 increased by almost 2 601, from 14 939 in 2004 to 17 540 in 2007. At the same time, the population in working age increased apparently from 30 830 in 2004 to 34 145 in 2007 but in terms of percentage observed while considering the total dependency ratio, it is decreased from 42.43 per cent in 2004 to 39.17 per cent in 2007. The proportion increased by only 0.35 per cent from 9.11 per cent in 2004 to 9.46 per cent in 2007. Because the population aged 65 and older and the proportion aged 15 to 64 increased and with different speed, the age dependency ratio is relatively increased with high speed. Because of the slight increase in fertility rates, a very slight increase in life expectancy is observed. Generally, the child dependency ratio of the national

population mostly depends on changes in fertility. It also depended on migration. Country with large in-migration rates tend to have low child dependency ratio and vice versa. Obviously, to the child dependency ratio, aged dependency ratio is not directly correlated with fertility but rather on mortality and migration. As most of South African migrants are in working ages, high positive net migration contributes to the low aged dependency ratio.

4.4.6.3 Interpretation Age-Sex Structure by Age using LFS 2005 at National level

As far as the LFS 2005 is concerned, the Figure 4.12 in appendices illustrates an age and sex structure with a very large proportion of children, a very small proportion of elderly persons and a low median age of South Africa population at national level. The sex ratio at birth is 104.26. Males under 5 have greater proportion than females at the same ages. The population under 5 represents 9.31 per cent of the total population while the population under 15 represents 32.55 per cent of the total population. This case reflects the poor quality of the health standards of living. It also shows the lack of birth control with the high fertility rate. The young dependency ratio is 52.66 per cent, for the elderly dependency ratio is 9.1 per cent; the total dependency ratio indicates 61.77 per cent while the working age represents only 38.23 per cent. Although, deaths are usually concentrated among young children and old ages, there is disproportionately less number of young adults who represent in majority the working age. The population aged 14 indicates the highest frequency of males (138) and also for females (130).

UNIVERSITY of the
WESTERN CAPE

The band at age 0 is greater than the 1 cohort and the population at age 2 is wider than the 1 cohort. The possible explanation is that, this case may be resulted from the increase in the number of births or may be the increase in the rate of mortality at age under 5. In addition, the decline in the fertility rate of the total population in the main reproductive ages 25 to 30 years in the previous years for instance in 2000 contributed to the decline in total births in 2005 because this category is still active and notably the cohort of children aged under 5 years of the 2005 will be also affected.

From age 20 to 35, the graph is showing the missing of males in these ages. A possible explanation is that may be this should be considered as errors in enumeration, or may be the young males emigrate in search of better education or life. The top of this pyramid (people aged

65 and above) which represents only 5.63 of total population indicates that females are more than males, due to greater life expectancy of females.

4.4.6.4 Interpretation Age-Sex Structure by Age using LFS 2007 at National Level

Considering the Figure 4.13 (Appendices), it reflects the typical characteristics for developing countries with an expansive pyramid. The life conditions are poor and children die at an early age. The population under 5 represents 9.54 per cent of the total population while the population under 15 years old constitutes 31.94 per cent of the total population. The sex ratio at birth is 97.69. For males under 5, there are slightly less females. The dependency ratio of young represents 51.35 per cent while the dependency ratio of old ages is 9.44 per cent. The total of dependency ratio represents 60.79 per cent while the working age indicates 39.21 per cent. The age 14 has the highest frequency of males (133) and also for females (130). To a certain extent, the population of South Africa is young. Both sexes are followed the same trend despite the slight improvement occurred. The dependency ratios have decreased from one to another with little distinction.

Observing this Graph 4.13, it reveals that from age 23 to 33 and 35 there are missing for male population. This missing can be considered as the fluctuation due to migration streams. Males migrate to the neighbouring countries for several reasons. This absence affects seriously the age-sex structure. At the old ages, males are fewer than females due maybe to biological factors. All these discrepancies in the age structure indicate that the quality of age and sex distribution of LFS 2007 is likely not accurate. The preference for these digits among males may be attributed to the greater tendency to overestimate the age, whilst for females; it may be due to underestimation of their age.

Comparing both surveys LFS 2005 and 2007; the Table 4.15 shows that the proportion of the population aged 0 to 14 has decreased by only 0.61 per cent from 32.55 per cent in 2005 to 31.94 per cent in 2007. Within two years from 2005 to 2007, the number of children aged 0 to 14 decreases by almost 527 from 17 994 in 2005 to 17 467 in 2007. The population in working age decreases slightly from 34 168 in 2005 to 34 015 in 2007. The young dependency ratio decreases by 1.31 per cent from 52.66 per cent in 2005 to 51.35 per cent. The old dependency ratio increases by 0.34 per cent from 9.1 per cent in 2005 to 9.44 per cent in 2007. Because the

proportion of the population in young age decreases, the child dependency ratio reduces from 52.66 per cent in 2005 to 51.35 per cent in 2007. The population aged 65 and older increases while the proportion aged 15 to 64 decreases at different speed, the total dependency ratio also declines by 0.98 per cent from 61.77 per cent in 2005 to 60.79 per cent in 2007. Knowing that the child dependency ratio of the national population is not only a single factor to explain changes in fertility, but also mortality and migration have an impact on the fertility.

The child dependency ratio is much higher than the aged dependency ratio. The trends in the total dependency ratio are similar to trends in the child dependency ratio.

4.4.6.5 Population Pyramids Census 2001

This section is analyzed per population groups to highlight how population trends and issues arising from their declaration on age affected the well-being of people (refer to Table 4.2.3.80 in appendices). Fertility, mortality, migration and other demographic indicators play a crucial role in determining what type of opportunities and challenges people face according to their ethnic groups across the country. The observation is made on the basis of population pyramid of Africans/blacks, followed by Coloureds group, then Indians/Asians. White population group is the last to be analyzed.

UNIVERSITY of the
WESTERN CAPE

4.4.6.5.1 Pyramid of Africans/Blacks using Census 2001

The Figure 4.14 illustrates the age and sex distribution for Census 2001 of the African/Blacks population at national level. However, the pyramid of Africans/Blacks is an expansive pyramid presenting a wide base which implies a high proportion of children with a rapid growth and a low proportion of older people. A steady upwards narrowing indicate that more people die at each higher aged. This population is characterized by a high birth rate, high death rate and a short life expectancy. It is the characteristics of developing country where population has little access to use birth control, negative environmental factor; for instance lack of clean water, poor access to health care. In fact, the population under 5 represents 10.64 per cent of the total population while the under 15 represent 34.49 per cent of the total population. The sex ratio at birth is 98.62 per cent which is less than the normative. The total dependency ratio represent 65.49 per cent of working age while for the young people is 60.12 per cent of the working age

and for the old age is 5.37. The half of this population group is young and economically they depend on only less than 35 per cent of its population. Males under 5 and under 15 has greater proportion than females which is showing the problem of sex selectivity or sex preference. When observing the young dependency ratio, it reveals that the level of fertility is relatively high among this population group. The difference between the proportion of young males and females indicates that females are more exposed to the risk of death than males.

The non-access to the health services, the lack of education of parents with cultural believes also contributes to the issue of sex preference and especially in the context of African population. The pattern of gender imbalance is observed in this population group at younger age; males usually exceed females at birth, but experience different mortality rates due to many possible reasons such as differential normal death rates, deliberate gender control, calamities or epidemics, migration of parents and so on. Seeing (378) the child woman ratio for the black ethnic, it reveals the level of fertility which is probably high among this group. Considering the old dependency ratio, it is visible that the rate of mortality is likely higher among males compared to females. Women live possibly longer than males. The age 10 has the highest frequency of males (1.28) and also (1.28) for females. Observing the graph, it shows some missing for males in the working age from age 20 to 34. This missing can be considered as a pattern of migration among young males. They move out of their provinces or country in search of better conditions or opportunities in employment. The shift occurred in the age structure tend also to affect the structure of demand or needs of population itself. The effect of the pattern of employment across population groups requires a substantial change or increase in labour mobility in order to accommodate the structural change observed. Definitely all these facts observed in the age and sex structure of Africans/Blacks for census 2001 indicates that the quality of data on age is likely inaccurate.

UNIVERSITY of the
WESTERN CAPE

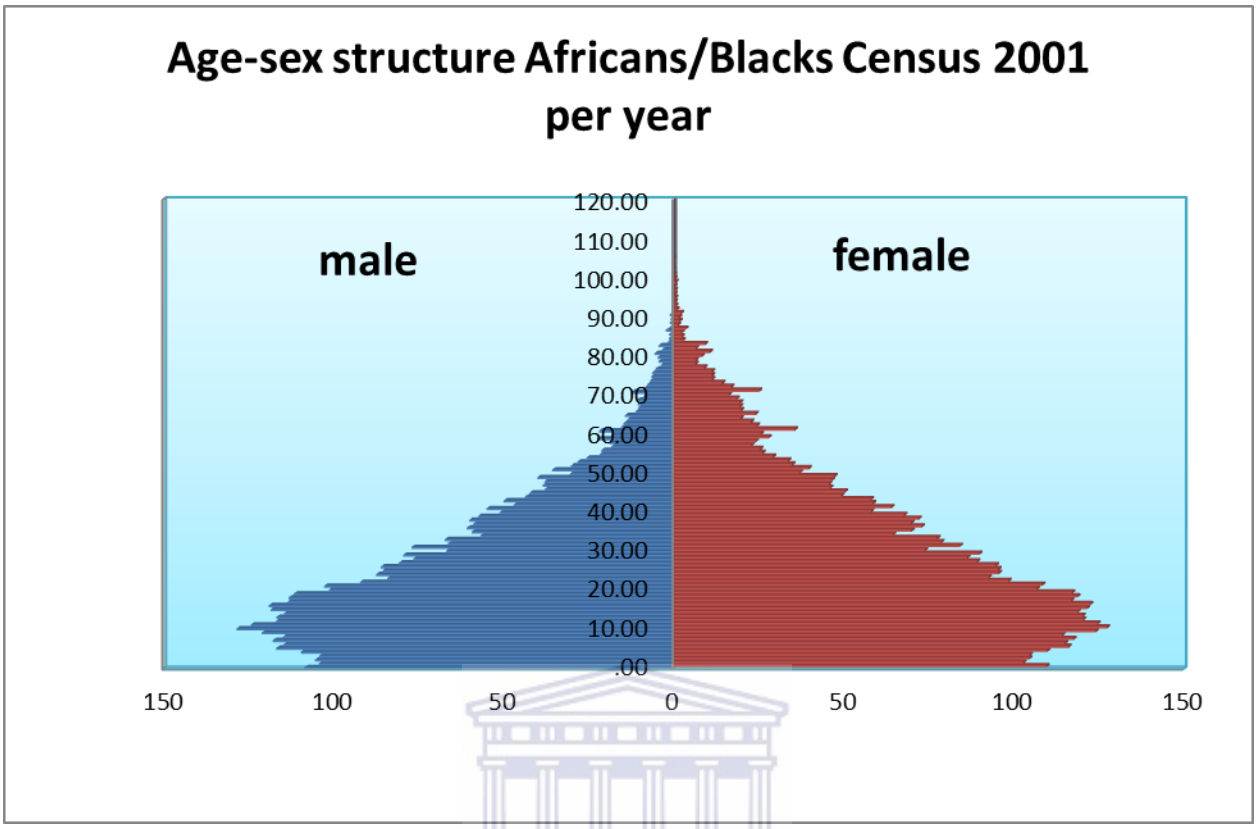


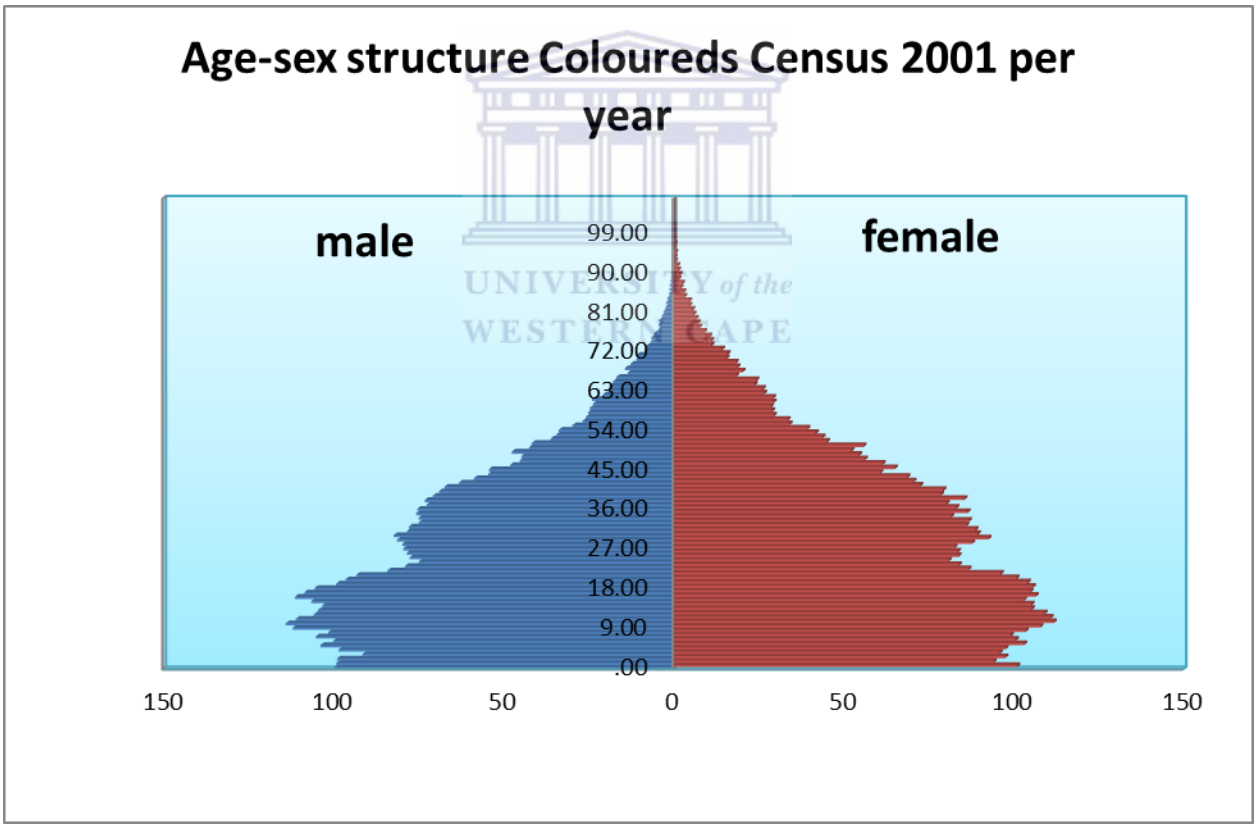
Figure 4.14 Age-Sex Distribution of African/Blacks using Census 2001

UNIVERSITY of the WESTERN CAPE

4.4.6.5.2 Interpretation Population Pyramid per year of Coloureds using Census 2001

Based on data of Table 4.11, the graph reflects the age and sex distribution of Coloured population group by single age which is a narrow shape at the top showing a low proportion of people living into old age and a high death rate. A wide base indicates a large proportion of children with less baby girls than baby boys. The age and sex structure of Coloured group is still in transition, they are moving from type of developing country to the type of developed country, but they are still on process, they have not yet reached because the proportion of young dependents is more than 50 per cent on the working age which mean this population group has a great need of support in particular the younger age. However, old age Coloureds population group represents only 5.19 per cent of the dependents males against 7.41 per cent for females. The level of fertility is still high. Children continue to out-number older persons in their contribution to the total dependency. The sex ratio at birth is 98.35 per cent revealing the pattern of preference of sons among this population group. Those under 5, the females are slightly less

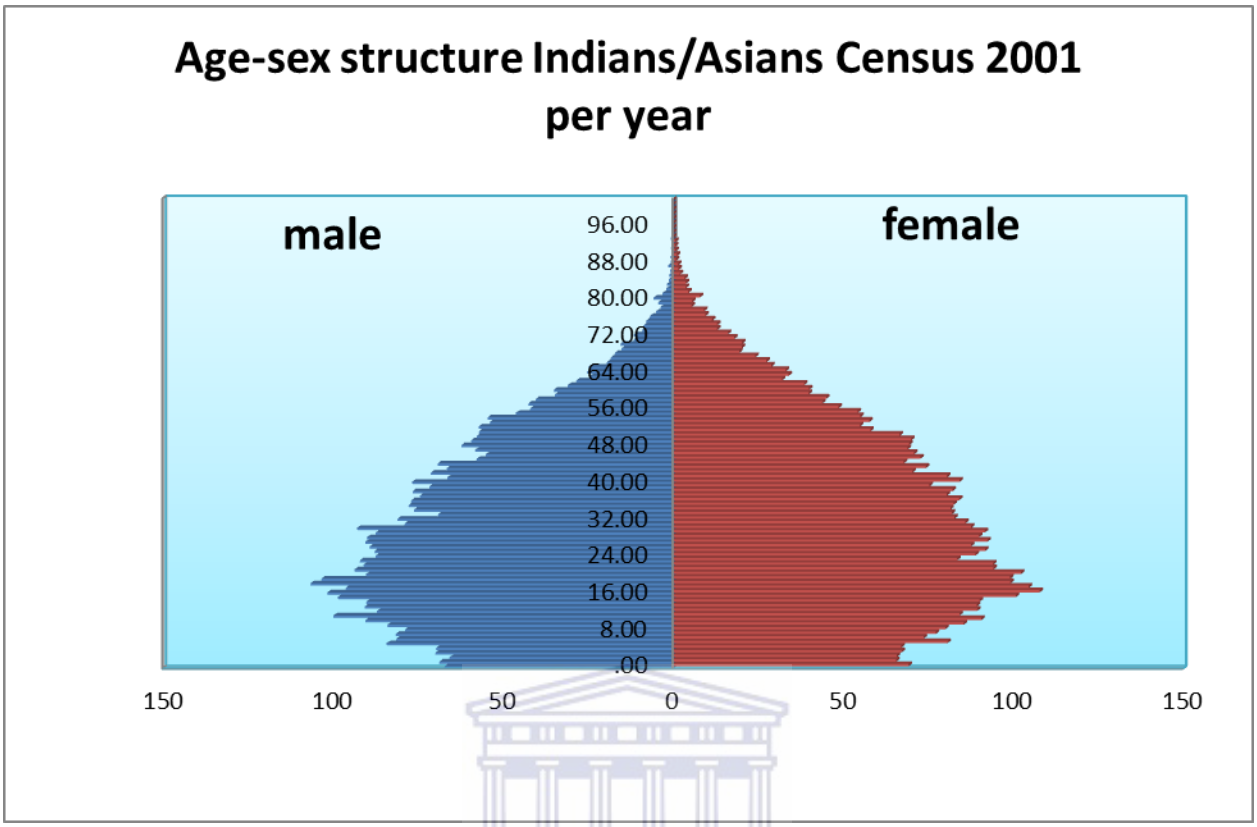
than males. The population under 5 consists of 9.74 per cent of the total population while the under 15 represents 30.91 per cent of the total population. The quality of age and sex data is likely not accurate. The age 10 has the highest frequency for males (1.14) and (1.12) for females at the same age. The shape is showing that the life expectancy of this group has increased people are living longer than before even in the working age the size has increased. There have been divergent trends for the child and old age dependency ratios in recent decades with the child ratio generally falling and that of older persons increasing. Based on the information from the figure, it is observed that males in working age 20 to 34 are missing in the distribution. This could be explained as the pattern of labour mobility within the country or abroad. In the same vein, the slight increase observed around age 50 indicates the return of labour migrants for the preparation of its retirements.



4.16 Figure Age-Sex Distribution of Coloureds using Census 2001

4.4.6.5.3 Interpretation Population Pyramid by Age Indians/Asians using Census 2001 Data

The Figure 4.17 illustrates the age and sex distribution of Indians/Asians population group of South Africa. Its structure is closer to the Whites population group with the narrower base, but with a low proportion of old ages. The low proportion of old ages is the first difference observed with the white population group. The sex ratio at birth is 96.63 and the child women ratio is 226 children per 1000 women of child-bearing age which are revealing some patterns. It could be established that the level of fertility is low but it seems that all births are not captured may be the issue of preference is dominating in this population group, or may be the level of mortality is high which is confirmed with the dependency ratio of old ages. However the value of the dependency ratio of old ages is much close to the Coloureds one than for the Africans/Blacks population group. This population group also faces the problem of missing males in working age. Observing the figure, it seems like the immigration is much pronounced in this population group which is growing especially at the working age, there are a concentration of people between 16 to 50 years (the density is bigger than in other ages). The out-migration is not even visible because the phenomena of in-migration are high. The female old dependency ratio is greater compared to males which could be explained by the greater life expectancy of females compared to males. All these facts indicate that the quality of age and sex structure of Indians/Asians population group is likely inaccurate.



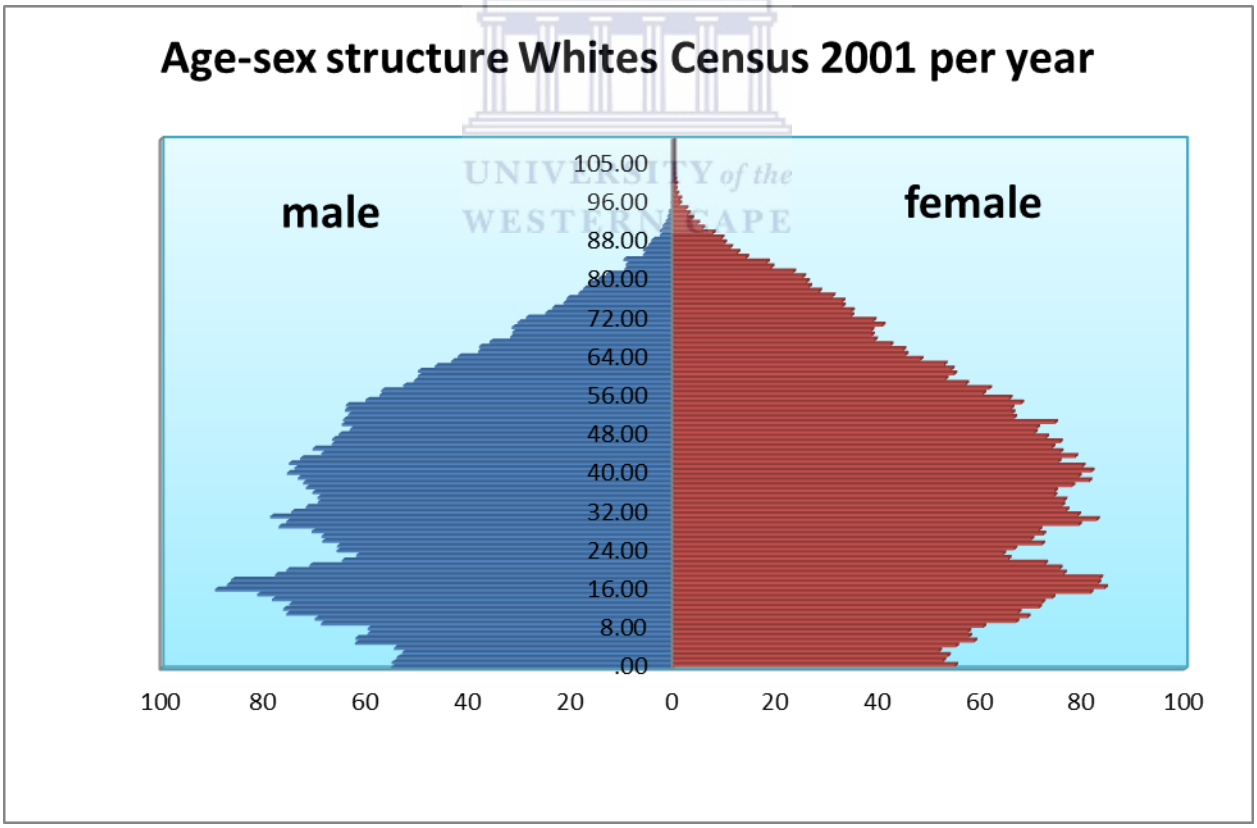
4.17 Figure Population Pyramid of Indian/Asian using Census 2001

UNIVERSITY of the WESTERN CAPE

4.4.6.5.4 Interpretation Population Pyramid per year Whites using Census 2001

The age and sex structure of the White population group is a narrower base and wider summit. Ageing is most pronounced in whites compared to the total of older persons in the country. More girls are born than boys. The sex ratio at birth is 99.46 per cent indicating lower birth. Observing the graph, the death rate is also low. The figure also shows that, at under 5 boys are slightly more boys than girls. After age 30, the rate of mortality of male is probably higher than for women which implies that women live longer in these ages (after age 30). The low fertility rate that is observed in this population group is confirmed by the child women ratio which is equal to 104 children per 1000 of women aged 15 to 49. The population under 5 represents 5.40 per cent of the total population. Looking at the table of calculation of dependency indicators, the proportion of dependency ratio of young ages is 27.24 per cent, hence for females is 28.47 and 26.08 for males. However, the total dependency ratio represents 43.14 per cent for males while females show 46.12 per cent. These ratios reveal that the population of working age is able to cover the

whole needs of the dependents. The living conditions are good quality. This group has a lot wealth with a good health environment which contributing in the health of this population group. The figure is showing a loss of young children under age 10 and also a drain brain at age between 22 and 40. This category of emigrants is generally well educated, skilled workers and productive. This emigration impacts the size of the population of this group and also would be affected negatively the economy of the country. There would be the need for a huge entry of skilled workers to fill the gap. However, due to the inequalities of the past, the education of most of the black children could not afford to solve this issue. The problem of emigration of knowledge and expertise which are not transferred to other population groups is destabilizing the entire economy of the country. Even the imported skills are often subject of constraints of deadlines, so very little skills and knowledge take place. The age 16 has the highest frequency of males (0.89) and for females (0.84). The life expectancy of this group is high for instance at age 70, males represent 0.31 while females showed 0.41; there are still more people at that stage.



4.15 Figure Population Pyramid of Whites using Census 2001

4.4.6.5.5 Comparison Inter-Ethnic Variations Population Pyramids Census 2001

Looking at the proportion of dependency ratios per population group, there are considerable differences. Africans/Blacks score the highest (63.84), follow by Coloureds (53.95), then Indians/Asians (40.08) and Whites (44.67). These differences among age and sex structure which are showing some features of each population group in particular mortality and fertility. Considering the sex ratio at birth, Whites population indicate 99.46, follow by Africans/Blacks group with 98.62, then Coloureds with 98.35 and the last is Indians/Asians with 96.63. There are showing the irregularities in the age and sex distribution of all these population groups. The percentage of old dependency ratio of White population group is confirming the living conditions of this population group. Moreover, the life expectancy of whites is probably much greater compared to all of other population groups. The dependency ratio of those under the age of 15 is indicating the level of fertility in each population group which could be confirmed just by comparing the child woman ratio of each group. Although, Whites score 104 children per 1000 women in childbearing age whilst Indians obtain 226 children, Coloureds 336 children and Africans/Blacks 378 children per 1000 women in childbearing. The clarity is that African/Blacks population group has a high level of fertility, follow by Coloureds, then Indians and the last position is White population group. The Asians population is somewhat similar to higher proportion of older women in the Africans/Blacks and Coloureds population groups. The gender imbalance could have an impact on the social and financial well-being of older women in the female population group compared to the proportions of older persons in the total population groups. There is also need to mention the high dependency ratio for old ages and children in Africans/Blacks. The Asians population group could be considered as an intermediate position between Whites and Coloureds while the Coloureds population group is in-between Africans/Blacks and Indians/Asians.

However, it is advisable to mention that the indicators value is limited as not all persons over 65 and less than 15 years require support and not all persons aged 15 to 64 years old provide support. Despite its limitations, the dependency ratio is often used as a helpful indicator of the level of potential support needs in a society.

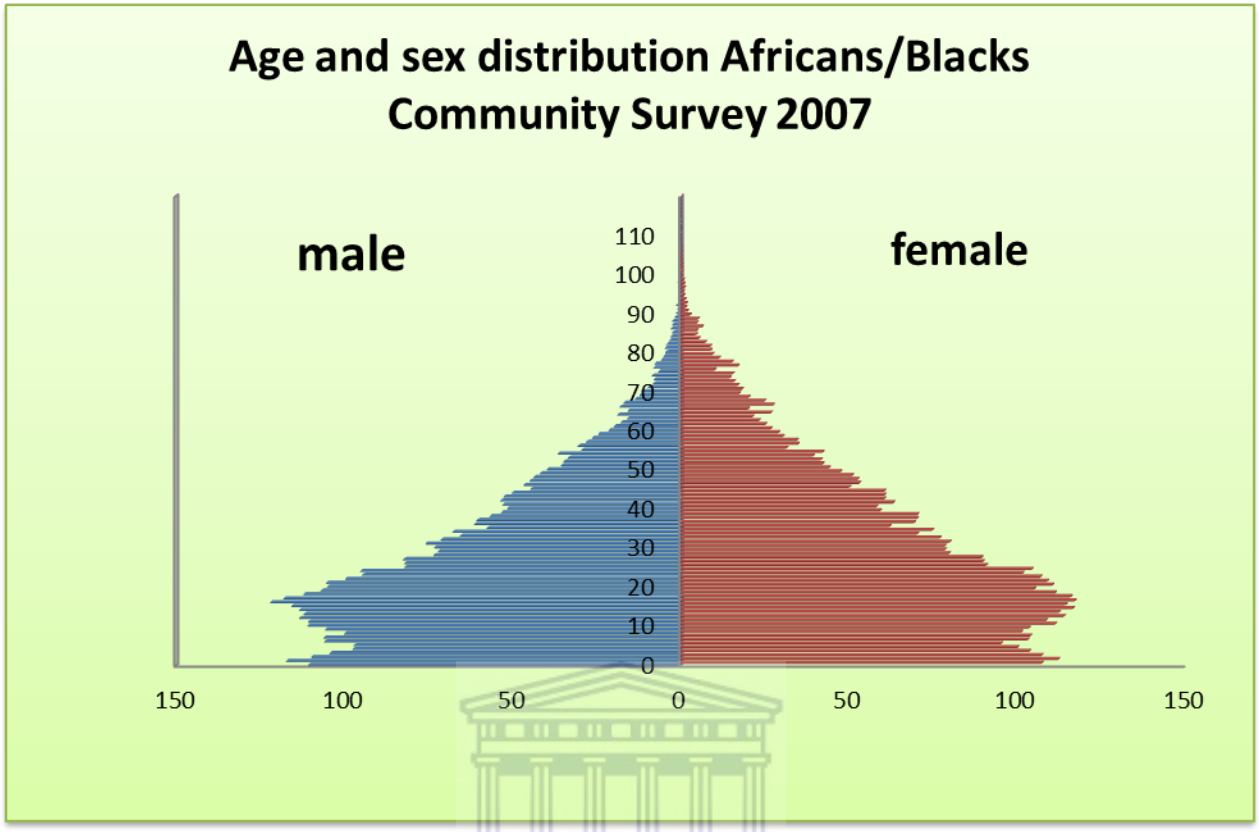
4.4.6.6 Population Pyramids Community Survey 2007

This section is emphasized for each ethnic group the changes observed, trends and issues arising from the quality of reporting their ages. These outcomes in general influenced the well-being of the population. The demographic components and some external factors indicated problems, challenges and opportunities people faced according to their ethnic group in South Africa. The analysis is made use of population pyramid of each ethnic group (African/Blacks, Coloureds, Indians/Asians and Whites).

4.4.6.6.1 Population Pyramid of African/Blacks Community Survey 2007

Considering Community Survey 2007, the Figure 4.18 reflects the age-sex distribution by single age of the total population of African/Blacks enumerated in South Africa. However, the graph presents a characteristic of developing country where the life expectancy is short and the living conditions are inconsiderable. The shape indicates a very large base of young ages and a decreasing proportion of old ages. The sex ratio at birth is 102.77 which is not reflecting the reality. This indicates that the quality of age reporting is probably inaccurate. For under 5, males are greater than females. The age 16 registered the largest frequency for males and for females. Under 15 males represent 33.64 per cent of total males while females of this age are 30.69 per cent of total females. The proportion of old ages is only 3.9 per cent of total males while female counterparts show 6.28 per cent of female population. The dependency ratio of young highlights 51.15 per cent while for the old age represents only 8.19 per cent. The total dependency ratio which is the sum of both previous equals to 59.34 per cent of the working age. However, more than half of the total working population of Blacks is dependents. This indicates the weaknesses of this population group. The variation of young dependency ratio mostly depends on changes observed in fertility rates. Migration and mortality also influence dependency ratios hence in developing country living conditions constitute a huge problem. Nevertheless, South African migrants are in working ages, an increase in net migration (positive) enhances to decrease dependency ratio.

UNIVERSITY of the WESTERN CAPE

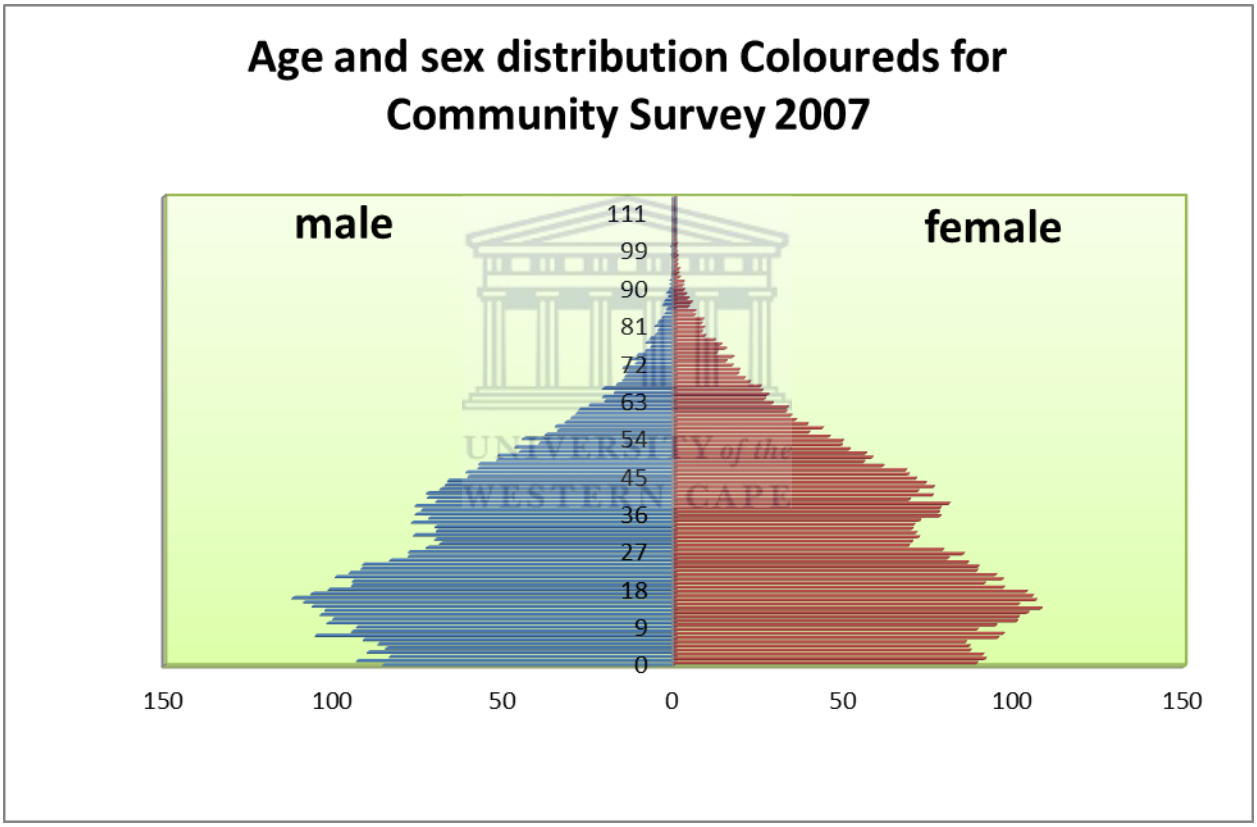


4.18 Figure Age-Sex Distribution of African/Blacks using Community Survey 2007

4.4.6.6.2 Population Pyramid of Coloureds Community Survey 2007

Looking at the Figure 4.19 it reveals the characteristics for developing countries with an expansive pyramid. The health conditions are poor; the mortality rates are high at early ages and high fertility rates. The proportion of population under 15 represents 28.43 per cent of total subgroup. The sex ratio at birth is 96.11 which increases to 101.89 at age 1. This can explain the inconsistencies in the quality of data. This population group seems to have very little access to use birth control with poor access to health care emphasize the short expectancy life of this people. Under 15, the frequency of males is close to females the difference is not significant. This can reveal some misreporting of births due to lack of existence registration offices in some areas in the country. As deaths are concentrated in the under 15 years of life, the change operates with the demographic transition maintain the high fertility rates and more of these children enter reproductive cycle of their lives; this phenomenon accelerate the growth rate. The old age males are greater compared to old female ages. This can reflects the results of HIV/Aids. However,

female population is exposed and vulnerable to sexual diseases. The total dependency ratio represents 51.17 per cent of working age while for the young people is 42.98 per cent and for the old age is 8.19 per cent. The half of this population is economically dependent on only 66.15 per cent of its population. The problem of gender imbalance is observed at younger age with frequency of males who exceed females. The age 16 has the highest frequency for males while the age 13 registers the highest for females. The missing of young males can be explained as a pattern of migration among this population in search better opportunities in employment. However, the quality of reporting age is inaccurate by Coloured respondents.

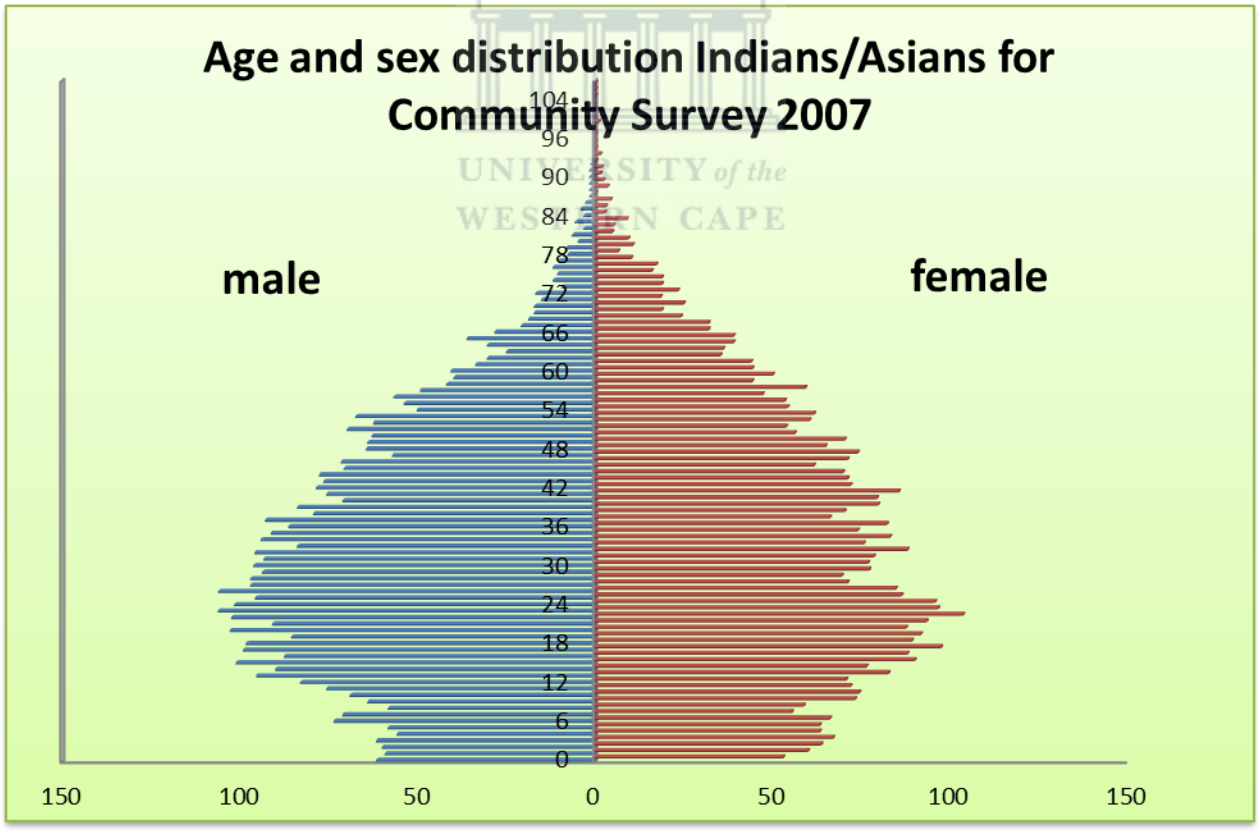


4.19 Figure Age-Sex Distribution of Coloureds using Community Survey 2007

4.4.6.6.3. Population Pyramid of Indian/Asians Community Survey 2007

As far as the Figure 4.20 is concerned, the age-sex distribution of Indians indicates that at younger ages, males are greater than females. Under age 15 represents 20.35 per cent for total males while females are 20.31 per cent for total females. Old ages constitute of 5.36 per cent of total males while females highlight only 7.47 per cent for total females. The structure of the

Indian population group has a narrow base with a low proportion of old ages. The sex ratio at birth is 115.5 showing that the level of fertility is very high. At age 1, the great decrease observed reveals the high level of infant mortality or, maybe all births are not captured. Furthermore, from age 2 to age 5, the vulnerability of male infants is highlighted. From age 11 to 18, the preference in sons occurs with high frequency of males in these ages. The young dependency ratio is 27.75 per cent while for old age dependency ratio is 8.74 per cent of the working age. The total dependency ratio of Indian population group when using Community Survey 2007 represents 36.49 per cent of total working age. The missing males in working age are also observed in this ethnic group. Referring to the graph, it reveals that the immigration is really most pronounced in this population especially in the working age. According to the results observed in table of dependency, the ratio of young male is greater compared to young female while old female ratio is greater than the ratio of old males. This reveals that the quality of age and sex structure of Indian/Asian population group is probably inaccurate.



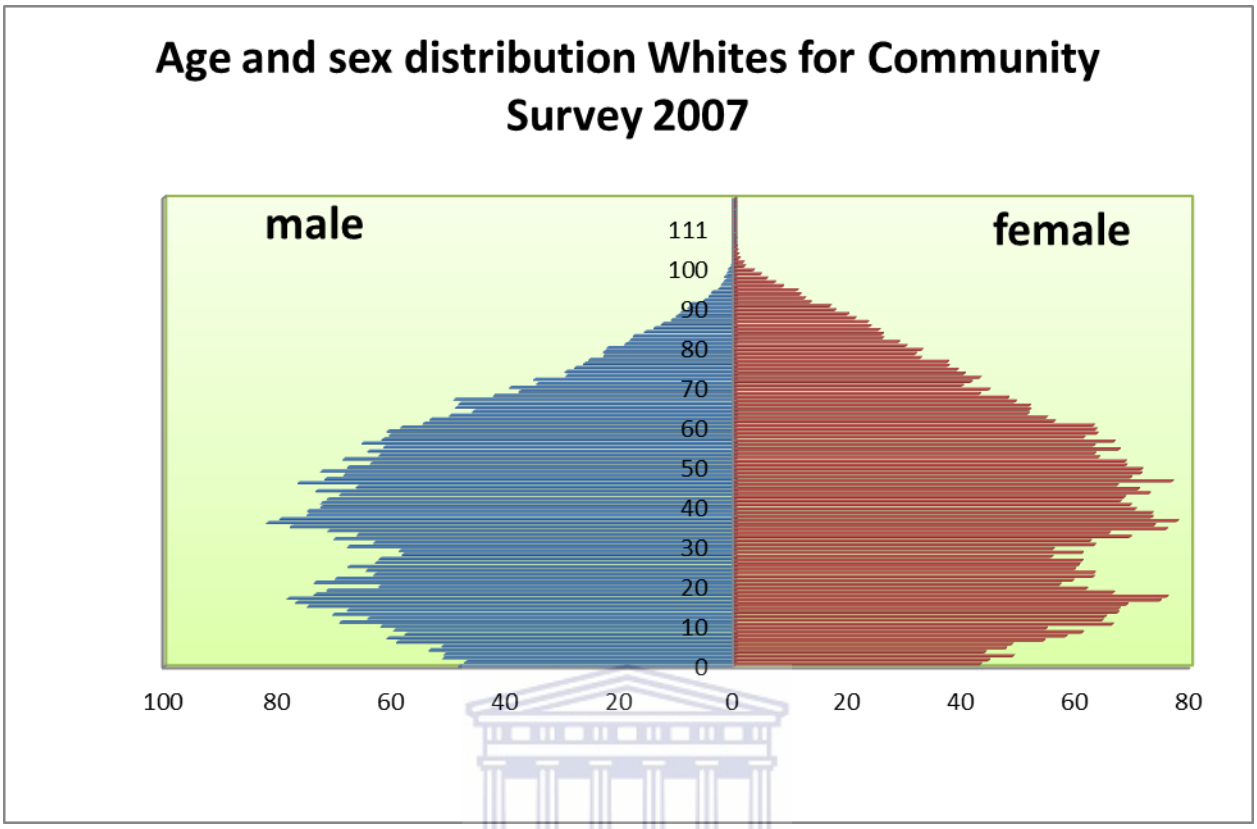
4.20 Figure Age-Sex Distribution of Indians/Asians using Community Survey 2007

UNIVERSITY of the WESTERN CAPE

4.4.6.6.4 Population Pyramid of Whites Community Survey 2007

Looking at the age-sex distribution of Whites, it indicates a steady base narrowing smoothing to the top. At young age (till under 15), females are less than males and represent 16.47 per cent while males are 17.67 per cent for total of males. However, the exception is observed at age 8, 10 and 12 where females' frequency is greater compared to males' frequency. Old ages show 14.36 per cent males while 18.12 per cent of total females. Working age indicates 67.97 per cent of male population while 64.70 per cent of female population. The sex ratio at birth is 111.93 per cent revealing a high level of fertility which decreases considerably at age 1 (105.76), then increases greatly at age 3. This fluctuation of sex ratios reveals some inconsistencies in the age-sex structure of this population group. The dependency ratio of under age 15 is 25.73 per cent while for the elderly is 25.06 per cent. However, both dependency ratios are closed and the total represents 50.79 per cent of the working age. The White population pyramid is revealing the same shape as Indian/Asian population and a missed of people observed at age 20, 22 to 34, 38 to 40. A reason could be the emigration / brain drain rather than mortality purposes especially in this population group.

UNIVERSITY of the
WESTERN CAPE



4.21 Figure Age-Sex Distribution of Whites using Community Survey 2007

4.4.6.6.5 Comparison Inter-Ethnic Variations Population Pyramids Community Survey 2007

Looking at the figures of the four population groups, Indians registers the highest sex ratio at birth (115.5), follow by Whites (111.93), then African/Blacks (102.77) and Coloureds (96.11). These differences highlight some features among age and sex structure of each population group in particular in fertility and mortality. These irregularities can be also observed in the proportion of dependency ratios per ethnic group. However, African/Black registered the highest proportion of young dependency ratio (51.15), followed by Coloureds (42.98), then Indians/Asians (27.75) and Whites (25.73). This reveals the quality and the living conditions in each population group. The lack of registration of births is visible in rural areas and the African/Blacks experience this issue because they are living in regions majoring in rural areas. Even the problem of preference in sons is dominating Indian group and also African/Blacks. According to the old age dependency ratio, Whites registered the highest proportion (25.06), followed by Indians/Asians (8.74), then Coloureds and African/Blacks show the same proportion (8.19). However, except

the Whites who registered the considerable proportion of old ages, the three other subgroups are still at the beginning process of ageing. The similarity is that the higher proportion of old females is observed in all subgroups. The gender imbalance could influence the social and financial well-being of older females in their different groups. Black and Coloured females are more involved in employment compared to males and these females have to facing issues of skills, hence they are engaged to work in clerical jobs (security, cleaner and so on…). Educational status is another challenge to consider knowing their realities. What come may, the high dependency ratio for young ages in African/Blacks and Coloureds reveal that most of respondents as in the population pyramids are in the young ages. Coloured and Indian young ages present different picture. The largest frequency is observed at age 16 for males and at age 13 for females of the Coloured population. Considering Indian/Asian, the highest frequency is observed at age 26 for males and at age 22 for females while Whites registered highest frequency at age 36 for females and at age 37 for males. Concerning African/Blacks, the highest frequency appears at age 16 for both sexes. The investigation of high mortality in the age beyond 64 years reveals that it is not as high as that of the young population; the base of the pyramid is narrower because of the fertility rate in South Africa.

UNIVERSITY of the
WESTERN CAPE

CHAPTER FIVE

DISCUSSION OF THE RESULTS

5.1 Introduction

The study emphasizes on the assessment of the data quality on age and sex in South Africa. It comes across specifically on the backdrop of the United Nations recommendations around the collection of high quality demographic statistics that should inform population policies. Indexes measuring the quality of data on age and sex were fundamental in this study. Therefore, the findings obtained from the analysis are discussed and interpreted in this chapter. The discussion covered three sections following the research design. The first focuses on the main procedures carried out in this study; the second is based on new facts contributing to the improvement of knowledge and the last part presents the recommendations which will contribute to policy formulation.

5.2 Main Procedures followed in the Research Design

UNIVERSITY of the
WESTERN CAPE

The quality of data on age and sex collected from census and surveys is quantitative and mainly built on normative comparison design with the aim of not only detecting and explaining variations observed in the statistics but also to help improve new facts or methods to adjust in the future. The application of Census 2001 and Community survey 2007, General Household Surveys 2004 and 2007, Labour Force Surveys 2005 and 2007 was carried out to achieve the objectives of this study. Secondary data obtained from Statistics South Africa was also used in this study.

During the previous data collection operations, the approach of enumeration in census and surveys involved a face to face and household self-completion questionnaire applied throughout the country. The questionnaire was used to gather information relating to the ages, backgrounds and past experiences. A post enumeration survey was conducted to determine the degree of undercount in Census 2001, Community Survey 2007, General Household Surveys 2004 and 2007, Labour Force Surveys 2005 and 2007. A sample survey was also conducted by means of personal interviews. The merger of the house file, worker's file and personal file using the

household questionnaire was used in collecting information. The information collected concerned backgrounds, living conditions and past experiences. The Census, Community survey, GHS and LFS files were obtained in SPSS format. Excel was used for age analysis using it's built in formula structures, age pyramids, age and sex ratio curves as this made it possible to have an overview constructed assessment on the quality of data.

The study focused on the quality of age and sex declaration in South Africa and took into consideration the nine provinces including the population groups in the analysis.

Some contributive factors to this purpose included errors arising from inaccessibility and cooperation with respondents, difficulties in communication, lack of boundary descriptions and change of addresses. Content errors were observed where characteristics such as age, sex, marital status, economy activity, level of educational status, and so on of a respondent in a census or survey are incorrectly reported or tabulated. In addition, content errors have two sources: firstly, an enumerator records incorrect response and secondly, a respondent provides a wrong response.

In each of the data collection instruments used, the question asked about age can be derived by "date of birth" or "completed number of years." From these two questions, different ages can be obtained. Both ways are applied in the recording of ages. According to this study, the evaluation of the quality of data made use of some indirect methods following Whipple's index, Myer's index, Age ratios, Sex ratios, Combined Index of United Nations and the Pyramids of population. In some instances, all the methods were applied while in others, only a few were used, depending on the nature of data at hand.

UNIVERSITY of the
WESTERN CAPE

5.3 Evaluation of Content Errors through the Attitudes of Respondents

The content errors are evaluated through the attitudes of respondents in order to report ages ending with a certain digit in preference to other digits. The main objective of the study was to describe and compare age and sex structure tools such as indexes, figures and population pyramids.

5.3.1 Evaluation of General Household Surveys 2004 and 2007

5.3.1.1 Assessment using the Whipple's Index

5.3.1.1.1 National Level

The General Household Survey (GHS) provides important information on the general living conditions across South Africa. The results of Whipple's index at national level from the observations in chapter four indicate that more males declared the better responses on age ending with 0 or 5 as compared to females at both dates of GHS. Perhaps females still need to be encouraged in terms of responses on age. This might be a possible explanation to the fact that females did not have a good understanding for the purpose of the survey. These findings confirm the literature which indicates that the questionnaire must be accessible, simple and clear to everyone. Females are more likely not to access the information easily hence the quality of their declaration is inaccurate. Feskens *et al.* (2006) emphasized on this issue and suggests that, it is more useful to make questionnaire more understandable to reduce the bias in the response. However, the GHS 2004 encountered better results compared to GHS 2007 at national level. As part of some developmental goals, there ought to be specific policies geared to the promotion of female's education in the country. Therefore, an intensification of sensitization among females during the pre-Census enumeration is critical in improving their awareness to this problem.

UNIVERSITY of the
WESTERN CAPE

5.3.1.1.2 Provincial level

For the GHS, the results of the analysis at provincial level attempted to confirm the hypothesis in which the quality of age and sex distribution is more accurate at national level compared to provincial level. However, a slight improvement is observed in the quality of data on age and sex distribution in the Eastern Cape, Northern Cape, Free State and Limpopo for the males, while in Eastern Cape, KwaZulu-Natal, North West, Gauteng and Mpumalanga for females. A possible reason for this is that the patterns of digit preference is consistent among females and vary from one province to another. Generally in the province with rural dominance, the patterns of preference for digits ending with 0 or 5 are more pronounced due to the factor of education as respondents could not access information easily and properly. For instance, KwaZulu-Natal

recorded the highest Whipple's index in both surveys. Moreover, this province is mostly rural and poorer.

The paradox arose with Gauteng which is affected by these patterns of preference. The alternative answer could be people did not have any interest in the survey, which revealed the weaknesses of the campaign surrounding the operation of data collection or political patterns could also have created distortion in the reporting of age amongst Gauteng male and female respondents. However, the GHS 2004 was more likely accessible compared to GHS 2007; hence, the results of males in six provinces (Western Cape, Northern Cape, KwaZulu-Natal, North West, Gauteng and Mpumalanga) were lesser in 2004 than in 2007 and for females in five provinces (Western Cape, Northern Cape, Free State, KwaZulu-Natal and Limpopo) were also lesser than those occurred in 2007. These issues raised some questions towards government's policy implementation as it should be revised according to specific strategies developing in provinces.

5.3.1.1.3 Cross-ethnic group differences in age heaping

An analysis of how the ethnicity differences affect the quality of declaration on age results were obtained through Whipple's index based on gender in order to measure the variable of index amongst population groups (African/Black, Coloured, Indian/Asian and Whites). The results show that both GHS data observed the strong preference to digit 0 and 5 (increase of indexes) among African/Black, followed by Coloured, Indian and female white. Only the index of male white had decreased.

Though the analysis in the previous chapter shows that Indian males and females have the highest index of Whipple, the issue highlighted by Graham et al (2006) is that cultural and ethical factors influence the declaration of age. All this shows that culture is an obstacle in terms of declaration of age and still prevails in the Indian/Asian society. In general, the overall picture is that the index of women increases substantially across all population groups from 2004 to 2007 although the African /Black realized moderate increase as well from 105.67 to 107.53. However, the GHS 2004 was probably more reliable than GHS 2007. Hence with regard to historical background, women are still experiencing issues of inequality in education due to some coercive policy instruments which do not take women into account.

5.3.1.2 Assessment using the Myer's index

5.3.1.2.1 National level

Myer's blended index is used in ensuring the accuracy of single age data, detect age heaping and digit preference in order to compare statistics among gender, population groups and across provinces from different dates and instruments. It also helps to measure the improvement achieved on the quality of declaration on age. In attempting to evaluate at national level the Myer's index in both GHS, the statistics revealed that the values of Myer's index are very low and have been progressively improved. It reached the highest value in the survey of 2004 at only 8.506 for males and 6.387 for females. It was reduced to 3.68 for males and 4.179 for females in GHS 2007. The results of Myer's index attempted to confirm the hypothesis in which the quality of data on age and sex collected from surveys was modified over years across gender. The trend has been modified for example in 2004, females' declared their age better compared to males. The reverse is observed in 2007 where males declared their age better than females. There are numerous possible reasons for that outcome, for instance, males and females were primarily informed about the issue of the survey in 2004. The females were sensitive to age declaration while males were likely reserved in responding. In 2007, female's interest reduced in responses or they did not see the change expected while males became more aware of the importance of the survey. If the change occurred in the orientation of the thinking of the population, it will benefit females especially in case they declared their true age. In fact, the GHS of 2007 was probably more comprehensive and accessible than the GHS of 2004. Improvement was observed in both sexes even though females were more involved compared to males in GHS 2007. According to Myer's index, GHS 2007 is a better instrument as the government will be able to plan appropriately as to what is meant for the population according to age. There would not be any shortage of resources for any particular age or age group.

5.3.1.2.2 Provincial level

National level tends to hide differentials at lower level and thus a reason for a break down at provincial level is important. For Myer's index, the results at provincial level revealed an attempt to confirm the hypothesis in which the quality of declaration of data on age and sex is poorer

UNIVERSITY of the
WESTERN CAPE

when compared to national level. In fact, progressive improvement is achieved by the male population in seven out of nine provinces namely Western Cape, Eastern Cape, Northern Cape, Free State, KwaZulu-Natal, North West and Limpopo. The same progressive improvement is also observed in females in eight provinces except Western Cape. However, the patterns of preference have changed the trend. In 2004, this pattern was more pronounced among females than males and in provinces of rural dominance. In 2007, the reverse was observed. The GHS of 2007 was designed in such a way to be more accessible by the respondents compared to the GHS of 2004. This could be the reason of the improvement observed. In the seven provinces, males improved their responses better than in 2004 except in Gauteng and Mpumalanga. For females, the improvement occurred in eight provinces except Western Cape. The findings revealed that the revision of the questions over years which made it simple and clear can be a reason for such outcomes. This means that the literature in which Jones *et al.* (2010) insists that appropriate training have to be achieved with particular challenges in various settings; every single question must be explained to avoid difficulties on the field of duty as stated above was supported by the findings. Government needs to encourage respondents in these particular provinces which are showing no interest to ameliorate the quality of their response by organizing supplement communication campaigns.

UNIVERSITY of the
WESTERN CAPE

5.3.1.2.3 Cross-ethnic group variations

In identifying the differences among the ethnic groups, an attempt has been made to assess how ethnicity differences affect the quality of declaration on age. Results of the questions on issue related to this point were obtained through the analysis of Myer's index based on gender in order to measure the variable of index amongst population groups. The results in cross-ethnic variations, for Myer's index pointed out the evidence of age heaping per gender like the other aspects of South African demography. The distortion observed on the reporting of age among respondents of the different population groups might be explained by the approach the enumerators used. It did not convince for instance the Indian population group which is influenced by the cultural backgrounds as cited above. The progressive improvement achieved in African/Black, (male and female), White (male and female) and male Coloureds is in agreement with Jones *et al.*'s (2010) as mention above (Section 5.3.1.2.2).

In the application of hot decks imputation to assess the age of a partner for example, when the respondent was not able to remember his/her exact age, the age of the partner could help to assess. However, the real issue of marital status in South Africa is that the change operates to improve the recognition of marriage which is influencing the declaration on age in some couples especially those belonging to African/blacks, knowing that this population group is the one more affected by the marriage pattern. However, sociological perception of the people and legal definition of marriage is different. Respondents have different cultural understandings as to what constitutes marriage which is a practice problem in South Africa. These deficiencies suggest that, registration of marriage like that of births and deaths might not be recorded according to legal responsibility as much as it should. The new law comes to clarify these features which indicate some discrepancies. Whichever way of looking at it, the GHS 2007 seems to be likely more accessible for population groups than the GHS 2004, though males of three sub-groups have improved the quality of their responses over years. It is crucial to initiate more comprehensive studies to establish appropriate policies to cope with the increase in preference especially in the Indian/Asian population group.

5.3.1.3 Assessment of Combined Index of United Nations

UNIVERSITY of the
WESTERN CAPE

5.3.1.3.1 National level

In attempting to evaluate at national level the CIUN in both GHS, the outcomes showed that the values of CIUN are high. The quality of data was not accurate over years. A slight improvement is observed and the decrease of indexes could not show a visible change in the quality of data on age and sex distribution. However, the GHS 2007 was likely more accessible compared to GHS 2004. A possible reason could be attributed to the effect of migration in neighboring countries (Botswana, Swaziland, Zimbabwe and so on) which occurred during the end of apartheid and continues to influence the sex and age indexes hence the CIUN is related. Another reason could be the pattern of mortality also observed at the same period because of colonialists' atrocity. The review of policy on employment is likely to be more attractive to both males and females in regard to legal age to work and will bring some light in the regulation of this issue.

5.3.1.3.2 Provincial level

The evaluation at provincial level of both GHS revealed high values of the index in Western Cape, KwaZulu-Natal and Gauteng between the dates. The low values of index over years are observed among the rest of the provinces compared to the previous. This increase and decrease could be explained by the influence of sex selective migration of the working age population which affects the ratio of male to females in different age groups. This corroborates the literature review cited by Chaurasia (2005). The results obtained in GHS 2007 were probably lesser in six provinces compared to those in GHS 2004. The improvement is revealing that the GHS 2007 was likely more accessible than GHS 2004. However, the quality of population structure by sex and 5-year age groups is inaccurate in four provinces namely Western Cape, KwaZulu-Natal, North West and Limpopo while it is very inaccurate in the rest of provinces. This may indicate a problem of inadequate targeting. Therefore, closer inspection is needed to ensure that this is not the case. When looking at geographical relation, there are disparities between provinces in terms of education. Most of females in rural areas are none educated mainly because of the environment they live in. When compared to males there are not many differences in this regard. Geographic factors also play role in the quality of declaration on age and sex in South Africa.

UNIVERSITY of the
WESTERN CAPE

5.3.1.3.3 Cross-ethnic evaluation

Population ethnic group remains a very informative demographic variable. It was used to evaluate the extent of age and sex assessed among African/Blacks, Coloureds, Indians/Asians and Whites according to gender in the 2004 and 2007 GHS using the CIUN. The results show a decrease of the index among African/Blacks and Coloureds. The slight decrease of index on Coloureds has shifted them in other category indicating the quality of responses on age is inaccurate. Indians and Whites have respectively increased. The preference is more pronounced. The White population with high level of education provided bias responses on age which is revealed by the increase of their CIUN (from 45.63 in 2004 to 74.59 in 2007). A possible reason could be the political issues as is confirmed with the literature in which McCarty et al (2007) has mentioned that some political issues have a significant role on the informant and they can influence the response rate. Whatever the reasons are, the GHS 2004 was possibly more accessible for males than females, while the GHS 2007 was designed in such a way so that

females were likely more comfortable than males. The government can revised the tax policies to ensure confidence in the White population group. To achieve this, there is need for government agencies and NGOs to strengthen education and community outreach programs. There is also a need for exciting dialogue on the economic and political constructions that shape the gender norms. This can be incorporated into various forums such as community meetings and church activities.

5.3.1.4 Age ratios variations

For both GHS, an attempt was made to evaluate the variations observed in age ratios at national level. When comparing both graphs of age ratios for the GHS 2004 and GHS 2007, it is evident that the GHS 2007 seems to be more reliable than GHS 2004. This irregularity still remains even after adjustment. The findings confirm Chaurasia's (2005) assertion that there is no guarantee that huge fluctuations in age ratios are due to unexpected changes in the fertility, mortality rates. He also established that there is no automatic link between the pattern of migration and age misreporting of respondent of a specific age or age group at the time of the enumeration. This pattern of age ratio can possibly be due to the weakness of enumerator bias in addition to misreporting in the age information. It will be appreciated not to consider age ratios as a measure of assessment because it quantifies only the quality of age and sex information. It is most likely, the result of migration, although the contribution of age mis-statement and omissions to the distortion cannot be ruled out.

UNIVERSITY of the
WESTERN CAPE

5.3.1.5 Sex ratios curves

In attempting to analyze sex ratios by age for the whole population in South Africa, according to both GHS, the findings showed a low rate of sex ratio at birth, a lower sex ratio in the ages under 15yrs indicating under-enumeration of children and unusual shortage of males of working ages. The low sex ratio at birth is abnormal in a developing country such as South Africa. This may be due to the under-enumeration of births, or due to the infant mortality rate or the decrease in fertility. These findings reveal some features in the sex distribution hence it is contrary to the expectation. Even the literature indicates that the sex ratios at birth range between 104 and 106. Sex ratios decrease with increased age as in normal population. Males at every age have higher

mortality rate than females (Siegel *et al*., 2004). When comparing both curves of GHS, it indicates that the GHS 2004 was likely more reliable than GHS 2007. The natural biological endowments and existing evidence in defiance to all the efforts has led to decrease in sex ratio in general. The development in technology was expected to facilitate a healthy result of the physiological process, but the orientation has been regularly changed. Generally, females who went for sex determination without medical advice are educated and have finance possibilities. The selection of sex is causing the increase or decrease in sex ratio. It will be important to encourage policies, information and communication campaigns affirming and promoting women's status in the family and society.

5.3.1.6 Population pyramids

The assessment of GHS using age and sex distribution at national level over years was attempted. The results reveal a steady growth in the whole population from 97 153 in 2004 to 109 831 in 2007. The graph has a broad base in 2007 as compared to 2004. This is a type of developing country's example where birth rates are high but health standards of living are low. The shape is different with the increase observed at different ages of the population. The greater proportion of younger age is observed. The rapid shift in the proportion of the population aged from 15 to 64 has to be brought about by either or both of an increase in fertility rates, or significant improvements in adult mortality rates. The increase in the working age population can be seen in the distinction emerging between 2004 and 2007. The change over years is important in such a way that it provides some strong evidence that the time of observing and capturing variations has little improvements in the standards of living and environment.

UNIVERSITY of the
WESTERN CAPE

With the high level of child and early adult mortality in South Africa, women need to ensure that at least one daughter survives to become a mother herself. A further consequence of this increasing fertility rate is that the rate of population growth is swift. The great unknown is the level of international migration into South Africa, and predicting future trends in this variable is impossible. These features reveal that the quality of distribution on age and sex GHS is not accurate. This reflects the gradually changing age structure of the South African population and increasing fertility rates, but the policy implications are significant. In each and every year after 2007, more children will be enrolled at schools. If budget assigned to schooling is not increased

but continues to decrease in real terms, the real per capita expenditure per pupil will slow rapidity. In this regard, the population aged 10 to 24 will reach its absolute maximum before the current year. Improving the funding of schools, however addresses only one aspect of quality related to content. What relevant policies have been made about the education in this country? Additional effort should be expended in reducing the child and infant mortality rates (keeping children alive must represent one of the most significant objectives in the country. Public health policies could be considered to reduce the burden of disease (communicable infections and diseases of lifestyle) among young adults. However, the change in the overall distribution of the GHS 2007 in South Africa will be slow to appear given the lead time required for changes in education policy and practice to have an impact on the adult lives of current and future learners.

5.3.1.7 Dependency ratios

Closer inspection of the variations was made in age and sex structure at national level for both GHS 2004 and 2007. The question arose from the analysis in previous chapter which states: "How useful are dependency ratios?" In fact, the child dependency ratio is measured as the ratio between the population below working ages (15 years of age) and the working age population (from 15 to 64 years of age) multiply by 100. This indicates the number of children dependents that need to be supported for every 100 people in working ages. In the same way, the old dependency ratio provides the number of old age dependents that need support for every 100 people in working ages. The results of the findings in GHS showed that the child dependency ratio likely increased faster compared to the old dependency ratio and the total dependency ratio which is the sum of the child and old dependency ratios that has the same trend as the child dependency ratio. The challenge is that the ratios do not reflect whether the people of working ages are actually economically productive or whether the older person and children are economically dependents. For instance, many old people are financially and physically independent whereas there are substantial portions of the working age population who may not earn incomes because they are unemployed, or unable to work, or they are in school, or in prison, or have opted out of the labour force entirely. Though, it is very difficult to consider intra-family financial assistance in an overall measure of social support. It is feasible to include employment characteristics of the population in the relevant age groups. However, fewer young dependents in the short run will be translated into fewer future workers in about two decades. At that time, the

UNIVERSITY of the WESTERN CAPE

net effect will be that the old age dependent will be increasing in the numerator while the number of working age persons to support them will be falling in the denominator.

Considering that there are economic, social and demographic problems in the general utilization of scarce and limited developmental resources of which there are several competing claims and demands; the projection of population under conditions of low and high fertility demonstrate the long run effect of reducing fertility. Not only will there be a significant fall in the level of childhood dependency and consequently total dependency but a noticeable improvement in the supply of education, health, employment and other facilities for the population. In addition, public health programs for the control of infections and contagious diseases and the improvement of sanitation and environmental hygiene have to benefit a great deal. Coupled with the effect of rising standards of nutrition, the health programs will invariably lead to a faster than anticipated reduction of the death rate in all ages. The government has to consider the lack of family planning programs and earmark substantial resources for education at all levels and in all fields of human endeavor. All these actions will create the variation in the size of the various functional segments of the population.

5.3.2 Evaluation of data from the Labour Force Surveys 2005 and 2007

UNIVERSITY of the
WESTERN CAPE

5.3.2.1 Results using the Whipple's index

5.3.2.1.1 National level

Looking into the age and sex structure of data produced from the Labour Force Surveys (LFS), the inspection of single year age-sex data using Whipple's index shows concentration of population at ages ending to digits 0 and 5. Comparing Whipple's index of South Africa with indexes obtained from other developing countries, it was seen that the age-sex data of South Africa seem to be relatively good at national level. The results of LFS 2007 showed the improvement observed, suggesting that the LFS 2007 was likely more reliable compared to the LFS of 2005. Although both sexes still need to be encouraged in terms of responses on age, women showed lower preference in declaring age ending with 0 or 5. A possible explanation to these results is that females easily change their mind; they are likely more sensitive when declaring their ages while males are reserved in their attitudes. Over time, one can therefore

admit that some improvements were made in the design of LFS with regard to data collection on age.

The outcomes for both LFS revealed that the progressive improvement was observed over years and the change in the trend of declaration of which more females provided better responses on age than males. There has been a great effort from the government to engage in the whole process of collecting data that is of good quality. These strategies started from the organization, the recruitment of personnel, training, techniques and different methodologies applied as approaches of collection guidelines and advices to enumerators and so on. Examining age and sex patterns in LFS for both years, it can be deduced that the decrease in indexes in males is likely less than for females. The results of LFS 2005 revealed that the low participation of females in Labour Force influences their interest in the survey operations; females were in lower numbers.

New developments in the economy over year's have encouraged female labour participation and results in the increase of females in the survey. Their implication in the Labour Force helps them to become more expressive compared to males in 2007. From a participative point of view, the government has to articulate great concern about the internal imbalance of the Labour Force participation between males and females as well as build valuable stimulus in labour participation of females. The integration of females in the Labour Force will bring a light in their attitudes hence the confidence is acquired with times. This policy will unlikely increase the relative size of the economically active population by higher capital intensity; labour productivity will need to increase over and above this mechanism in order to compensate for the impact of female participation in Labour Force. Whatever the reasons are, there will be need for more education and training to speed up human capital formation. The shift in the age structure will also change and have an effect on the patterns of employment across different sectors of the economy.

5.3.2.1.2 Provincial level

For LFS, the results of the analysis at provincial level attempted to confirm the hypothesis in which the quality of data on age and sex distribution collected from this type of surveys in South Africa has been improved differently over years across the provinces. Considering the LFS 2005,

the quality of data on age and sex distribution was likely better in five provinces namely Western Cape, Northern Cape, Free State North West and Mpumalanga for the males, while in LFS 2007, Western Cape and Northern Cape encountered inaccurate quality of data on the age for females. However, the LFS 2005 was more reliable for males compared to females while the LFS 2007 was likely more accessible for females than males. A possible reason is that the patterns of digit preference are consistent among males in LFS 2007; males were likely less expressed than females as progressive improvement occurred among females. The results of analysis showed that the indexes for males vary from 101.12 (North West) to 119.54 (KwaZulu-Natal) in 2005 while for females vary between 101.03 (Western Cape) and 127.39 (KwaZulu-Natal). According to LFS 2007, the indexes vary from 101.76 (North West) to 109.02 (Northern Cape) for males while females range between 102.27 (Free State) and 111.42 (Northern Cape). The slight increase in the inferior limit highlights a small observation in the quality of data of 2007 and in that particular province, hence the consistent decrease obtained in the upper limit reveal an improvement realized in those particular regions. Despite the fact that the LFS 2007 was not of better quality for males, there are some spectacle improvements achieved in three provinces namely the Eastern Cape (from 110.48 in 2005 to 102.25 in 2007); Gauteng (from 109.89 in 2005 to 102.42 in 2007); and Limpopo (from 113.79 to 97.37 in 2007).

UNIVERSITY of the
WESTERN CAPE

5.3.2.1.3 Cross-ethnic group differences

For both LFS, an attempt was made to evaluate the variation observed in Whipple's index across gender among population groups. The results of the findings had confirmed one the specific research questions investigated: "How ethnicity differences affect the quality of declaration on age and sex?" The results obtained in the previous chapter reveal the pattern of differing preference among the population groups and across gender. The indexes of Indian, White males and African/Black population group have decreased whilst the indexes of Coloured males, Indian and White females have increased. The pattern of preference is most pronounced among Indian females followed by White females. The relationship between educated females, labour force participation and quality of declaration on age are complex. Some statistical studies (Graham *et al.*, 2006; McCarthy *et al*., 2007) found lower reporting on age with more participation in the labour force while others found the opposite. Such inconsistency is not surprising given demographic and household characteristics, cultural forces, socio-economic and political

circumstances around the country. While statistical researches into females labour force participation and quality of declaration on age had not produced clear findings, the conceptual links are clear. However, the government has to encourage female education to greater levels which occupies a unique place in demographic discourse and policy because educated females play a role of exposing other females to new ideas about culture, act as sources of information, social support, social pressure that diffuse their lifestyles and ideas to other females.

5.3.2.2 Assessment using the Myer's index

5.3.2.2.1 National level

The examination of single year age–sex data using Myer's index indicated concentration of population at ages ending with 0, 2 and 5 for both LFS at national level. The respondents were fewer for ages ending with the digit 1. The reporting of age among males was better compared to females in both surveys, but females showed better improvement of the quality of their responses than their counterparts in 2007. The results obtained in the analysis have confirmed the hypothesis in which "Gender differences are associated with the poor quality of declaration on age and sex." The possible reason to this failure linked with the question asked for the year of birth where many respondents tend to report being born in the years with terminal digit 0 as for instance, 1940, 1950, 1960, 1970 and so on. This is particularly observed with the respondents who do not precisely remember their date of birth and year close to the year ending in 5 or 7. The survey is conducted in year 2005 ending with 5 and the second in year 2007 ending with 7. Since survey took place in 2005 when age is calculated, about three quarters of those who declared being born in the years ending with zero (1940, 1950, 1960…), especially those who do not remember precisely when they were born close to the years with these digits (years ending with 5, 7 for example) will end up having their age ending with 0 and the remaining quarter will have their age ending with 5. This distortion is due to incorrect responses of age by the respondents. The government will need to continue the actions of information on importance of surveys by organizing at different stages campaign of vulgarization on the quality of data (trying to explain the necessity and role of survey in the country).

UNIVERSITY of the WESTERN CAPE

5.3.2.2.2 Provincial level

In attempting to evaluate at the provincial level, the Myer's index in both LFS results indicated that the values of Myer's are small showing high accuracy in responses on age per province. The outcomes of the findings had not confirmed the hypothesis in which the quality of declaration on age and sex is poorer at provincial level compared to national level. In 2005, the indexes for males are ranged from 4.29 to 7.39 and for females from 4.37 to 7.91. In 2007, the index for males varies between 3.92 and 6.94 and for females varies from 4.54 to 6.86. The difference between the highest and the lowest for males is 3.1 while for females is 3.54 in LFS 2005. In the same vein the differences observed between the highest and the lowest index for males is 3.02 and for females is 2.32. This small difference reflects high reliability in the quality of responses on age. Considering the LFS 2005, five provinces have lowest values of Myer's and four with the highest Myer's values for males while three have lowest values of Myer's and six highest values of Myer's for females. This shows that the LFS 2005 got better responses for males compared to their counterparts. Applying the LFS 2007, three provinces obtained lowest values of Myer's and six with the highest values of Myer's for males while two provinces showed lowest values of Myers and seven provinces obtained highest values of Myer's. This reveals that the LFS 2007 got better responses for females than males hence the reporting on age was greater among females compared to males. The decrease of indexes was observed in five provinces for males and in six provinces for females in LFS 2007 indicating the progressive improvement in the quality of reporting age. The LFS 2007 was doubtless, well designed and likely more accessible to respondents than LFS 2005.

Apart from problems affecting the delivery of birth certificates in many provinces, the high values of Myer's in some provinces are also affected by migration. Age and sex selection are the feature of migration. Generally, migrants are young and males. This selection changes the normal sex and age structures of both place of origin and destination. Therefore, it is possible to have higher index at the provincial level than at the national level. Eastern Cape, KwaZulu-Natal, North West, Gauteng, have higher index than at the national level in LFS 2005 while Western Cape, Eastern Cape, Northern Cape, Free State, North West, Mpumalanga and Limpopo have highest index than at the national level in LFS 2007. It can be concluded that migration has a strong impact on Myer's index at provincial level. However, high index indicates high inter-

provincial migration and low index shows low inter-provincial migration. The high inter-provincial migration disturbs the age and sex structure of the population. There is a need to encourage policies to prevent the increasing sex ratio by organizing communication campaigns affirming and promoting women's importance towards development in society. This measure will reduce the surplus of male population in the near future in some provinces.

5.3.2.2.3 Cross-ethnic group differences

Investigating both LFS's, Myer's index was used to assess the extent of the declaration on age amongst African/Blacks, Coloureds, Indians/Asians and Whites according to the gender. The results of the analysis per population groups attempted to confirm the hypothesis in which: "The ethnicity differences have varying influence on the quality of data on age and sex collected from surveys in South Africa." The results revealed that the Myer's index using LFS 2007 has poor declaration among Coloureds population group and Indian males while Indians females, African/Blacks and Whites population groups showed a decrease of its Myer's index. This decrease of Myer's index indicates the improvement realized which reveals a great change of attitudes among White population, Indian females and African subgroup. As mentioned above, the patterns of preference occurs as a recurrent factor in Indian population group. Nevertheless, besides that recurrence within Indian group population, the policies cited above remained identical. However, the government will need to promote education in all categories of age in order to break down those cultural barriers among Indian population. The programs will be made in such a way so that every category group of age must fill more confident in declaring their ages.

UNIVERSITY of the
WESTERN CAPE

5.3.2.3 Assessment of data using the combined index of United Nations

5.3.2.3.1 National level

An attempt was made to evaluate the CIUN in both LFS at national level. When comparing the results of CIUN for LFS 2005 and LFS 2007, the quality of data deteriorated over years. The increase of CIUN observed in 2007 is a direct consequence of great increase of age specific indicator of males and the slight decrease of age specific indicator of females. The decrease of females' indicator also affected the variation of sex ratios which increased automatically. The

transfer of females from older ages to younger ages and for males from younger ages to older ages could create these distortions or also the feature of out-migration into neighbouring countries might be a possible reason to this deterioration of CIUN. However, government needs to follow actions in terms to improve the labour sector and apply more effective methods to create new opportunities with efficient environment. This can discourage workers who are still in need of new challenges.

5.3.2.3.2 Evaluation per province

For CIUN, the results at provincial level revealed an attempt to confirm the hypothesis in which the quality of data on age and sex is poorer at provincial compared to national level in both LFS. The decrease of CIUN is observed in three provinces namely the Eastern Cape, Free State and Gauteng. The slight improvement of Gauteng and Eastern Cape is not visible because the CIUN remain in the same category. Nevertheless, the CIUN of Free State has decreased, changed the range despite the fact that the quality of its distribution is not accurate. Their respondents still have to improve the quality of their responses on age. The increase of index in five provinces (Western Cape, Northern Cape, KwaZulu-Natal, North West and Limpopo) indicated the change of range; the quality of sex and 5-year age distribution which was inaccurate has become somewhat very inaccurate. The Mpumalanga province has increased but still in the same range indicating that the quality of data on age is highly inaccurate. The patterns of preference are more pronounced among males than females in both surveys but the LFS 2005 was more likely accessible to respondents than LFS 2007 where the age reporting was possibly greater in LFS 2007 compared to 2005 indicating the distortion in the age and sex distribution. The deterioration of indexes is observed among males than females and in provinces with White population group dominance (Western Cape), Indians (KwaZulu-Natal) and rural areas (Mpumalanga, Limpopo, North West and Northern Cape). A possible answer is that males tend to be more open than females but over years the inverse is observed. Females become more open while males tend to be more reserved. Although both sexes still need to be encouraged in terms of responses on age.

UNIVERSITY of the
WESTERN CAPE

5.3.2.3.3 Cross-ethnic group variations

Investigating both LFS's, CIUN was used to evaluate at what extent the declaration on age was assessed among African/Blacks, Coloureds, Indians and Whites according to gender. The results of the analysis per population groups attempted to confirm the study research question: "how do ethnicity differences affect the quality of declaration on age?" The purpose of this question was to determine the influence of ethnicity on the quality of declaration on age using the LFS. The 2007 results revealed the data absence of regularity of ages declared by African/Blacks, Coloureds and Whites population groups. The results obtained in the previous chapter revealed an increase of the values of this index indicating there is no improvement realized hence no change is observed in the attitudes as far as the patterns of regularity are concerned. Only the index of Indians has slightly decreased. This slight improvement could not however modify the range of the index. As a consequence, the quality has remained the same, in that it is still highly inaccurate.

Concerning the pattern of preference, the age reporting of males was improved among Indians and Whites. In general, for all the ethnic groups, the female population was more pronounced in preference compared to males in 2007 except African/Blacks. However, the age reporting was likely more accessible among females in LFS 2005 than males; while in 2007, the age reporting was more accessible among males than females. When comparing the differences between indexes using both surveys, it reveals that the LFS 2007 was probably more reliable compared to LFS 2005 (For 2005: 32-21=10, 41-32 = 9, 77-41 = 36; for 2007: 46-30 = 16, 60-46 = 14, 70-60 = 10). In fact, 16, 14 and 10 are closer than 10, 9 and 36. A possible answer to this issue of preference is that the lack of education in female population, the non- consideration of female in the Indian society or inequality observed also in African/Blacks and some political issues for Whites continue to affect CIUN. Possible reason could be attributed to the uncertainty of the political situation in the country and the backlash if the black government is ushered in. Subsequently, other factors such as increase in insecurity, crime become important. Another possible reason is that the gender of enumerators could influence the response of female Indians. If more males are assigned to collect data, females tend in general to reject due to cultural reasons.

UNIVERSITY of the WESTERN CAPE

5.3.2.4 Age ratios variations

The age range considered is 0 to 70 years. The age ratio index is given by the mean of the deviation from 100. The results of applying age ratios are shown in Table 4… Normally if age reporting was accurate, then age ratio should deviate very little from 100. The calculated age ratios exhibit some differing patterns. Comparing both surveys, the age ratios for males show surpluses at ages 10, 14, 16, 18, 30, 32, 38, 40, 42, 45, 50, 52, 60, 65 and 67. The deficit occurred at ages 1, 3, 15, 17, 19, 29, 31, 33, 39, 41, 51, 61 and 66. The age ratios for females indicate surpluses at ages 10, 14, 16, 30, 32, 35, 40, 42, 45, 50, 52, 55, 60 and 65 while the deficit of females appear at ages 1, 9, 11, 15, 29, 31, 33, 34, 41, 43, 46, 47, 51, 56, 58, 61, 66, and 69. The LFS age data designate the presence of substantial errors. These errors arising from age mis-statement, digital preference and omissions may have caused distortions in the data. If the examination of age ratios for LFS (see table 4.2.4), the surplus population was reported with some small expectation in five year old age groups ending to digit 0 and 5.

In previous discussion of digit preference, the existence of apparent preference for digit 0, 2, 4, 5, and 8 are for males and 0, 2, 4, 5 for females. Thus it seems that digit preference has contributed to some of the distortions in the data. The deficit of male population reported in the young adult ages as revealed by the age ratio analysis might be due to the out-migration of long term male migrants. In fact, another part of the distortion might be authentic. The issue that the age ratio index present better age reporting for females than for males may be due to the fact that the age distortion caused by migration may be more for males compared to females.

UNIVERSITY of the
WESTERN CAPE

5.3.2.5 Sex ratios variations

An attempt was made to assess the sex ratios by age for the South African population according to LFS. The results of the findings had confirmed the literature in which National Statistics Bureau (2008) explained that the sex ratio patterns could be observed in the population where male population exceeded female population with the peak in the working age because of the effect of migration which is sex selective. The sex ratio patterns for LFS 2005 can be summarized as follows:

a. Excess of males in ages 0, 1, 4, 5, 9, 14, 15 and 18.

This might be due to under enumeration of female children aged at 0, 1, 4, 5, 9, 14, 15 and 18. Or it might be due to the transfer of female children of age 1 to 2, 4 to 3, 5 to 6, 9 to 10, or transfer of female children of age 14 to age 13, from 15 to 16 and from age 18 to 17.

b. Excess of females in ages 3, 6, 12, 17 and 19.

As presented in the LFS 2005, a possible explanation could be the under reporting of males aged 3, 6, 12, 17 and 19. This might be due to overstatement of the ages of females aged 5, 8; some of whom were transferred to the next higher age 12 or 17 and 19. The effect of puberty with the development of physiology pushed some teenagers to overestimate their age in order to be considered responsible.

c. Excess of females from 21 and above.

This can be divided into two parts: excess of females of working age (21 to 39) and excess of females aged 40 and above. The surplus of females of working age is common in developing countries and a possible reason might be out-migration of young adult males or might also be overstatement of females, omission or understatement of male ages. The surplus of females aged 40 and above might be explained by the practice of males to overstate their age which brings a shift into older ages and for females to understate their ages which shift them also into younger ages. It might be explained by the life expectancy of female which is likely greater compared to males.

The sex ratio pattern for LFS 2005 obey the rules of the typical developing country sex ratio model, but the only deviation is in the ages 3, 6, 12, 17 and 19 where the sex ratio of South Africa (LFS 2005) is below 100 while for developing country it exceeds 100 indicating that the quality of data is not accurate.

The sex ratio patterns in LFS 2007 can be grouped as follows:

a. Excess of females in ages 0, 1, 3, 5, 8, 11, 13 and 20.

This might be due to under-enumeration of male children aged 0, 1, 3, 5, 8, 11, 13 and 20; or transfer of male children aged 1 to 2, 3 to 4, 5 to 6, 8 to 7, 11 to 10, 13 to 12 and 20 to 19.

b. Excess of males in ages 2, 4, 7, 9, and 12.

The under-reporting of females aged 2, 4, 7, 9 and 12 might be due to the overstatement of males. Some females were transferred to the next or higher level; for instance 2 might be transferred in 3 or 4 in 5, 7 in 8, 9 in 10 and 12 in 14, 15.

c. Excess of males in ages 16, 17, 18 and 19.

This is a regular feature in developing countries and a possible reason might be the overstatement of females in puberty aged 16, 17, 18 and 19 because they are physically more developed, they are considered adults or, omission, or understatement of male ages which shifted them into higher ages, or, out-migration of young adult males.

d. Increase of sex ratio in ages 42, 45, 47, 50, 52, 56, 62 and 65.

This situation could be explained by the phenomenon where male migrants return home to prepare for retirement.

e. Excess of females aged 20, 22 years and above.

Generally, as the age increases, the life expectancy of female is likely greater compared to male. This situation might be due to biological factors which tend to increase the mortality rate of males. It might also be explained by the tendency of women of over 40 years of age to understate their age because they wish to remain young (they are shifted into younger ages), especially to gain more respect which shift them in older ages.

UNIVERSITY of the
WESTERN CAPE

Definitely, the sex ratio pattern for LFS 2007 follow the structure of sex ratio model of developing countries except the deviation observed in only ages 0, 1, 3, 5, 8, 11 and 13 where the value of sex ratio was below 100 while it's supposed to exceed as in developing countries. However, The LFS 2007 is more reliable than the LFS 2005. The respondents were more opened in 2007 compared to 2005; the quality of their responses was better than for LFS 2005. The presence of all these features is revealing that the quality of data on age and sex was not accurate. Despite the fact that the knowledge practices have not changed, the girl child remains neglected; there is a need for media and the NGOs to make concerted efforts in increasing the awareness levels and work, putting the girl child at the same pedestal.

5.3.2.6 Assessment of population pyramids

Investigating both LFS, population pyramid was applied to assess to what extent the declaration on age was accurate among gender at national level. The results of the findings attempted to confirm the specific research question in which "To what extent over years gender differences have affected the quality of declaration on age?" The LFS results showed a decline in the total population from 110, 545 in 2005 to 109, 387 in 2007 representing a decrease in majority of ages despite some few cases of increase registered here and there in the distribution. In 2007, the diagram reflects an expansive pyramid which reveals the characteristics of developing countries; where the life conditions are still poor with high infant mortality rate. This shape is completely different with the increase observed in younger ages, the decrease at old ages, the increase in working males and decrease in working female populations. More than half of the populations are young. However, the decline observed in younger ages (0, 1, 7 to 11, 13 and 14 for males while 3, 4, 5, 7 to 14 for females) is greater than in older ages 65 and above. Considering gender, the decrease observed in males is greater than for females at younger age which confirm the literature in which biological factors contribute to the vulnerability of males. As fertility and mortality rates fall, a window is opened up as the proportion of the population that is old and young (65 and over with those aged under 15). The increase in the working age population (15 to 64) emerging between LFS 2005 and LFS 2007 explained the shifts of the younger and older ages. This could be the implications of improvement in adult mortality rates or decline in fertility rates. However, the mechanics of demography are a poor predictor which influences the quality of age and sex structure of the population. This also could be linked to subsequent economic performance. The effect of HIV/AIDs epidemics in the population pyramids for 2005 and 2007 is observed with a slow increase in the proportion of working age males which is not equivalent with the decrease in the proportion of working age females, but it affects the quality of distribution on age and sex which is not accurate for LFS. A further consequence of this declining fertility rates is that the rate of population growth has slowed. The combination of migration factor which is not predictable and the decrease in fertility rates and decrease in mortality of adults has an important implication in the policy. The budget assigned to education for instance must be revised accordingly to the situation observed. The decrease observed in younger age must be considered in such a way the per capita expenditure per pupil has to be

controlled. This could affect the young adult to attend their maximum after the following year. The improvement of education has to consider all aspects related. Generally, the fertility rates change slowly over time, or a decline in fertility is balanced usually by an increase in adult mortality. The low fertility decline in South Africa is evident from the narrowing of the base of the pyramid from 2005 to 2007.

5.3.2.7 Dependency ratios

A deep understanding of the variations observed in age and sex structure at national level in both LFS in percentage for population under the three broad age groups revealed that the growth of the size of the population sub groups by age have social and economic implications for planning education, manpower, housing and other social requirements. Analysis of these will concentrate on the social and economic implications of the total dependency, school age population and labour force. The outcomes showed that the child dependency is decreased from 52.66 per cent to 51.35 per cent while the old dependency increased from 9.1 per cent to 9.44 per cent. The net effect of dependency will be that the old age dependence will be increasing. The total dependency also decreased but not at the same rate. In general, it is recognized that the developing countries of the world have high total dependency which impede their economic development. The high total dependency as a result of a large number of children suggests that government will have to invest large amounts of funds in servicing and maintaining the population.

In South Africa, the lower age limit of entry into primary school is seven years. Primary and secondary education takes 6 years and 4 years respectively. The primary school age population is defined to include those aged 7 to 13 years and the secondary age population as those aged 14 to 18. The primary and secondary school age population for the 13 years projections will tend to show a reduction in the primary school age population as a result of fertility decline. However, the government‘s aim for the future is universal primary education. It seems certain that the rapid and substantial growth in the primary and secondary school age populations in conjunction with the government policy of universal primary education will call for a heavy drain on national resources. Educational services measured in terms of the numbers of teachers, classrooms and

additional facilities are expensive and sometimes inaccessible because of the time and cost of training and developing the needed resources. Scarce as the resources might be, they have to be provided in proportion to the school age population, taking into account the effective resource per pupil ratio needed for achieving good results and the replenishment of wasted resources. All these will no doubt, take up a large chunk of the available resources for national development.

5.3.3 Evaluation of 2001 Census and Community Survey 2007

5.3.3.1 Assessment of Whipple's index

5.3.3.1.1 National level

Because the 2007 Community Survey served in the mid-period Census, an attempt was made to evaluate at national level the data from the Census 2001 and the Community survey using Whipple's index. The findings showed that the index had decreased across years. The results of Whipple's index attempted to answer the specific research question which was asked: "Over years, has the quality of data on age and sex been improved for censuses or surveys across gender"? In the census, the preference for digits 0 and 5 was lower among females compared to males while in the 2007 Community Survey the reverse was observed and males' preference was lower than their female counterparts.

UNIVERSITY of the
WESTERN CAPE

The results from the data obtained from the Census and Community Survey concur with the findings summarized under this section and further articulate the quality of age and sex distribution of both females and males sound accurate at national level. The age reporting in the South African Census 2001 and Community Survey 2007 has been very good. From the Census results of many African countries, it is difficult to obtain values of Whipple's index below 110. The findings however suggest some differences in that, more females declared the ages more properly than males for the population Census while the contrary is observed for the Community Survey. The possible reason for these gender differences can be that females might become more conscious compared to males or the information was more accessible across gender or the procedure to ask historical events especially to those who are unable to remember their age had contributed strongly in the assessment of the quality of responses on age in the whole country. Another possible reason is the fact that the age variable was derived from the information on date

of birth has been a strong factor in ensuring accurate age reporting. The accuracy of index of females in Census can also be linked to the fact that Census generally provides better statistics compared to survey.

5.3.3.1.2 Provincial level

In attempting to evaluate Census 2001 and Community Survey 2007, the Whipple's index was used. The findings highlight the answer for the specific research question which was the quality of declaration of data on age and sex poorer at provincial level compared to national level? In general the results are good in both instruments except in three provinces (Free State, North West and Limpopo). But, when compared, it reveals that at provincial level, for males the slight deterioration of index is observed in Western Cape, Northern Cape, KwaZulu-Natal, and Gauteng while for females it is observed in Eastern Cape, Free State, KwaZulu-Natal, North West and Gauteng. However, the slight improvement is also observed in the rest of the provinces. Males had improved the quality of their responses in five provinces while females did in only four provinces.

Males of Western Cape provided good reporting of ages over years, after six years they still have the smallest value of Whipple's index which means they are well informed and this could be the result of a lot of guidance and awareness according to the campaigns surrounding the operation of data collection; the involvement of everyone in the process which is a great concern in this province. Northern Cape females arise with the smallest index as this could be explained by the new policies which try to involve female participation in all sectors of activities; the programme of education for all is also bringing to light in the attitude of females to express freely their opinion. The position of Limpopo with the highest index for males and females could be explained by the fact that this region is predominate by rural areas, people who live there, the majority have low education. In seven out of nine provinces, males had better reporting on age compared to female counterparts. Totally, Community Survey 2007 indicates only slight avoidance in the attitude of respondents while Census 2001 highlights slight preference in Western Cape and slight avoidance in the rest of the provinces.

5.3.3.1.3 Cross-ethnic group variations

Considering the differences observed among the ethnic groups, an attempt was made to measure "How the variable ethnic group could influence the quality of declaration on age?" The evaluation of Census 2001 and Community Survey 2007 applied Whipple's index. The results in cross-ethnic differences indicated the evidence of age heaping per gender in both instruments. The distortions observed on the reporting of age among African/Blacks in particular might be explained by the level of education which was the consequence of poverty that this population group was facing, related to the issues surrounding the lack of education in rural areas. However, Coloured, Indian/Asian and White females realized a progressive improvement over years in their declaration on age while African/Black females deteriorated in their quality of responses. This problem (education) should be considered at all ages in such a way as comprehensive programs could help this population group to come out of this issue. Indian and White males response quality deteriorated as this could be explained by the approach used by the enumerators. Surveys used the same questions, but there are some differences in the details of the questions asked; these are source of variability in the responses. African/Black and Coloured males registered a slight improvement in the quality of their responses. However, Community Survey was more reliable for females compared to males while Census was more reliable for males compared to female counterparts. According to ethnic group using Whipple's index, Census 2001 was likely better designed than Community Survey 2007. The possible explanation relate to the accuracy of the census data could be due to different approaches used in these designs.

5.3.3.2 Results of Myer's index

5.3.3.2.1 National level

The inspection of one year age data of Census 2001 and Community survey 2007 using Myer's index showed a positive aspect in the findings. It can be seen that the values of Myer's index are below 5 which is rare to have in developing countries. However, the results pointed to a decrease of the index which signals a progressive improvement for both sexes. The male index was however better compared to females. The pattern of digit preference was overall more

pronounced among females than males. However, females were slightly attracted to numbers with digit ending 4, 6 and 7 while avoided to digit ending 5, 0, 9 and 3; hence males were attracted to ages with digit ending 7, 6, and 4 while repulsive to ages with digit ending 5, 9, 3 and 0. In census 2001, males and females who obtained the same result, were avoided to digit ending 0, 5 and 3 while attracted to digit ending 1, 9 and 8. The quality of reporting age was more accurately for the Community than for the Census 2001.

5.3.3.2.2 Provincial level

The assessment of Census 2001 and Community Survey 2007 at provincial level was made using Myer's index indicated a slight improvement observed in more than eight out of nine provinces for male respondents. In general, the value of Myer's index was less than 5 which was more close to 0 than 90 the full range. However, Myer's index in both instruments reached the highest value in the population Census 2001 at only 4.12 for males in Eastern Cape while in the Community Survey 2007; the highest value is observed in Eastern Cape and Northern Cape (3.58) for males. Hence, for females, Census registered the highest in Limpopo (4.45) while Community Survey indicated Eastern Cape (4.75). These values were considered very small compared to full range of this index which varies from 0 to 90. These indexes confirm that age heaping at age ending 0 and 5 in Census and Community Survey in South Africa is insignificant. Eastern Cape, Northern Cape and Limpopo are provinces with rural predominance. A possible reason could be some cultural factors which characterized the attitudes of those living in these areas. Only male respondents of Northern Cape out of 9 provinces registered an increase of their Myer's index. The deterioration in the reporting of age revealed might be a non-confidence in the government showing that the populace has lost in trust in government promises. This was clearly highlighted when one could see in the engagement to revise the questionnaire, organization of the collection process over years. Even the problem of poverty which was severe in these particular provinces with high indexes indicated one reason of non-accessibility to education. Not all families were able to provide education to its members.

UNIVERSITY of the WESTERN CAPE

5.3.3.2.3 Cross-ethnic group variations

An attempt was made in terms to evaluate Census 2001 and Community Survey 2007 using Myer's index, the findings indicated the improvement observed in the quality of responses on age over years in all population groups except Whites and female Indians who realized an increase of their indexes. However, the improvement observed could be explained by all efforts the government engaged in different areas such as publicity, amelioration in training programs, and revision of questionnaires. After six years, the change was visible in such a way that the indexes obtained at national level were higher than those registered in Coloured, Indian/Asian and White population groups. Despite the fact that African/Blacks showed the highest indexes, at least these indexes decreased, the quality of their distribution on age was likely accurate. According to Census 2001, White population presented best indexes and considering Community Survey 2007, White females registered the best index despite the deterioration observed in the quality of their declaration on age. Census 2001 and Community Survey 2007 seemed to cope with the White population group. However, Census 2001 was likely more reliable for White population group compared to others while male Whites tended to be more comfortable than female Whites. Regarding the Community Survey 2007, it revealed that it was probably better designed for males compared to females. Despite the existence of the above mentioned phenomena, the age-specific Myer's index in 2007 yielded the highest value at 3.26 (within the possible range of 0 to 10). It can be concluded that age preference in responses to the population Community Survey in South Africa were not a concern.

5.3.3.3 Results of the combined index of United Nations

5.3.3.3.1 National level

In attempting to evaluate the CIUN at national level in Census 2001 and Community survey 2007, the results of the study indicated a slight improvement in the quality of data which revealed the index decreased at 2.52 points showing the amelioration in the quality of responses. The migration patterns could be the principal reason for this feature. Census and Community survey were two different instruments which presented some particularities, that is, period during

which the census took place, how the collection was organize. The scope of both instruments was not the same.

5.3.3.3.2 Provincial level

An evaluation of the Census 2001 and Community Survey 2007 using CIUN was made to highlight the variations observations over years at provincial level. The CIUN explained how accurate the population structure by sex and 5-year age groups was. The findings revealed that the decrease of indexes appears only in three out of nine provinces namely Western Cape, KwaZulu-Natal and Mpumalanga. However, the slight improvement observed in KwaZulu-Natal and Mpumalanga did not provoke the change of category which meant the quality of its data distribution remained probably inaccurate. Only Western Cape managed to prevent the accuracy of its data distribution after six years. Statistics in Table 4.9 indicated that the values of CIUN for both instruments in South Africa were between 20 and 40 points except Western Cape, Free State and Gauteng with CIUN less than 20 for Census 2001. The highest value recorded was 30.13 (Limpopo) for the Census in 2001 and 33.58 (Limpopo) for Community Survey 2007. Evaluating the reporting on age, it revealed that Community Survey 2007 was likely more reliable for females compared to males.

UNIVERSITY of the
WESTERN CAPE

Census 2001 was probably better designed than Community Survey 2007. The great change in the attitudes of female respondents could be explained by the confidence acquired over years, the policy related to the education for all and the involvement in Labour force participation encouraged female respondents to be more expressive. The deterioration of male index could be seen as discouraging, or might be the misreporting. The missing of males could be the consequence of migration, hence young males moved in search of better life (employment or education). This feature of migration is age and sex selection. It might be the main reason of high value of index. Migration has a strong impact on CIUN at provincial level. However, the provinces with low CIUN show either small interprovincial migration or the age structure of in-migrants is very similar to the population structure of the destination province. In contrast, provinces with high values of CIUN have high interprovincial migration (either out-migration or in-migration). These provinces include Limpopo, North West, Eastern Cape, and Free State. The high levels of out-migration and in-migration disturbs the ordinary sex and age structures of the

population. Thus, high CIUN values in these provinces can be explained by the unusual sex and age structures of the population rather than discrepancies in age and sex reporting in the Census or Survey. However, Census 2001 was probably more reliable compared to Community Survey 2007.

5.3.3.3.3 Cross-ethnic group variations

In attempting to evaluate the ethnic similarities and differences of Census 2001 and Community Survey 2007 using CIUN, the findings indicated a very slight improvement observed in African/Black population. This could be explained by the application of awareness and advices enumerators received during the training stage. The improvement of African/Blacks highlights that their index did not change the category hence, the quality of reporting on age remains likely inaccurate. The increase observed in the values of CIUN Coloured, Indian/Asian and White population revealed deterioration in the quality of declaration on age by respondents. This could be explained by the high level of in-migration of Indian/Asian population with the influence of cultural factors. According to Whites, this could be explained by the drain brain observed in this population group due to the lack of security and safety across the country.

UNIVERSITY of the
WESTERN CAPE

5.3.3.4 Age ratios variations

The distribution of age ratios was applied to assess the quality of data derived from Census 2001 and Community survey 2007. The motivation of using age ratios was based on Arriaga (1994) and Chaurasia (2005) who both argued that the age ratios should be fairly similar across the age categories if there was no extreme variation in the past vital events. However, the results of the findings showed considerable fluctuations when comparing both curves of males' age ratios and also for females. The age range was 0 to 69 years (to minimize the effect of random fluctuation due to small numbers in the Survey and census, the age band 0-69 covered about more than 95 per cent of estimated population). The smallest frequency of males' age ratio had increased from 85.29 (at age 50) to 89.80 (at age 63) while the highest frequency had decreased from 121.38 (61 years) to 112.60 (54 years). Considering the females' curves, the smallest frequency had decreased from 86.23 (60 years) to 81.86 (65 years) and the highest frequency had also

decreased from 130.23 (61 years) to 119.85 (64 years). Some peaks and troughs were observed in these curves indicating preferences in age reporting (for a certato digit ending) and under reported ages respectively. The possible reasons to explain these patterns of age ratios could be due to missing information or insufficient information reported. In some cases, the low age ratios observed in previous and precedent cases could also explain the high age ratios. Migration is another possible reason especially in this case where two instruments of different sources are compared at different time period (2001 and 2007). In the same vein, the differential pattern of mortality observed in both instruments could be contributive to the explanation of mis-reporting in ages. However, the high age ratios observed in the working ages could be probably due to migration of people in search of new opportunities.

5.3.3.5 Sex ratios variations

An attempt was made to evaluate the sex ratios by age for the South African population according to Census 2001 and Community survey 2007. The outcomes of the analysis in the previous chapter confirmed the literature in which Chaurasia (2005) described that a clear age specific patterns can be seen in the sex ratios. For instance, the sex ratio revealed obviously in which age the sex ratio was favourable to males or to females. He also found that the female specific abortion and female infanticide tended to affect the size of the female population. Comparing both instruments, the analysis of sex ratios revealed an increase in deficit of females observed at ages 0, 1, 2, 3, 5, 6, 7, 9 and 11 while a deficit of males persisted at ages 4, 8, 10, 12 to 14. Low sex ratios in the ages under age 15 indicated under enumeration of children hence sex ratios should be high at such ages. The possible explanations could be the fact that generally females aged over forty tended to underestimate their age which could have shifted them into younger ages while the males overestimated their ages which shifted them into older ages. The high sex ratios could be mostly attributed to the low living standards and lack of education which lead similar mortality rates between males and females. However, the cohorts which had low sex ratio 6 years before still have low sex ratios 6 years later. The highest sex ratio had increased from 101.56 (at age 1) to 103.92 (at age 16) while the smallest had also increased from 59.31 (at age 68) to 72.36 (at age 66) indicating the slight improvement of the life expectancy. The concavities observed in working ages could be explained by the patterns of migration which was sex-selective. This pattern substantially increased the size of males for instance, in the case;

UNIVERSITY of the WESTERN CAPE

males migrated from rural to urban areas. The excess of females aged 40 and above might be due to biological phenomena and also the higher rate of mortality of males at the same ages. Community survey 2007 was likely more reliable than Census 2001. The quality of responses on age in 2007 was probably better accurate compared to Census 2001. It was difficult to ascertain whether some observed features in age and sex distribution were due to international migration or not. Migration data at national level could be available from census and surveys but the accuracy was in doubt. In this especially case of comparing age ratios and sex ratios of Census 2001 and Community survey 2007, migration could not be ignored and also the differential pattern of mortality should be considered. The government had to increase communication in terms of supporting women's status in the society and the improvement of the health sector will lead to reduced mortality rate.

5.3.3.6 Variations in population pyramids

UNIVERSITY of the
WESTERN CAPE

The comparison of the results from Census 2001 and Community survey 2007 was possible at national, provincial levels and per population groups in such a way it highlights similarities and differences observed in the quality of responses on age which influences the age and sex distribution of the South African population. However, even this comparison should present some limitations due to the sampling units used for selection, the responding units that provide the information and the units of enumeration and analysis in which information was sought in the survey might be different and difficult to compare.

An attempt was made to assess the age-sex distribution of the Census 2001 and Community Survey 2007 per population groups. The pattern of age–sex distribution in the country varied with population group as indicated in both instruments. The population pyramid of Africans was similar to that of South Africa as a whole with wide base which revealed poor living standards, high fertility and high infant mortality rates which led similar mortality rates between males and females. A surge of youth, many to poor families with limited prospects for good education was making the challenge of ramping up for instance public education to the right level. This merely reflected the dominating influence of the African population on the overall population of the country. The younger population had higher proportion in the younger ages and decreasing proportions with increase in age. However, the variation observed in young dependency ratio

indicated a slight deterioration in this population group. The ratio moves from 56.5 to 51.15 which highlight a slight decrease in fertility. This is contradicting with the sex ratio at birth which increases from 98.62 to 102.77. A possible explanation could be the misreporting of birth in some rural areas where there is a lack of vital registration system during Census 2001. Age–sex distribution of Coloured was very close to the African with the similar shape showing expansive pyramid. The decline observed in working age could reflect HIV/AIDs deaths which was affecting the productive age. The highest proportion of children and the lowest proportion of working age population indicated multiple needs to provide with limited resources (schools, jobs, houses, health and social facilities…). This population group also highlights a decrease in sex ratio at birth and in young dependency ratio. This could reveal the legacy of abortion in South Africa which impacts the rate of fertility. However, the high dependency ratio indicates the great need of support.

The White population in South Africa presented a very different age pattern. The pyramid for the White population is bell-shaped, typical of a developed country with fewer children and more elderly people. Whites had the smallest proportion of their population under 30 and the highest proportion for ages over 40. The increase of elderly in this population group needs provision of social services. The brain drain observed in working age revealed the potential need of skilled workers to maintain performance in the economy of the country. Asian population distribution was closer to that of the White population. The variation observed in the proportion of dependency ratios and in the sex ratios at birth per population group revealed some features in fertility and mortality such as the sex ratio at birth confirmed irregularities in the age and sex structure. The dependency ratio of under age 15 indicated the level of fertility in each group which could be verified through the living conditions per population group.

UNIVERSITY of the
WESTERN CAPE

The gender imbalance could reflect in the social and financial well-being of older women in the female population group compared to older males in the total population groups. Considering both instruments, the missing of young males in African/Black and Coloured population groups could be the result of out-migration into neighbouring countries in search of better employment. According to Indian/Asian, the increase of number could be seen as the effect of in-migration which is more expressed in this community. However, African/Blacks showed high dependency ratio, followed by Coloured population group, than Indians and Whites. For the ageing

population, one could see a positively skewed distribution towards higher ages with higher peaks while for the younger populations, one could see a less positively skewed distribution with lower peak.

Various aspects of the age-sex distribution estimated by Statistics South Africa remained challenging. In attempting to bring more precisions in these issues, this study evaluated the age and sex distribution from 2001 to 2007 using indirect methods (Whipple's index, Myer's index, CIUN, age ratios, sex ratios and population pyramids) and also initiates new approach to adjust age-sex assessment as a means of evaluating age-sex distribution from the Census 2001, Community Survey 2007, GHS 2004, GHS 2007, LFS 2005 and LFS 2007, and future Censuses or Surveys across the country. Planning and policy decision can be impaired if information about the age-sex distribution of the population was inaccurate. It was important to appraise age-sex distribution from a Census or Survey to accurately inform planning and decision making.


UNIVERSITY of the
WESTERN CAPE

CHAPTER SIX

Conclusion and Recommendations

6.1 Conclusion

Based on data from the Census 2001 and Community Survey 2007, General Household Surveys 2004 and 2007, Labour Force Surveys 2005 and 2007, the statistics on the quality of age and sex population of South Africa were examined. This section of the thesis summarizes the overall outcomes from the research relating to this study and how it was investigated through variables such as gender, age, population groups, and provinces.

These sources of demographic data reveal some relevant features regarding coverage errors and content errors. However, the coverage errors include omission or duplication resulting in under or over enumeration occurrence, non-reporting or misclassification of age which contributes to age heaping. Also included in this are errors arising from the following: difficulties in communication, inaccessibility and cooperation with respondents; lack of, boundaries descriptions and change of addresses. The content errors are considered where characteristics such as age, sex, population group, provinces and so on of a respondent in a census or survey are incorrectly reported or tabulated. The two possible sources of content errors are when, a respondent provides a wrong response and an enumerator records an incorrect response.

UNIVERSITY of the
WESTERN CAPE

In each of these instruments (Census 2001 and Community survey 2007, General Household Surveys 2004 and 2007, Labour Force Survey 2005 and 2007), questions which include “date of birth” or “completed number of years,” may result in different ages. The study has applied some indirect methods following Whipple’s index, Myer’s blended index, Combined Index of United Nations, Age ratios, Sex ratios, and Population pyramids which have been discussed in previous chapters.

The overall objective of the study was to describe and compare age and sex structure tools such as cross tabulations, graphs or figures and population pyramids. Different features have been discussed. The above contributive factors are of significance to the study as they assist in attempting to answer research questions in the first chapter of this study. By comparing the

results from analysis of the Census 2001 and Community survey 2007, GHS 2004 and 2007, LFS 2005 and 2007 data, the variations have been presented. The study evaluated the set of data without adjustments.

The theoretical framework was drawn from the literature on sustainable framework. The study first evaluated the GHS 2004 and GHS 2007. The examination made use of Whipple's index, Myer's index, CIUN, age ratios, sex ratios and population pyramids at national, provincial level and per population groups. At national stage, considering GHS, it was found that in 2004, the males' Whipple's index was 104.54 and this has increased to 106.86 in 2007 while females' Whipple's index also increased from 105.69 to 108.18 for the same years; both indexes revealed deterioration in the quality of reporting of age. With regard to provinces, the highest Whipple's index for males was found in KwaZulu-Natal, followed by Gauteng and Eastern Cape while for females, it was found in KwaZulu-Natal, Free State, Mpumalanga and Gauteng.

Considering the population groups, the highest index for males was found in Indian/Asian population group followed by Whites and African/Blacks. For females, the highest index was found in Indian/Asian, followed by Coloureds, African/Blacks and Whites. According to Myer's index, the study also evaluated at national, provincial and inter-ethnic variations. It was found that the value of index for males and females was low in general showing that progressive improvement was observed in 2007. Myer's indexes have confirmed that age heaping at ages ending with 0 or 5 in the GHS population in South Africa is insignificant hence its values are very low compared to the full range of this index which varies from 0 to 90.

These findings support the specific research question in which the differences in the quality of data on age and sex for provincial and national levels are gauged. However, African/Blacks females indicated a small Myer's indexes in 2007 than other categories while Coloureds also presented for males. The highest index was observed by Whites, followed by Indian/Asian and African/Blacks for males in 2004 while Indian/Asian followed by African/Blacks and Whites for females in the same year. In 2007, the highest index occurred in Indian/Asian, Whites and African for males while Indian followed by Coloureds and Whites for females. However, Whipple's and Myer's indexes had confirmed that age heaping at ages ending with 0 or 5 in the GHS in South Africa is insignificant for males in Western Cape, Northern Cape, North West (also females in this province), Mpumalanga and Limpopo with both surveys.

With regard to GHS 2004 only, both indexes confirmed the insignificance of age heaping at ages ending with 0 or 5 for males at national level in Free State while for females include Western Cape, Northern Cape and Limpopo. The Females' index of Mpumalanga in GHS 2007 also confirmed this insignificance. The patterns of preference which are highlighted by the education, cultural and political factors are observed in both surveys. The quality of reporting on age depends on the education of the respondents; the cultural factor is seen as an obstacle and the political affiliation also. The results of the population groups reinforced the trend hence Indian/Asian dominated KwaZulu-Natal, Whites in Gauteng province and African /Blacks dominated all the provinces with rural area predominance. As the literature indicates, it is true that it is more useful to make questionnaire more understandable to reduce errors in the responses. This means that though more females than males are less educated, males are likely to report their age better compared to females. In the same vein, in regard to the culture, females cannot easily express their opinions because they are undermined by males (less consideration in the society). In addition, the direct consequence of the end of apartheid is highlighted by a White population group that is not confident with the new South Africa government. The study concludes that education, cultural and political issues play a crucial role in the quality of declaration of age and sex though they have limited the quality of reporting on age.

UNIVERSITY of the
WESTERN CAPE

Furthermore, the findings on the quality of reporting on age using the CIUN suggest that for GHS, the effect of migration may be a factor that helps respondents in a manner of having been missing in the age-sex distribution of the population. Apart from the consequence of apartheid, the CIUN is also affected by migration. The feature of migration is age and sex selection hence the majority of migrants are young people and males. This selection affects sex and age structures of both destinations. The CIUN by province is much higher than at the national level. The provinces with the highest CIUN include Mpumalanga, Northern Cape, Free State and Eastern Cape. Low CIUN value show small inter-provincial migration or the age structure of the in-migrants is very similar to the population structure of the destination province. In contrast, provinces with high values of CIUN have high inter-provincial migration (either in-migration or out-migration). These include Gauteng, North West, KwaZulu-Natal, Limpopo and Western Cape. The high level of in or out-migration perturbs the usual sex and age structures of the population. The results of CIUN per population groups confirmed Indian/Asian and Whites with highest indexes as people are permanently mobile. These high values of CIUN are much

explained by the unforeseen sex and age structures of the population rather than discrepancies in age and sex reporting in the GHS.

According to data on age and sex, the evaluation of its quality points out some features such as age heaping (underestimation of females and overestimation of males), over-age reporting for elderly people, under reporting of children under age five, missing of young working age from 20 to 34. Through analyzing the statistical data of GHS on age patterns, it can be deduced that generally the deficit of males' population reported in young adults ages as revealed by the age ratio are due to the out-migration. There are some disparities with regards to the ages, when analyzed the ratios of males show surplus at ages 7, 10, 14, 16, 18, 25, 30, 32, 40, 42, in both surveys while for females appear at ages 7, 10, 14, 30, 32, 40, 42, 50, 52, 60, 62, 65. The surplus is detected in the same age for both sexes and also the issue of rounding ages is observed with some similitude in ages 10, 14, 16, 18, 30, 32, 40, 42, 50, 52, 60 and 62. It is more pronounced for females than males with their highest values. The deficit of male population in young adult ages as indicated by the age ratio analysis may also be due to differential omissions of respondents of a specific age or change in the levels of fertility and mortality, or due to misreporting in the age information. The GHS 2007 was more reliable than the GHS 2004 in regard to age's assessment. The study concludes that the quality of responses on age was inaccurate for both surveys.

UNIVERSITY of the
WESTERN CAPE

Like with other aspects of South African demography, sex patterns, preference patterns, migration patterns, fertility and mortality patterns presented in the section of sex ratios assessment show variations, by gender, per age. However, the findings indicated a low sex ratio at birth, an under enumeration of children (under 15) and a shortage of males in the working ages for both GHS. The results revealed that GHS 2004 was reliable than GHS 2007. The increase observed in sex ratios is probably due to the decrease in the fertility rate and increase in mortality rate, or due to labour migration, or due to biological factors which results in an excess of male births. The unusual shortage of males is maybe due to migration stream (sex selective), or errors in the enumeration reports or preference for sons. There is a need to sensitize the health workers regarding the issue of declining sex ratio (through training) so that they can sensitize the Community.

Considering the population pyramids at national level using GHS indicates an expansive pyramid with high birth rate, low health standards of living and high mortality rate at child and early adult. The level of international migration is unknown in South Africa. The policy implications are important. There is a need to improve the supply of education, health, employment and other facilities for the population. For instance, public health policies must be revised in terms to reduce communicable infections and improve the lifestyle of people.

This study secondly evaluated the LFS 2005 and LFS 2007. The inspection made use of Whipple's index, Myer's index, CIUN, Age ratios, Sex ratios and Population pyramids. Regarding Whipple's index, the results for LFS at national level indicated that there was improvement. As mentioned above, these results were somewhat relatively accurate, very close to the limit 105. There was a decrease to digital preference from 2005 to 2007 among males and females. A possible reason might be the fact that Statistics South Africa is to some extent not constant in terms of improvement in the quality of its services. It has been changed or revised many times that might have an effect on the collection and reporting of age and sex in this country. If one system or method would be figured out to work better, somehow there might be a positive effect even on Whipple's index at national level.

UNIVERSITY of the
WESTERN CAPE

Gender disparity is also one of the factors that need to be looked into in terms of what can be done to improve the quality of declaration on age. Generally, females are affected more in provinces with rural predominance than those with urban predominance, with regard to the participation in Labour force. However, the study could not dwell on urban or rural areas. Nonetheless, literature indicates that there are inequalities between these areas. Females in rural areas face more difficulties in their participation in Labour force; they are affected by socio-cultural economic factors compared to those in urban areas. Furthermore, at provincial levels, the results confirmed the hypothesis. This is not surprising as the data indicates that there is a decrease in females' index than for males in most of the provinces. This indicates that policy implementation is more targeted at females than males. The persistence of preference in some provinces such as Free State, Northern Cape, KwaZulu-Natal and Mpumalanga, underlines all the shortcomings in reaching the objectives, and more efforts to reduce the vulnerability mechanism in attraction in ages with a certato digit ending.

Overall, the study has attempting to answer the research question, "How ethnicity differences affect the quality of declaration on age and sex?" As the results revealed, there were disparities in attainment across the population groups. The disparities indicate that there is improvement in Indians and Whites males, African population group also have decreased while the Indians, White females and Coloureds males are vulnerable. These results support the deterioration of indexes in Whites and Indians as this increased the pressure on Whites to enter cash economy and migrate to foreign lands for new opportunities. Both political and economic attitudes of Whites changed radically. In 2004, the study found that the reporting on age in Coloureds population group was of high quality hence in 2007, the reporting on age was of the high quality in female Whites and male Coloureds. Therefore, one can conclude that there are still preferences in the declaration on age and sex with regard to population groups.

The assessment of LFS using Myer's index at national level was attempted. The study concludes with an improvement of indexes over years per gender. This has confirmed the hypothesis in which "gender differences are associated with the poor quality of declaration on age and sex. Government has to emphasize more actions using information on the importance of surveys by multiplying campaign vulgarization on the quality of data. Apart from the issues relative to the birth certificates in many provinces, age and sex selection are features of migration. These features could be the possible reason of the higher value of Myer's index at provincial level. This was observed in the Western Cape, Eastern Cape, Northern Cape, Free State, North West, Mpumalanga and Limpopo. There is a need to highlight policies to stop the raising sex ratio by promoting female importance towards development in society. The results of the analysis per population groups indicated that the patterns of preference are recurrent factors in Indian communities. The government needs to promote education at all categories of age in order to break down those cultural barriers among this population in sense that every group of age must fill more confident in declaring its ages. The evaluation of LFS using CIUN at national level was attempted. The results revealed that the shift of females from older ages to younger ages and for males from younger ages to older ages could be a possible reason of the distortion observed in the age and sex structure of data.

The out-migration in neighbouring countries might be another reason to this increase of CIUN. Government has to initiate more effective actions to create new opportunities with efficient

UNIVERSITY of the WESTERN CAPE

environment. The results of CIUN at provincial level showed the patterns of preference greater among males than females in both surveys. A possible reason could be that the attitudes of respondents changed over years, females became more open while males tend to be more reserved. There is a need to encourage both sexes to improve the quality of their reporting on age. Exploring LFS, CIUN was used to evaluate the quality of declaration on age among population groups in South Africa. The results revealed the female population was more pronounced in preference compared to males in 2007 except African/Blacks. The age reporting was likely greater among females in LFS 2005 than males while the reverse was observed in LFS 2007. The lack of education, non-consideration of female in the Indian society, inequality observed in African/Blacks and the uncertainty of the political situation which affects the attitudes of Whites, contribute to impeding the CIUN. Consequently, increased insecurity and crime become significant in the country. Investigation was attempted through LFS using age ratios per year at national level. Apparent preference was observed in the data, deficit of males reported in the young adult ages which might be due to the out-migration of the long term male migrants. The age ratios showed better age reporting for females than for males. Some omissions could also be the reason of distortion in the structure of data. These features contributed to the deterioration of the quality of data on age.

UNIVERSITY of the
WESTERN CAPE

Referring to the analysis of sex ratios, the findings indicated that the sex ratio pattern in LFS followed the type of developing countries except the deviation occurred in ages 3, 6 and 12 for LFS 2005 and in ages 0, 1, 3, 5, 8, 11 and 13 for LFS 2007. At these ages, the sex ratios were supposed to exceed 100 as the case in developing countries. The quality of the structure of these data was not accurate, but the LFS 2007 was more reliable compared to LFS 2005. The government and NGOs need to conjugate efforts in such a way that female can be considered equally in the society as male. The application of population pyramid per year was attempted to evaluate both LFS. The findings revealed the shape was different but both surveys showed the characteristics of developing countries with poor condition of life, high infant mortality and fertility rates. The shifts of the younger and older ages explained the increase observed in the working age population between LFS 2005 and LFS 2007. This could be the result of the improvement in the adult mortality and fertility rates. The variation of these factors combined to migration has serious implication in the policy of the country. For instance, the improvement of health has to consider all aspects related. The high total dependency ratio observed in both

surveys implies that the government has to invest in education, health, manpower, housing and social requirements for achieving good result for national development.

Furthermore, the study thirdly evaluated Census 2001 and Community survey 2007. The findings indicated at national level using Whipple's index the increase of indexes over years. Despite this deterioration, the quality of declaration of respondents remains very accurate for males and females. The preference pattern was more pronounced among males than females in the Census 2001 while the reverse was observed in the Community 2007. Considering Myer's index, the results showed an improvement for both sexes. The slight preference is more pronounced among females compared to males. However, Community 2007 was more reliable than Census 2001. Males and females were attracted to ages with digit ending 4, 6 and 7 while avoided in ages with digit ending 0, 3, 5 and 9 in the Community Survey 2007. Both sexes were attracted to numbers with digit ending 1, 8 and 9; avoided in 0, 3 and 5 regarding the Census 2001. The CIUN was applied to evaluate Census 2001 and Community survey 2007. The findings showed a decline observed in the index. The slight improvement indicated the amelioration in the quality of responses.

Knowing that Census and Community are two different instruments with each particularity, migration patterns might be the important reason to explain the change observed. Using age ratios, the quality of data on age distribution was evaluated for the two instruments. The results revealed some peaks indicating preferences in the age reporting (for a certato digit) and some troughs showing under reported ages. The missing information or insufficient information might be possible reason or the observation of low age ratios in previous and precedent might also have contributed to the high age ratios.

Furthermore, migration could also clarify this situation. However, the comparison of two instruments with different sources, at different time period (2001 and 2007) resulted in difficulties because they follow different patterns of mortality which contribute to explain the mis-reporting in ages. Considering the sex ratios, Census and Community were evaluated. The findings revealed that the females' abortion and infanticide affect the size of the female population. The low sex ratio are observed in ages under 15 showing under enumeration of children hence sex ratios must be high at such ages. The concavities in working ages are the result of migration patterns (sex selective). The excess of females aged 40 and above could be

due to the higher rate of mortality of males at the same ages. The government needs to reinforce communication in order to highlight women's status in the society and ameliorate the programme of health as it will contribute in reducing the mortality rate. The lack of data on region and population groups was observed for both instruments. It made it impossible for the comparison of population pyramids for both instruments.

However, this study attempted to evaluate Census 2001 using population pyramid according to population groups. The findings revealed the population pyramid of African/Blacks with a broad base of young and a small apex of old. Government needs to increase schools, jobs, health facilities and houses. The age-sex distribution of Coloureds is close to the African/Blacks, but with different age pattern. The population pyramid of Whites revealed lower proportion of children below the age 15, a higher proportion of population in ages 15-64 and 65 and above; more people are eligible for social security. To reduce the issue of drain brain observed in White population pyramid, the government needs to ensure safety and security of life.

In addition, this research explored the SA data quality on two variables, age and sex. It seemed that the evaluation of age and sex distribution of the South African population arouse a great concern about the variables especially when it is using the Post Enumeration Survey (PES). The researcher started with interviews through emails and face to face interaction in order to apprehend the method used by Statistics SA to assess the quality of its data collected from censuses and surveys. During the interview, the researcher asked the respondents how they handled the non-responses, the missing information on age, the differences observed in method of data collection in censuses, the GHS, and LFS; when and where the adjustment took place... As the SA has been applying the PES, the researcher decided to apply the indirect method to see if there were similarities or differences in the results. The results revealed some inconsistencies in the quality of data on age and sex in South Africa.

Indeed, various features of the age and sex distribution estimated by Statistics SA remain controversial. In attempting to shed more light on these issues, the study evaluated the age and sex distribution from 2001 to 2007 using indirect methods (Whipple's index, Myer's index, CIUN, age ratio, sex ratio and population pyramids). It also introduced new approach or method of "legal age" to adjust age-sex distribution from the 2001 census 2004 GHS, 2005 LFS, 2007 GHS, 2007 LFS and 2007 Community survey, and future censuses and surveys across the

country especially in case of logical imputation. However, the "legal age" was not applied in this study and no adjustment was done. Planning and policy decision can be impaired if information about the age-sex distribution of the population is inaccurate. It is important to appraise age –sex structure from a census or a survey to accurately inform planning and decision making. For instance, it was observed that there was an incomplete coverage of respondents in certain age groups in a census or survey according to education; this can result in a particular population sub group being underestimated in population projection, which may affect adequate allocation of resources in planning for this sector. In the same vein, the incomplete coverage of respondents in certain age group perhaps 20 to 24 in a census or survey, if not corrected could lead probably to wrong conclusions about the impact of HIV / Aids which could influence planning in the health sector. It was assessed during the interview that the age-sex in consideration of indirect methods had never been undertaken (considered). New approach was developed to adjust the existing data as a means of evaluating age-sex distribution for future censuses and surveys in the country.

In exploiting PES for the census purpose, checking a sub sample of households after do household surveys compare the results for selected questions, SA is relaying only for the checking. This is not enough despite this practice; the researcher detects mistakes by just relaying on these indirect methods, there is no insurance of the quality hence some attractions and avoidance are still there. The responsibility of the government should be to build population policy and design policy measures to help officials cope with the problems arising from a society's particular fertility pattern, mortality pattern, migration pattern and so on. It is important to distinguish between those behavioural and biological factors that have a direct impact on the quality of data on age and sex, and those socio-political, economic, and cultural factors that affect the quality of data on age and sex only indirectly through the proximate determinants. Information of the relationship between the direct and indirect fertility determinants allows a clear perception of specific opportunities for effective policy interventions. A young woman's education is one key socio-economic variable that has a great impact on the intermediate variables and on fertility. It should therefore be considered seriously in fertility related policies.

UNIVERSITY of the
WESTERN CAPE

6.2 Recommendations

The study has shown that the preference among gender is a persistent phenomenon which prevails the inaccuracy of data quality according to the Census 2001, Community survey 2007, both GHS and both LFS. Urgent policy issues should therefore be addressed if the South African vision on quality data prevalence on age and sex is to be achieved. The following are some recommendations that can be implemented, monitored and evaluated through relevant policy interventions.

1. All births should be registered under the relevant birth certificate. Therefore, government has to initiate a premium on the birth of a girl child (incentives, recognition).

2. Self-motivated pregnant women should not be entertained for sex selective procedures, come what may.

3. It is necessary to carry out more stringent policies to prevent the increasing sex ratios at birth. However, South Africa needs not only to prevent but also to cope up with the situation of surplus male population in the near future. Maybe it will be better to examine the experience of Taiwan, China to help in setting up appropriate policies.

UNIVERSITY of the
WESTERN CAPE

4. Different reasons are given for preference of a son because the son is needed to sustain a family tree. This is a common excuse. Knowing that son preference is the main reason in declining sex ratio, there is a need to create awareness in the community about change in the mentality of the people by religious leaders concerning this issue. It is important to examine the correlation between economic, cultural and social determinants and population issues for design of appropriate policies.

5. High quality information facilitates guidelines, understanding of best practice and ensures that resources are properly targeted. Likewise, province-based sensitization workshops should be organized in consultation with local NGOs and leader groups (parents) to provide education in general about the importance of data quality, including feedback about relevant data quality issues.

6. Changing existing gender norms can improve the quality of declaration; involving gender into different programs can be an opportunity that addresses the specific needs of girls, educating young women on social issues, encouraging them to become more involved in making decision, to freely express themselves.

7. South Africa population has been experiencing changes in size, quality and structure of data. These changes vary among gender, across provinces as well as population groups. Policies on population and development should be flexibly established and deployed to correspond to the different demographic patterns in South Africa, depending on gender, province, ethnicity, geographic and other factors.

8. To achieve to break down cultural background in the Indian population group, there is a need for government agencies and NGOs to strengthen teacher training, introduce parent education and community outreach programs. It is also important to establish dialogue on the cultural constructions that shapes gender norms in the community. Many prevailing society norms negatively affect Indians to access modern knowledge information.

The application of both methods (PES and indirect) to assess the quality of data could become operative and significant. However, the researcher would also have liked to combine these methods with the direct method of "legal age" especially as the logical imputation was required. This will help to reduce gap and reach a level where it can provide more relevant quality data. Therefore, a research combining PES, direct and indirect methods can be carried out in the future.

UNIVERSITY of the WESTERN CAPE

References

Adam, A., 2007. "Data Quality Issues in Surveys of the International Comparison Program for Africa (ICP-Africa)". *The African Statistical Journal,* vol. 5, pp. 49-67.

Amann, L.C., Phillips, J.M., Halene, T.B. & Siegel, S.J., 2008. "Male and female mice differ for baseline and nicotine-induced event related potentials.". *Behavioral Neuroscience; Behavioral Neuroscience,* vol. 122, no. 5, pp. 982.

Anderson, M. & Fienberg, S.E., 2002. "Why is there still a controversy about adjusting the census for undercount?". *PS-WASHINGTON-,* vol. 35, no. 1, pp. 83-84.

Andrews, P.E., 2007. "Big Love-The Recognition of Customary Marriages in South Africa". *Wash.& Lee L.Rev.,* vol. 64, pp. 1483.

Ansari, M., 2008. "Economic Growth, Rural Poverty and Human Resource in Bhutan: A Description of the Linkages in the Eastern Himalayas". *Saarc Human Resource Development Centre (SHRDC),* pp. 87.

Arnab, R. & Serumaga-Zake, P., 2006. Orphans and vulnerable children in Botswana: The impact of HIV/Aids. Vulnerable children and youth studies. 1. 221-229. Taylor & Francis.

Arriaga, E.E., Johnson, P.D. & Jamison, E., 1994. *Population analysis with microcomputers,* Bureau of the Census.

Austad, S.N., 2006. "Why women live longer than men: sex differences in longevity". *Gender medicine,* vol. 3, no. 2, pp. 79-92.

Bangha, M.W., 2008. "Estimating Adult Mortality in Cameroon from Census Data on Household Deaths: 1976-1987". *African Population Studies,* vol. 23, no. 2, pp. 223-247.

Banister, J., 2004. "Shortage of girls in China today". Journal *of Population Research,* vol. 21, no. 1, pp. 19-45.

Becht, M., Bolton, P. & Röell, A., 2003. "Corporate governance and control". *Handbook of the Economics of Finance,* vol. 1, pp. 1-109.

Bhagat, R.B., 2008. "Assessing the measurement of internal migration in India", *Asian and Pacific Migration Journal,* vol. 17, no. 1, pp. 91.

Bongaarts, J., Büttner, T., Heilig, G. & Pelletier, F., 2008. "Has the HIV epidemic peaked?". *Population and Development Review,* vol. 34, no. 2, pp. 199-224.

Brass, W., 1996. "Demographic data analysis in less developed countries: 1946–1996". *Population Studies,* vol. 50, no. 3, pp. 451-467.

UNIVERSITY of the WESTERN CAPE

Chaurasia, R A., 2005. The age and sex structure of tribal population in central of India. *Studies in Population and Development.* Shyam institute, *India vol. 1.*

Citro, C.F. & Kalton, G., 2007. "Using the American Community Survey: Benefits and Challenges, Panel on the Functionality and Usability of Data from the American Community Survey". National Academies Press, Washington, DC.

Clark, G., 2008. *A farewell to alms: a brief economic history of the world.* Princeton University Press, Princeton.

Cleland, J., 1996. "Demographic data collection in less developed countries 1946–1996". *Population Studies,* vol. 50, no. 3, pp. 433-450.

Crayen, D. & Baten, J., 2006. "Numeracy, Inequality, Age Heaping, and Economic Growth: New Estimation Strategies for Western Europe and the US (17th-19th centuries)". *University of Tuebingen. Working Paper*.

Cutler, D.M., Deaton, A.S. & Lleras-Muney, A., 2006. "The determinants of mortality". *NBER Working Paper No. 11963*.

Dandona, R., Dandona, L., Kumar, G.A., Gutierrez, J.P., McPherson, S., Samuels, F. & Bertozzi, S.M., 2006. "Demography and sex work characteristics of female sex workers in India". *BMC International Health and Human Rights,* vol. 6, no. 1, pp. 5.

Dekker, A., 2001. "Adapting new technologies to census operations". *Symposium on Global* Citeseer.

WESTERN CAPE

De Moor, T. & Van Zanden, J.L., 2010. "Girl power: the European marriage pattern and labour markets in the North Sea region in the late medieval and early modern period1". *The Economic History Review,* vol. 63, no. 1, pp. 1-33.

Diane, P. C., 1996. *Platform for Action and the Beijing Declaration.* DIANE Publishing Company.

Docquier, F., Marfouk, A., Salomone, S. & Sekkat, K., 2012. "Are skilled women more migratory than skilled men?". *World Development,* vol. 40, no. 2, pp. 251-265.

Ebenstein, A., 2007. *Fertility choices and sex selection in Asia: Analysis and Policy,* Unpublished article edn.

Elmhirst, R. & Resurreccion, B.P., 2008. "Gender, environment and natural resource management: New dimensions, new debates" in *Gender resource management: livelyhoods, mobility and interventions.* eds. R. Elmhirst & B.P. Resurreccion, Earthscan, Lomdon, pp. 3-22.

Ewbank, D.C., 1982. "The sources of error in Brass's method for estimating child survival: the case of Bangladesh". *Population Studies,* vol. 36, no. 3, pp. 459-474.

Feskens, R., Hox, J., Lensvelt-Mulders, G. & Schmeets, H., 2006. "Collecting data among ethnic minorities in an international perspective". *Field Methods,* vol. 18, no. 3, pp. 284-304.

Firebaugh, G., 1979. "Structural determinants of urbanization in Asia and Latin America, 1950-1970". *American Sociological Review,* pp. 199-215.

Franks, P., Muennig, P., Lubetkin, E. & Jia, H., 2006. "The burden of disease associated with being African-American in the United States and the contribution of socio-economic status". *Social science & medicine,* vol. 62, no. 10, pp. 2469-2478.

Frøen, J.F., Pinar, H., Flenady, V., Bahrin, S., Charles, A., Chauke, L., Day, K., Duke, C.W., Facchinetti, F. & Fretts, R.C., 2009. "Causes of death and associated conditions (Codac)–a utilitarian approach to the classification of perinatal deaths". *BMC pregnancy and childbirth,* vol. 9, no. 1, pp. 22.

Graham, J., Lewis, J. & Nicolaas, G., 2006. "Ethical relations: A review of literature on empirical studies of ethical requirements and research participation". *ESRC Research Methods Programme Working Paper No 30, Manchester: Economic & Social Research Council*.

Groves, R.M., 2004. *Survey errors and survey costs.* Hoboken, NJ. Wiley-Interscience,

Guilmoto, C.Z., 2009. "The sex ratio transition in Asia". *Population and Development Review,* vol. 35, no. 3, pp. 519-549.

Gunasekera, R.C., 2009. *An Evaluation of Age- Sex Data of Census of Population and Housing 2001- Sri Lanka. A* paper presented at the 24th Population and Census Conference, Hong Kong, China 25 – 27 March 2009.

Gupta, M.D., 2006. "Cultural versus biological factors in explaining Asia's "missing women": response to Oster". *Population and Development Review,* vol. 32, no. 2, pp. 328-332.

Haslett, S. & Jones, G., 2008. *Potential for Small Area Estimation and Poverty Mapping at Constituency and at Gewog/Town Level in Bhutan.* World Food Programme, New Zealand.

Heron, M., Sutton, P.D., Xu, J., Ventura, S.J., Strobino, D.M. & Guyer, B., 2010. "Annual summary of vital statistics: 2007". *Pediatrics,* vol. 125, no. 1, pp. 4-15.

Hobbs, F., 2004. "Age and sex composition" in *Methods and materials of demography.* eds. D. Swanson, J. Siegel & H.S. Shryock, 2nd ed. Elsevier Academic Press, San Diego, CA, pp. 125-173.

UNIVERSITY of the WESTERN CAPE

Hobbs, F., 2003. "Age and sex composition". *The Methods and Materials of Demography,* pp. 125.

Jdanov, D.A., Jasilionis, D., Soroko, E., Rau, R. & Vaupel, J.W., 2008. "Beyond the Kannisto–Thatcher Database on Old-Age Mortality: An assessment of data quality at advanced ages". *MPIDR Working Paper WP 2008-013, Max Planck Institute for Demographic Research*.

Jones, E.C., Murphy, A.D., Perilla, J.L., Pérez-Vargas, I. & Norris, F.H., 2010. "Design for Data Quality in a Multisite Cross-Sectional and Panel Study". *Field Methods,* vol. 22, no. 3, pp. 250-269.

Joubert, J., Rao, C., Bradshaw, D., Dorrington, R.E., Vos, T. & Lopez, A.D. 2012. "Characteristics, availability and uses of vital registration and other mortality data sources in post-democracy South Africa". *Global Health Action,* vol. 5.

Khalfani, A.K., Zuberi, T., Bah, S. & Lehohla, P.J., 2005. "1. Population Statistics". *The Demography of South Africa,* pp. 3.

Khan, N., 2010. *An assessment of retrospective birth history reporting for the measurement of fertility in South Africa.* University of KwaZulu-Natal.

Kiregyera, B., 1982. "On sampling frames in African censuses and surveys". *Journal of the Royal Statistical Society, Series D (The Statistician),* vol. 31, no. 2, pp. 153-167.

Kiregyera, B., 1982. "Some Sources of Nonsampling Errors in African Census and Survey Work: A Review". *International Statistical Review/Revue Internationale de Statistique,* pp. 305-316.

Leggieri, C.A., 1994. "Census Bureau Plans To Build A Master Address File". *Proceedings:... Annual Research Conference.* US Department of Commerce, Bureau of the Census, pp. 3.

Lester, P., Peterson, K., Reeves, J., Knauss, L., Glover, D., Mogil, C., Duan, N., Saltzman, W., Pynoos, R. & Wilt, K. 2010. "The long war and parental combat deployment: effects on military children and at-home spouses". *Journal of the American Academy of Child & Adolescent Psychiatry,* vol. 49, no. 4, pp. 310-320.

Leow, B.G. & Koh, E.C., 2001. "Combining Survey and Administrative Data for Singapore's Census of Population 2000". *53rd Session of the International Statistical Institute, Seoul 2001*, 4 July 2001, pp. 1.

Malley, J., Netten, A. & Jones, K., 2007. "Using survey data to measure changes in the quality of home care. Analysis of the Older People's User Experience Survey 2006". *PSSRU Discussion Paper 2417/2*.

Manzel, K., 2007, "Long run development of human capital and the impact of the "hungry forties" in Spain". *University of Tuebingen Working Paper*.

UNIVERSITY of the
WESTERN CAPE

Mathews, T., Miniño, A.M., Osterman, M.J.K., Strobino, D.M. & Guyer, B., 2011. "Annual summary of vital statistics: 2008". *Pediatrics,* vol. 127, no. 1, pp. 146-157.

Mba, C.J., 2003. "Assessing the reliability of the 1986 and 1996 Lesotho census data". *Journal of Social Development in Africa,* vol. 18, no. 1.

McCarty, C., Killworth, P.D. & Rennell, J., 2007. "Impact of methods for reducing respondent burden on personal network structural measures". *Social Networks,* vol. 29, no. 2, pp. 300-315.

Merry, S.E., 2005. *Human rights and gender violence: Translating international law into local justice.* University of Chicago Press.

Micklin, M. & Poston, D., 2005. "Prologue: The Demographer's Ken: 50 Years of Growth and Change". *Handbook of Population,* pp. 1-15.

Miller, L.S., Mitchell, M.B., Woodard, J.L., Davey, A., Martin, P., Poon, L.W. & for the Georgia Centenarian Study, 2010. "Cognitive performance in centenarians and the oldest old: norms from the Georgia centenarian study". *Aging, Neuropsychology, and Cognition,* vol. 17, no. 5, pp. 575-590.

Møller, V., 2007. "Satisfied and dissatisfied South Africans: results from the General Household Survey in international comparison". *Social Indicators Research,* vol. 81, no. 2, pp. 389-415.

UNIVERSITY of the
WESTERN CAPE

Opiniano, J.M., 2004. *Our Future Beside the Exodus: Migration and Development Issues in the Philippines.* Friedrich-Ebert-Stiftung.

Phillips, H.E., Anderson, B.A. & Tsebe, N.P., 2003. "Sex ratios in South African census data, 1970-96". *Development Southern Africa,* vol. 20, no. 3, pp. 387-404.

Poston, D.L. & Davis, M.A., 2009. "Population aging in the Koreas". *International Handbook of Population Aging.* Springer , pp. 173-190.

Poston, D., 2005, "Age and sex". *Handbook of Population*. Springer, pp. 19-58.

Poston, D.L. & Davis, M.A., 2009. "Population aging in the Koreas". *International Handbook of Population Aging,* pp. 173-190.

Preston, S.H., Heuveline, P. & Guillot, M., 2001. *Demography: measuring and modeling population processes.* Wiley Online Library - John Wiley & Sons.

Rao, C., Bradshaw, D. & Mathers, C.D., 2004. "Improving death registration and statistics in developing countries: lessons from Sub-Saharan Africa". *Southern African Journal of Demography.* pp. 81-99.

Richter, L.M. & Desmond, C., 2008. "Targeting AIDS orphans and child-headed households? A perspective from national surveys in South Africa, 1995–2005". *AIDS Care*, vol. 20, no. 9, pp. 1019-1028.

Robbins, J.A., 2005. "Green Industry Spring Tune-Up: Focused Training For The Retail/Landscape Sector". *HortScience,* vol. 40, no. 3, pp. 870-870.

Roberts, B., 2005. "18 'Empty stomachs, empty pockets': poverty and inequality in". *w.hsrcpress.ac.za. Vol. 18*, pp. 479.

Robinson, K., 1993. "False Hope or a Realizable Right-The Implementation of the Right to Shelter under the African National Congress' Proposed Bill of Rights for South Africa". *Harv.CR-CLL Rev.*, vol. 28, pp. 505.

Rodriguez, G., 2009. Breastfeeding duration, Bangladesh census. Demography Research Princeton. Available http://data.princeton.edu/eco572/digitpref.html Accessed on 21 July 2011.

Rodriguez, G., 2009. Assessment age heaping in Philippines census 1990. Available http://data.princeton.edu/eco572/digitpref.html Accessed on 21 July 2011.

Roos, L.L., Gupta, S., Soodeen, R.A. & Jebamani, L., 2005. "Data quality in an information-rich environment: Canada as an example". *Canadian Journal on Aging,* vol. 24, no. 1, pp. 153-170.

UNIVERSITY of the
WESTERN CAPE

Rutstein, S.O., 2008. *Finding the missing maternal and paternal orphans.* Macro International, Calverton, Md.

Salomon, J.A., Hogan, D.R., Stover, J., Stanecki, K.A., Walker, N., Ghys, P.D. & Schwartländer, B., 2005. "Integrating HIV prevention and treatment: from slogans to impact". *Public Library of Science Medicine,* vol. 2, no. 1, pp. e16.

Session, IV. 2007. "United Nations Statistical Commission and United States Census Economic Commission for Europe Bureau Conference of European Statisticians World Bank".

Setel, P.W., Sankoh, O., Rao, C., Velkoff, V.A., Mathers, C., Gonghuan, Y., Hemed, Y., Jha, P. & Lopez, A.D., 2005. "Sample registration of vital events with verbal autopsy: a renewed commitment to measuring and monitoring vital statistics". *Bulletin of the World Health Organization,* vol. 83, no. 8, pp. 611-617.

Shryock, H.S., Siegel, J.S. & Larmon, E.A., 1980. *The methods and materials of demography,* US Dept. of Commerce. Bureau of the Census: for sale by the Supt. of Docs. US Govt. Print. Off.

Shryock, H.S. & Siegel, J.S., 1976. Methods and Materials in Demography, condensed edition. Academic Press, New York.

Siegel, J.S., 1980. The methods and materials of demography, Department of Commerce. Bureau of the Census.

Sridhar, K.S. & Reddy, A.V., 2011. "Reasons for Rural–Urban Migration: Recent Evidences from Bangalore". *Migration, Identity and Conflict: India Migration Report 2011*, pp. 125.

Statistics, S. A., 2010. *South African Statistical Quality Assessment Framework (SASQAF).* Operational Standards and Guidelines. Statistics South Africa, Pretoria. 1st ed., p. 87

Statistics, S. A., 2009. 2009 Mid-year population estimates. Statistical Release P0302. Statistics South Africa: Pretoria.

Statistics, S. A., 2008. *South African Statistical Quality Assessment Framework (SASQAF).* Statistics South Africa, Pretoria.

Statistics, S.A., 2008a. Guide to Quaterly Labour Force Survey. Report No 02-11-01. http://www.statssa.gov.za/qlfs/docs/Quarterly-Labour- Force-Survey-Guide.pdf

Statistics, S. A., 2007. Community survey 2007. Statistical release P301. Pretoria.

Statistics, S.A., 2007. Using the 2001 census approaches to analyzing data, Stats SA and the Human Sciences Research Council. ISBN 978-0-621-37259-5. Pretoria Statistics SA.

UNIVERSITY of the
WESTERN CAPE

Statistics, S. A., 2003. Census 2001: investigation into appropriate definitions of urban and rural areas for South Africa, Discussion document. Pretoria: Statistics South Africa (Report No. 01-01-20) www.statssa.gov.za.

Statistics, S. A., 2003. *Census 2001: Census in brief,* Statistics South Africa.

Statistics, S. A., 2003. Census 2001: How the count was done. Pretoria, South Africa.

Statistics, S.A., 1999. "Towards Improving the Registration of Marriages and Divorces in South Africa".

Statistics, S.A., 1998. "The People of South Africa Population Census, 1996". *Pretoria, South Africa: Statistics South Africa*.

Statistics, S.A., 1998a. The people of South Africa: population census 1996, calculating the undercount in census 96. Report 03-01-18 (1996). Pretoria: Statistics South Africa.

Swanson, D., Siegel, J.S. & Shryock, H.S., 2004. *The methods and materials of demography,* Emerald Group Publishing.

Taha, M.M., 2001. *The Potential Role of GIS in the Development and Applications of Urban Indicators: The Case of Housing in Khartoum. Sudan,* Department of Civil and Environmental Engineering, Royal Inst. of Technology.

Trovato, F. & Heyen, N., 2006. "A varied pattern of change of the sex differential in survival in the G7 countries". *Journal of Biosocial Science,* vol. 38, no. 3, pp. 391.

UNESCO, 2003. Gender and Education for All: The Leap to Equality. *EFA Global Monitoring Report.* Paris, UNESCO.

UNESCO, 2005. Education for All: The Quality Imperative. *EFA Global Monitoring Report.* Paris, UNESCO.

UNFPA, 2011. Vietnam Population and Housing Census 2009, Monograph - Age-Sex Structure and Marital Status of the Population in Vietnam. Ministry of Planning and Investment, General statistics Office. pp 166.

United Nations, 2008. "Principles and Recommendations for Population and Housing Censuses, Revision 2". *United Nations Expert Group Meeting on the 2010 World Programme on Population and Housing Censuses 2008*.

United Nations, 2008. Principles and Recommendations for Population and Housing Censuses Revision 2 Statistical papers Series M No. 67/Rev.2. ST/ESA/STAT/SER.M/67/Rev.2

United Nations, Statistical Division, 2008. Designing Household Survey Samples: Practical Guidelines. United Nations Educational.

United Nations, Statistical Division & National Household Survey Capability Programme, 2005. *Household Sample Surveys in Developing and Transition Countries.* United Nations Publications.

United Nations, Division for Public Economics & Public Administration, 2000. *Professionalism and Ethics in the Public Service: Issues and Practices in Selected Regions.* United Nations Publications.

United Nations, 1998. Principles and Recommendations for Population and Housing Censuses, Revision 1. Statistical Papers, Series M, No. 67/Rev. 1. Sales No. E.98.XVII.8.

United Nations, Economic and Social Council & United Nations. Economic Commission for Africa, 1996. "Report of the regional working group on recommendations for the 2000 round of population and housing censuses in Africa".

United Nations, 1996. "Report of the fourth World Conference on women" Beijing, 4-15 September 1995, P. 218.

UNIVERSITY of the
WESTERN CAPE

United Nations, 1990. 1988 Demographic Yearbook. United Nations, New York.

United States, B. C., 1985. *Evaluating Censuses of Population and Housing*, Statistical Training Document, ISP-TR-5. Washington, D.C.: US Bureau of the Census.

United States, Bureau of the Census., 1964. Census of population and housing, 1960 public use sample: one-in-one-thousand sample. Washington, D.C.

United Nations, 1952. "Accuracy Test for Census age distributions Tabulated in five years and ten year groups" . Population Bulletin of the UN, No. 2, new York, pp. 59-79.

Velkoff, V.A. & Kowal, P.R., 2007. *Population aging in Sub-Saharan Africa: demographic dimensions 2006.* US Dept. of Commerce, Economics and Statistics Administration. US Census Bureau.

Visaria, L., 2004. "The Indian Census", *Population Census in South and West Asia,* .

Waldron, I., 1976. "Why do women live longer than men?". *Social Science & Medicine (1967),* vol. 10, no. 7, pp. 349-362.

Welzel, C., Inglehart, R. & KLIGEMANN, H.D., 2003. "The theory of human development: A cross-cultural analysis". *European Journal of Political Research,* vol. 42, no. 3, pp. 341-379.

Wittenberg, M., 2004. "The Analysis of Survey Data: A practical introduction using South African data". *School of Economics. University of Cape Town*.

Wittenberg, M., 2003. *Decentralization in South Africa.* Econometric Research Southern Africa.

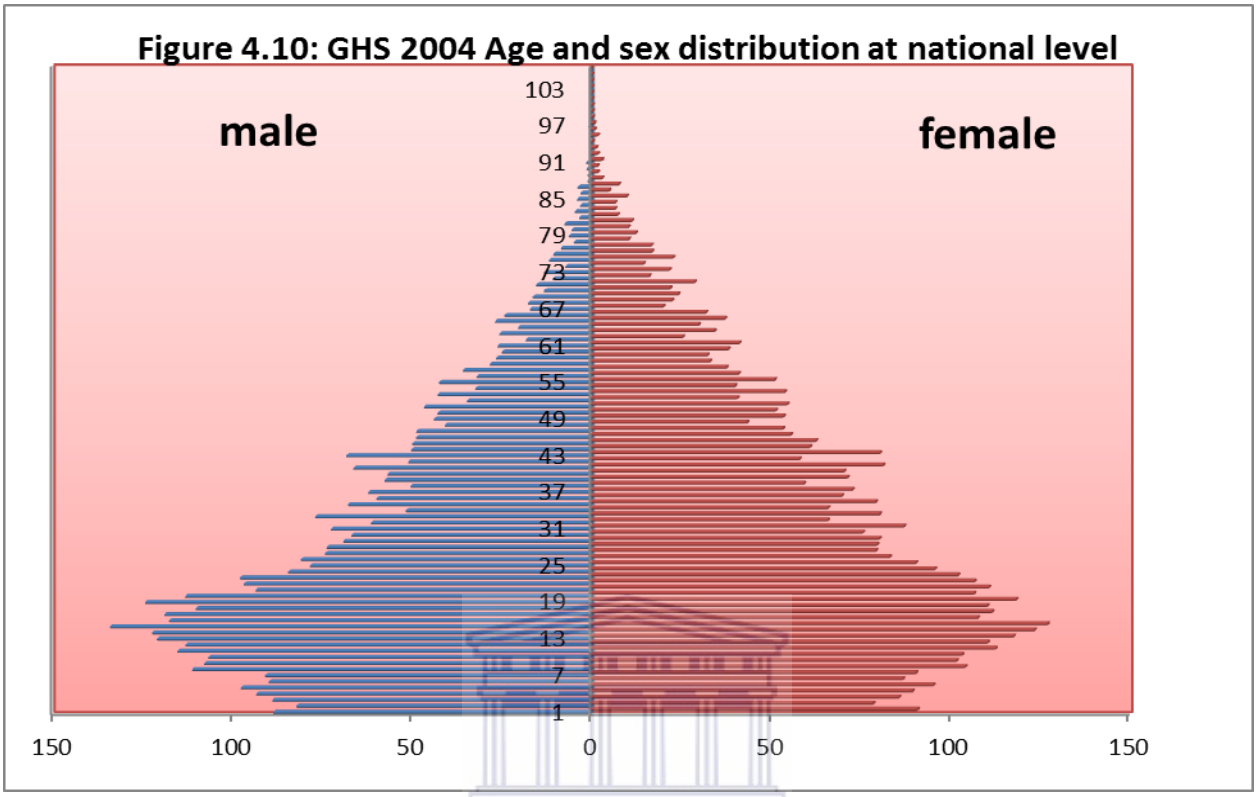
Wright, G., Noble, M., Barnes, H. & Noble, S., 2009. "The South African Index of Multiple Deprivation for Children 2007 at Municipality Level". Social Development, Statistics S.A. pp. 27.

Yearbook, D., 1997. "Historical supplement". *Special Issue (New York: United Nations, 1979). Table,* vol. 2, pp. 189-209.

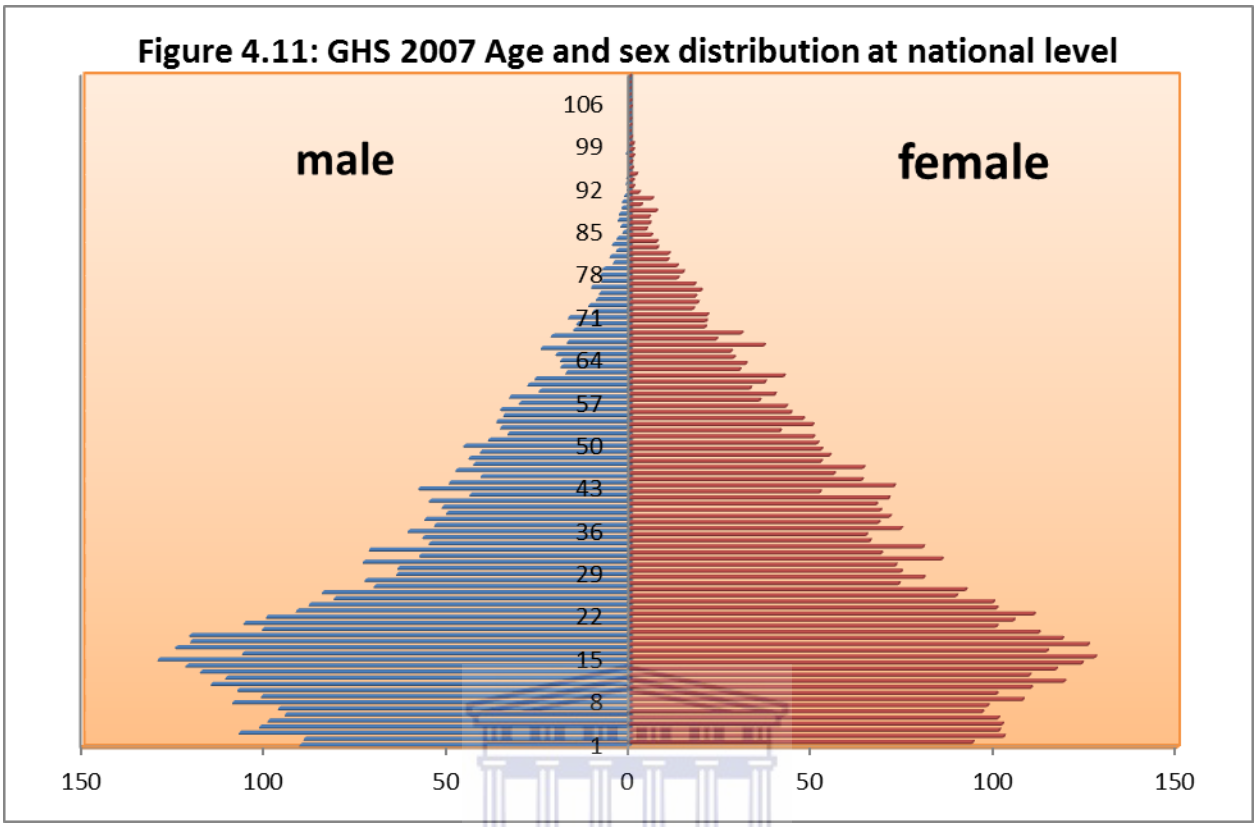
Zeleza, T., 1997. *A modern economic history of Africa: The nineteenth century*. East African Educ Press.

UNIVERSITY of the
WESTERN CAPE

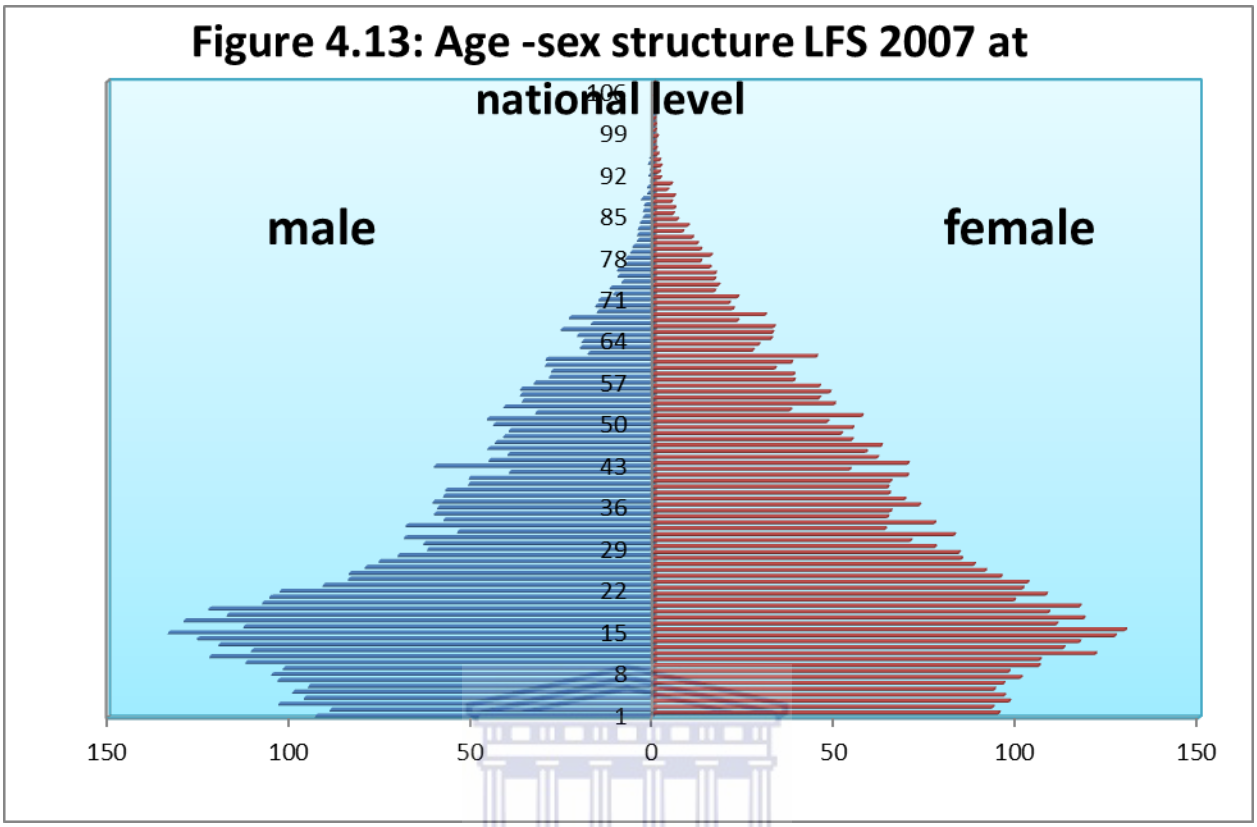
4.2.7 Figures in Appendices



UNIVERSITY of the
WESTERN CAPE



UNIVERSITY of the
WESTERN CAPE



UNIVERSITY of the
WESTERN CAPE

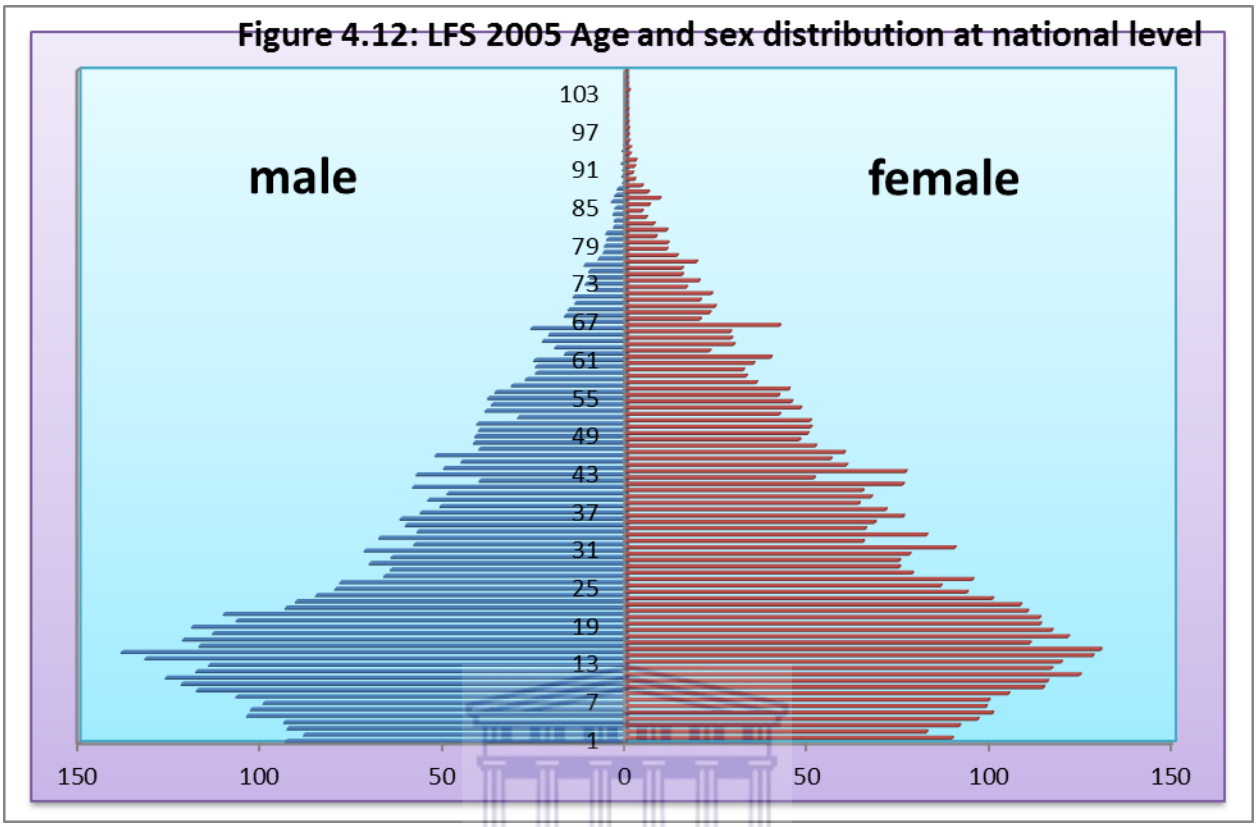


Figure 4.12: LFS 2005 Age and sex distribution at national level

UNIVERSITY of the
WESTERN CAPE

Appendices

E-mail Questionnaire applied to investigate the practical method Statistics SA use to assess the quality of Data on age and sex collected from censuses and surveys in South Africa.

Germaine Kamleu
Department of statistics
University of the Western Cape
7535 Bellville
E-mail: germainekam@gmail.com
Cell: 0765084511

Dr. Isabel Smith

Liberty Building, 3rd Floor.

22 Long Street Cape Town.

Dear Madam,

Request for Information

I have the honour most respectful to approach your high office to request for information. I am Germaine Kamleu, a research master's student at the University of the Western Cape writing a dissertation titled, "Assessing the data quality from Censuses, Labour force surveys and General household surveys in South Africa." Your cooperation is crucial and will go a long way to shed light and contribute to the success of this project.

As an authority in this field, I would be very grateful if you could with precision assist me in answering the following questions:

1. How does Stats SA handle missing information regarding Age, date of birth?
2. How does Stats SA handle non-response in censuses and surveys?
3. Why is the quality of data on Age and Gender so important?
4. What are the different methods used in collecting Age data?

5. What improvements have been done on the methods of collection of data on age and sex from 1996 -2010? collection
6. Are there differences in methods of data collection between censuses and surveys?
7. What techniques are used to assess the data quality by Stats SA?
8. Are there differences in the methods used to assess the quality of census and survey data by Stats SA?
9. Are there any adjustment to the survey data before report and how is it done?
10. Is adjustment done at the national or regional level and what accounts for the differences in quality of data at the national and regional level and why have these differences persisted.

Madam, the progress and completion of this project will depend on finding answers to these questions. I therefore plead that you treat it as a matter of urgency.

Yours sincerely
Germaine Kamleu

UNIVERSITY of the
WESTERN CAPE

Responses to Questions of Germaine Kamleu UWC

1. How does Stats SA handle missing information regarding Age, date of birth?
   For the GHS missing date of birth information is not imputed. However, limited age imputations are done based on information reported for relation to the household head (if either ages are supplied and the other is missing), whether you are married or living together and spouse/partner is living in the household (in which case the missing age of a partner can be imputed based on the reported age of the partner), and father or mother living in the household in the case of missing child age information (using the ages of the parents).

   Reported date of birth is used to calculate age information if missing. If there are discrepancies between reported age and date of birth, then a set of rules are followed to determine which one of the two should be used.

2. How does Stats SA handle non-response in censuses and surveys?
   In Census 2001 non-response was imputed to a 100% imputation rate. The QLFS imputes non-response 100%. The GHS conservatively imputes non-response but does not aim for 100% imputation.

3. Why is the quality of data on Age and Gender so important?
   The data on age and gender (and population group) are used to weigh/calibrate the sample to the population as a whole. The sampling experts use tables that have the predicted number of persons in each age group by sex and population group for each year to weigh the sample data so that it reflect the estimated population numbers for that year.

4. What are the different methods used in collecting Age data?
   Statistics South Africa uses two methods: asking for the age in years and also for the date of birth.

UNIVERSITY of the WESTERN CAPE

5. What improvements have been done on the methods of collection of data on age and sex from 1996 -2010?

   The main change has been to include a question on date of birth in addition to the one on age alone.

6. Are there differences in methods of data collection between censuses and surveys?

   Self-completion (drop off and complete) is used by the Census in selected areas. Surveys are always done face to face.

7. What techniques are used to assess the data quality by Stats SA?

   Various methods are used once the data is finalized but all follow the SASQAF guidelines (please check and download from website).

   Within each division we do the following while editing and preparing the data:

   1) During processing the Data Processing centre runs checks on the data they specifically check for large numbers of missing values in which case the data was not processed properly. They also check and manually correct images that could not be clearly identified by the scanner.
   2) We write editing programs to identify inconsistencies, violated skip instructions and extreme values.
   3) We do manual checks to verify selected inconsistencies to make sure that they are not the result of processing and scanning errors.
   4) Then we run automated editing and imputation programs to correct skip violations and impute missing data.
   5) We do final checks to look at imputation rates and evaluate the extent to which the data was modified or changed during editing and imputation.

8. Are there differences in the methods used to assess the quality of census and survey data by Stats SA?

   No they are primarily the same except that because of the volume of Census questionnaires they cannot do any manual checks all editing is done automatically.

UNIVERSITY of the
WESTERN CAPE

9. Are there any adjustment to the survey data before report and how is it done?
   We do not do any special adjustments. The only recent change for the GHS has been to weigh the person and household files separately.

10. Is adjustment done at the national or regional level and what accounts for the differences in quality of data at the national and regional level and why have these differences persisted.
    No adjustments are done to national and regional level. Our master sample design was done in such a way that stratification took place at national and provincial level. We get reasonably constant trends when analyzing at provincial level across time. However, particularly for the Western Cape we have found that changes in the estimates of the population that is used to weigh the household data has made the time series vulnerable to criticism. There are also some questions for which provincial level estimates have not provided the smooth trends that we expected and we find it difficult to explain. What is important though is that key questions such as those on education show great consistency in terms of proportions (not absolute numbers as that is influenced by the population estimates used for weighting) with that from the Department of Basic Education.

UNIVERSITY of the
WESTERN CAPE

Germaine Kamleu
Department of statistics
University of the Western Cape
7535 Bellville
E-mail: germainekam@gmail.com
Cell: 0765084511

Dr. Isabel Smith

Executive Manager

Social Statistics South Africa

Tel: 012 337 6379

Mobile: 082 884 4281

Dear Madam,

Request for Information

I have the honour to appreciate your contribution towards the investigation you made in making my work to be successful. As an authority in this field, I would be very grateful if you could with precision assist me in answering the following few questions:

1. In the case of GHS, when old people are not able to declare their ages properly especially when they are the head of household and living with their grandchildren what strategy do you take to get their approximate ages?

2. How does the GHS handle non-response in surveys? The GHS conservatively imputes non-response but does not aim for 100% imputation. Please can you be more specified and show examples on how you deal with it.

3. Non-response at national level; for example 1% of the total population represent the non-response. What do you do with this people? Will you take them into account? What procedure will you adopt? And if this people are from the same race or coming from the same region how will you manage solve this issue?

Madam, the progress and completion of this project will depend on finding answers to these questions. I therefore plead that you treat it as a matter of urgency.

Yours sincerely
Germaine Kamleu

Germaine Kamleu
Department of statistics
University of the Western Cape
7535 Bellville
E-mail: germainekam@gmail.com
Cell: 0765084511

MrsTasneemSolomons

STATISTICS South Africa

Dear MrsTasneem,

Request for Information

I have the honour most respectful to approach your high office to request for information. I am Germaine Kamleu, a research master's student at the University of the Western Cape writing a dissertation titled, "Assessing the data quality from Censuses, Labour force surveys and General household surveys in Africa." Your cooperation is crucial and will go a long way to shed light and contribute to the success of this project.

UNIVERSITY of the
WESTERN CAPE

As an authority in this field, I would be very grateful if you could with precision assist me in answering the following questions:

1. How does Statistical office in South Africa handle missing information regarding Age, date of birth?

2. How does Statistical office in South Africa handle non-response in censuses and surveys?

3. Why is the quality of data on Age and Gender so important?

4. What are the different methods used in collecting Age data?

5. What improvements have been done on the methods of collection of data on age and sex from the two last censuses collection?

6. Are there differences in methods of data collection between censuses and surveys in South Africa?

7. What techniques are used to assess the data quality in South Africa?

8. Are there differences in the methods used to assess the quality of census and survey data in South Africa?

9. Are there any adjustment to the survey data before report and how is it done?

10. Is adjustment done at the national or regional level and what accounts for the differences in quality of data at the national and regional level and why have these differences persisted.

Mrs, the progress and completion of this project will depend on finding answers to these questions. I therefore plead that you treat it as a matter of urgency.

Yours sincerely

Germaine Kamleu

UNIVERSITY of the
WESTERN CAPE