Methodology and Assessment of Data Quality of the Somali People

Volume 1

UNFPA

UKaid
from the British people

SWEDEN

This report should be cited as: UNFPA (2016). *Methodology and Assessment of Data Quality of the Somali People.*

The report can be downloaded directly from http://somalia.unfpa.org/publications/
Technical enquiries on the reports can be directed to: semu-banda@unfpa.org

United Nations Population Fund Somalia,
Block P, Level 2, UN Complex,
UN Avenue, Gigiri.
P.O. Box 28832 – 00200,
Nairobi, Kenya.
Tel: (+254) (0) 207 625 742
www.unfpa.org

MAPS AND DESIGNATIONS
The designations used for the maps in this report are the 1986 pre-war geographic regions and boundaries of Somalia. These do not imply any expression of opinion whatsoever on the part of UNFPA concerning the legal status of any administration and its authorities. It is important to note that regions and districts have changed and the newly established regions may have no link to the pre-war regional and district boundaries used in this analysis.

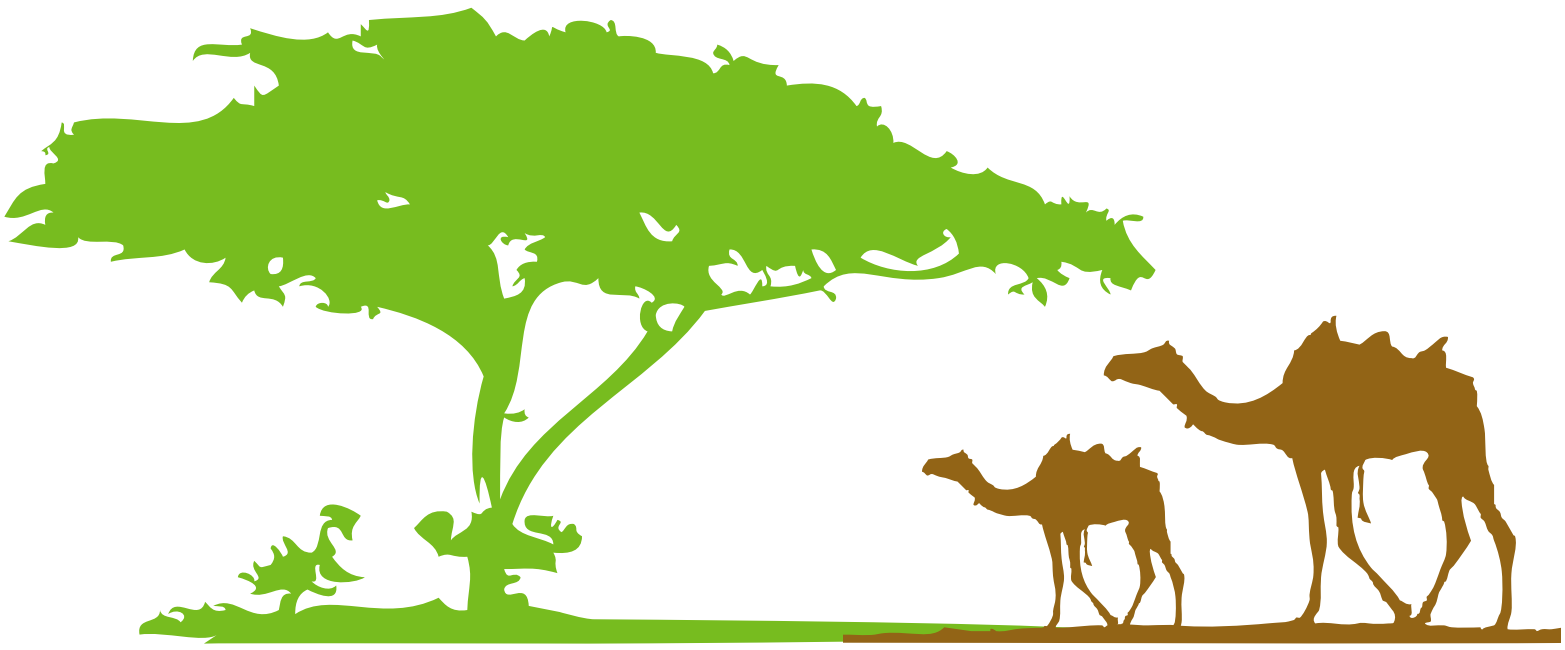
Photo credits
Cover page: ©UNFPA Somalia, **Photo 1:** AU/UN IST, **Introduction:** ©Keisha Rukikaire/IRIN, **Chapter Two:** ©UNFPA Somalia, **Chapter Three:** © UNFPA Somalia, **Chapter Four:** ©UNFPA Somalia, **Chapter Five:** ©UNFPA Somalia, **Chapter Six:** ©UN/Somalia F. Juez, **Conclusions:** ©UN/Tobin Jones

Design and layout by: UNON Publishing Services

The Federal Government of Somalia covers South-West, Juba-land, Galmudug and Puntland; Somaliland declared its unilateral independence in May 1991 and is yet to be recognised internationally.

Methodology and Assessment of Data Quality of the Somali People

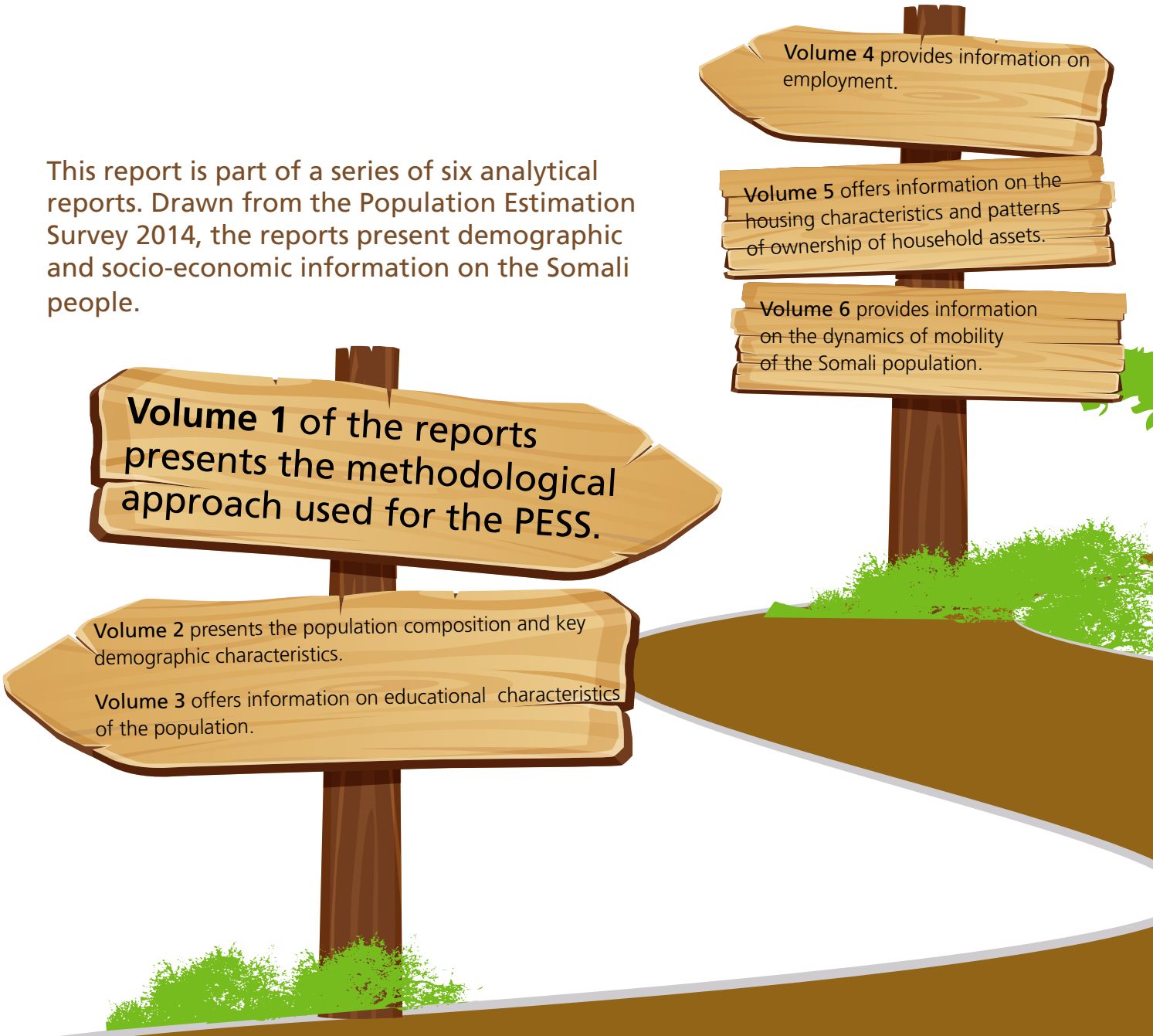
Volume 1



Data for a Better Tomorrow

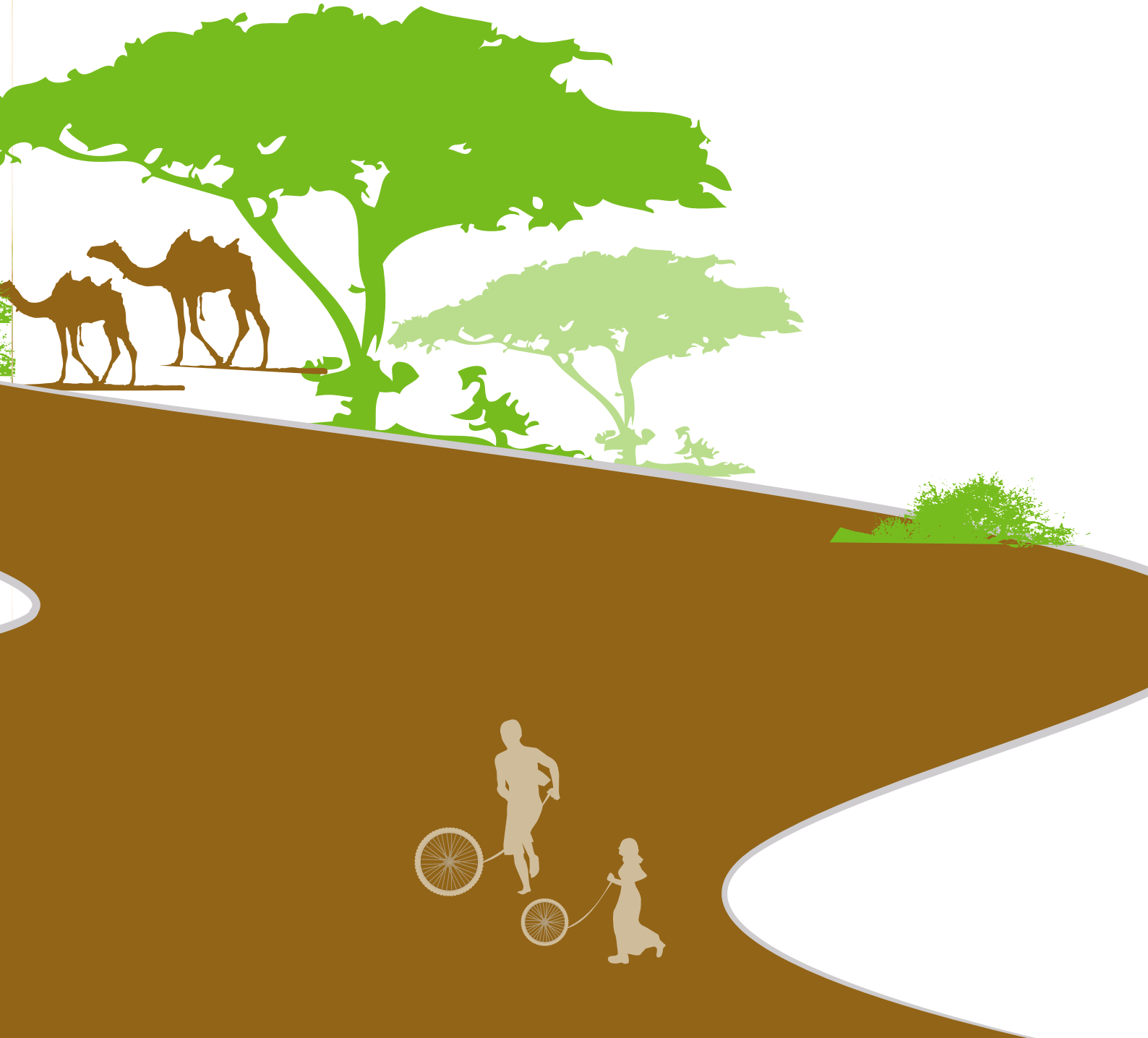
PESS 2016

This report is part of a series of six analytical reports. Drawn from the Population Estimation Survey 2014, the reports present demographic and socio-economic information on the Somali people.



“Without data, you’re just another person with an opinion.”

W. Edwards Deming



FOREWORD

These reports are expected to change the rhetoric on the absence of information about the lives of Somalis.

On behalf of the UN Country Team, it is my pleasure to present this volume, which is part of a series of analytical reports based on 2014 Population Estimation Survey (PESS) data. These reports are expected to change the rhetoric on the absence of information about the lives of Somalis. The PESS is the first large-scale household sample survey to be conducted to estimate the Somali population in more than three decades. Along with reliable population estimates, this series of analytical reports provides a comprehensive picture of Somalis and the lives they lead. It tells their story: how and where they live; how old their family members are; how many are men, women or children; how many have access to education; how many are employed; what kind of assets they own; their mobility patterns– among other crucial social and economic indicators. The United Nations Population Fund (UNFPA) took on this task jointly with the Somali authorities, and with the support from the United Kingdom (UK) Department for International Development (DfID) and the Swedish Embassy.

The aims of the series of the analytical reports are to provide a sound foundation of information for policymakers and political, economic and social actors to craft articulate strategies and to avail much-needed benchmark population data. The findings provide valuable insights into the challenges faced by the Somalis on the road to build a stable and peaceful future. For example, due to the high fertility, the proportion of children is very high, while that of the working age population (15-64 years) is relatively small. This creates a 'burden' for the working age population to cater for the needs of the young and the older persons. Somali's age dependency ratio (which measures this 'demographic burden') is higher than in most of the neighbouring countries. In addition, nearly half of the working age population is economically inactive which means that the economic dependency burden on the labour force (i.e. the employed and unemployed) is nearly doubled. Just under a quarter of the labour force is unemployed. The Somali unemployment rates are close to those of Ethiopia and Sudan.

In addition to the wealth of information that the analytical reports provide, the two years of meticulous planning, implementation of the survey, and analysis of information have left a great legacy for future generations, including a strengthened Somali statistical system and on increased capacity to conduct similar large-scale surveys. This is also a stepping stone towards a potential population and housing census in the future.

A mammoth task of this kind can only be the result of hard work, commitment and dedication of several individuals and institutions. They range from Somali authorities, who guided the undertakings, Somalis who allowed us to take a glimpse of their lives, enumerators walking from door to door at times under trying circumstances to collect information, to donor agencies providing support at every stage, among other partners.

I remain hopeful that Somalis and development and humanitarian agencies working to support them will be able to use this information to draw up effective plans and programmes that aim to improve the stories and lives of Somalis.

Peter de Clercq (signed)
Deputy Special Representative of the UN Secretary-General,
UN Resident and Humanitarian Coordinator for Somalia,
UNDP Somalia Resident Representative

PREFACE

The United Nations Population Fund has the honour of unveiling a milestone for the Somali people: a series of analytical reports based on the data of the 2014 Population Estimation Survey.

It has been a privilege for UNFPA to work closely with and be guided by the Somali authorities and experts in the preparation of these reports. I would like to commend the hard work that went into their production. The analysis helped to uncover and present crucial information on the Somali population. This would not have been possible without the cooperation of the numerous Somali officials and experts who supported the process and shared information with us, and those who braved through various circumstances to collect and record information.

The publication of these reports would not have been possible without the generous contribution from the UK Department for International Development (DfID). DfID helped to turn a Somali dream into reality, through capacity building for the Somali experts involved in the writing of the reports, and promoting the widest possible use of the PESS data. I would also like to thank the Swedish Embassy for their invaluable support through all stages of the PESS project.

We now have much-awaited information about the lives and needs of the Somali people, such as how many women, youth and children there are; where they live; who the most vulnerable members of the society are; what kind of educational levels they have had access to; what household assets they own; how many are seeking employment; and how many are moving across national and international borders, among other indicators. This information serves as a reference for development and a benchmark to measure the progress made.

I would urge Somali authorities, and their national and international partners, including institutions of higher learning to use the PESS data and the information these analytical reports present. Every number tells a story about a Somali household, and the life it leads.

From the numbers presented, it is evident that the country is demographically very young, with three-quarters of the population under 30 years of age. Only two out of ten children of primary schoolgoing age are currently enrolled in school. Two out of ten households are headed by women, with a further two in a thousand households headed by children. One in ten under-eighteens has been married at least once in their lives. Two in ten households have no access to a human waste disposal facility. For every one thousand Somalis living in the country, twenty-one are living outside the country.

So far, numerous attempts have been made to make progress in the humanitarian conditions and overall development of the Somali people. However, we have lacked information that would help steer us in the right direction.

This series of reports brings new, credible promises for the Somali people. Using the information offered, government officials will be able to better address inequalities between men and women, the wealthy and the underprivileged or vulnerable members of the society. Somali authorities will now be able to design and implement articulate, targeted and inclusive pro-poor policies and programmes. It will also enable development and humanitarian actors to plan, implement, and

We have a window of opportunity, and collective responsibility, to improve the lives and realities of individuals, families and communities.

monitor activities in an effort to direct aid to areas with the greatest need. We have a window of opportunity, and collective responsibility, to improve the lives and realities of individuals, families and communities.

In addition to the invaluable data about the Somali people at a critical juncture of their history, PESS leaves another important legacy—a strengthened statistical system and an increased capacity to conduct large-scale surveys and population counts. It is thus a stepping stone towards a future population and housing census, which will help put in practice the "one person-one vote" principle that underlies every stable democracy.

It is my hope that Somali authorities and their partners will acknowledge that behind each number presented in the reports is a human face and story. Let us ensure we listen and do justice to these unheard voices.

Nikolai Botev (signed)
UNFPA Representative

ACKNOWLEDGEMENTS

The efforts towards the successful production of the Analytical Thematic Reports 2016, from planning to data analysis and actualisation of the thematic reports, have been vigorous. This initiative has involved a number of organizations and individuals, both in the UN fraternity and donor organizations. On behalf of the UN, we would like to express our heartfelt gratitude to those involved.

First and foremost, we would like to acknowledge the Somali authorities for steering the process in such a professional manner and creating consensus at every stage. Particularly, we extend our gratitude to the in-country team, including Ahmed Elmi Muhumad (former Director General, Statistics), Nur Ahmed Weheliye (Post-PESS Coordinator), Hashim Sheikh Abdinoor (Technical Support), Mohamed Abdinur Mohamed (Technical Support) and Hussein Elmi Gure (Technical Support).

Likewise, we would like to single out Leo Thomas, Results Advisor (DfID), and his core team, Hannah Chira and Maimuna Abdalla, for their technical support and consistent guidance. We remain grateful for the generous contribution from UK DfID, which helped turn the dream of the Somalis into reality, in terms of capacity building for Somalis in gathering and analysing information, writing of the analytical thematic reports and ensuring the delivery of the information to the various users. We would also like to acknowledge the valuable inputs and advice from Vincent Kutai, Programme Manager, Water, Sanitation and Hygiene (WASH) and Statistical Support from the Swedish Embassy, particularly through the input of Jean-Pierre Ntezimana and the team from Statistics Sweden.

It would not have been possible to bring these reports through to completion without the guidance from Nikolai Botev (Representative, UNFPA Somalia), as well as the steady support of Grace Kyeyune (Deputy Country Representative, UNFPA Somalia). Special efforts made by the Population and Development (P&D) Manager, Mariam Alwi, for her devotion and patience in steering this work, are highly appreciated.

The technical team involved did a commendable job in their professional capacities, and their passion, commitment and dedication has been much appreciated. Team members include: Felix Mulama (Consultant Technical Lead), Richard Ng'etich (Technical Lead), Sammy Oyombe (Statistician), Zena Lyaga (Consultant Demographer), Susan Maina (Consultant Demographer), Umikaltuma Mohamed (Geographical Information System (GIS) Consultant), Samwel Andati (Data Management Assistant), John Okongo (Programme Associate) and Jaafar Adon (Programme Assistant). The team also benefitted from the technical contribution of Per Schoning (Norway Statistics).

We also express our deep gratitude to the editorial team: Namita Mediratta (Editorial Consultant), who copy-edited the reports, and ensured that the language is in a readable format for the various users, Osman Hussein Warsame (P&D Consultant) who reviewed the applicability of the information in the context of the country, and Emily Denness (Midwifery Specialist), who provided kind support and time in editing and proofreading the reports. We would like to thank Scadden Orina (Graphic Designer), who created the info graphics and illustrations used both for print and web.

Our gratitude also extends to the UNFPA internal reviewers for providing the first in-depth scrutiny of the reports. They include: Bakhtior Kadirov (Head, Garowe sub-office), Ezekiel Kutto (Monitoring and Evaluation Analyst), Salada Robleh (Head, Mogadishu sub-office), Pilirani Semu-Banda (Communications Specialist), Salad H Dualle (National Programme Specialist), Anas Jabir Babikir (Head, Hargeisa sub-office) and Ahmed Mihile (P&D Specialist).

We are also privileged and grateful to have benefitted from the extensive knowledge and expertise of the external reviewers: Jeremiah Banda former Chief, Demographic and Social Statistics Branch, United Nations Statistics Division (UNSD) in New York and former PESS Team Leader, UNFPA Somalia, Werner Haug (Former Technical Director, Regional Office for Eastern Europe and Central Asia UNFPA) and Eric Jager (Former PESS Demographer UNFPA Somalia). We also thank the UN family, particularly United Nations Development Programme (UNDP), United Nations Children's Emergency Fund (UNICEF), World Food Programme (WFP), Food and Agricultural Organization (FAO), United Nations High Commission for Refugees (UNHCR), International Organization for Migration (IOM) and United Nations Human Settlements Programme (UN-Habitat), and the donor community—in particular UK DfID and the Swedish Embassy—for their continued support.

These analytical reports would not have been a success without the contribution of several individuals and institutions, many of whom are acknowledged in the initial PESS 2014 report. We remain greatly indebted to each one of them.

EXECUTIVE SUMMARY

There has been only one successful census in Somalia, carried out in 1975. The 1986 census remained incomplete, and its results were never published. The PESS was conducted to provide updated information on the Somali population for planning purposes, as well as humanitarian work. It is the first comprehensive population survey to be carried out since the 1975 census.

To collect information, the PESS selected a representative sample of clusters through a one-stage stratified-cluster sample design. These clusters formed the primary sampling units (PSUs), which were referred to as enumeration areas (EAs) in urban areas, settlements in rural areas, camps for internally displaced persons (IDPs), and water points for nomadic populations. In each of these areas, except for the water points, the PSUs comprised 50 to 149 households.

The survey methodology entailed determining sample sizes for urban, rural and nomadic areas, and internally displaced person camps, allocating the sample to the 18 pre-war regions, which formed the first strata. This was followed by the selection of clusters for urban and rural areas, and IDP camps. For nomadic populations, water points were grouped by type, following which the water points were selected for enumeration. The estimated sample size was 2,535 PSUs, which was then adjusted to 2,735, comprising 868 EAs, 1,104 villages, 735 water points (for the nomadic population) and 28 camps for the IDPs.

Data was processed in a standard manner in statistical hubs in Hargeisa, Garowe and Mogadishu. The string of processes began with the receipt of the filled-in questionnaires, manual code checking, data entry and processing/exporting of data from Census and Survey Processing System (CSPro) into Statistical Package for the Social Sciences (SPSS), and finally the construction of weights, which were included in the data set.

The PESS team evaluated the extent of inaccuracies in age and sex reporting in the data by applying the single year age and sex distribution, Whipple's Index, Myer's Blended Index and UN Age Sex Accuracy Index. The assessment of the quality of age and sex data indicates an undercount in the age group 0-4 years, "age heaping" on ages with terminal digits "0" and "5", and relatively low numbers of the male population in the ages 20-39 years, which may indicate emigration of men in the working ages.

Statistical methods of "smoothing" were applied to restructure the reported age and sex data to address age misreporting. Different methods/approaches are presented in this volume. Strong smoothing provided ideal estimates for the age and sex data of Somali people. The use of smoothed data on the macro-level is appropriate for national population projections and other specific applications of national planning. The smoothing showed that reported ages are relatively inaccurate at younger ages, and is more pronounced for males than females.

However, for the calculation of complex demographic indicators and the breakdown of indicators by regions and type of residence, it is not possible to use smoothed data at this stage, as additional work and analysis would be required. The analytical volumes of the PESS include, therefore, only data that are not smoothed.

The computed standard errors indicate confidence in and high reliability of PESS estimates. The Coefficient of Variation (CV), a measure for reliability for most estimates is assessed to be good (CV <= 15%) based on a classification scheme developed for the American Community Survey (ACS) conducted by the United States Bureau of the census. The Coefficient of Variation for most estimates was less than 10 percent, which is below 15 percent, the upper margin for estimates assessed to be good, which generally indicates that the quality of the PESS meets international standards.

CONTENTS

1 INTRODUCTION 2
1.1 History of census taking 2
1.2 Rationale for conducting the Population Estimation Survey 2
1.3 Filling the data gap 2

2 SURVEY METHODOLOGY 5
2.1 PESS planning and organization 5
2.2 Sampling frames 5
2.2.1 Urban sampling frame 5
2.2.2 Rural sampling frame 5
2.2.3 Internally displaced persons' sampling frame 6
2.2.4 Water points' frame 6
2.3 Sample stratification, design and selection 6
2.3.1 Stratification 6
2.3.2 Sample design 7
2.3.3 Sample size for urban and rural population 7
2.3.4 Sample allocation to regions and sub-strata 7
2.3.5 Selection of PSUs 8
2.3.6 Selection of segments from large settlements 9
2.3.7 Sample selection of water points 10
2.3.8 Population count 10
2.4 Data collection 11
2.4.1 Selection and training of field staff 11
2.4.2 Stages of data collection 12
2.4.3 The household survey questionnaire 12
2.5 Survey limitations 13

3 DATA PROCESSING 15
3.1 Data centres 15
3.2 Data coding 15
3.3 Data entry 15
3.4 Data editing 15
3.5 Tabulation 16
3.6 Dissemination 16

4 ESTIMATION 19
4.1 Estimation 19
4.2 Adjustment for non-coverage and non-response 19

5 QUALITY ASSESSMENT AND ADJUSTMENT OF AGE AND SEX DATA 21
5.1 Introduction 21
5.2 Assessment of age heaping 21
5.2.1 Single year age and sex distribution 22
5.2.2 Whipple's Index 24
5.2.3 Myer's blended method (Myer's Index) 25
5.2.4 UN Age-Sex Accuracy Index or UN Joint Score 27

5.3 Five-year age and sex distribution 29
5.3.1 Population by broad age groups 30
5.3.2 Age ratios 31
5.3.3 Smoothing of age and sex data 32

6 SAMPLING ERRORS 37
6.1 Introduction to sampling errors 37
6.2 Sampling errors for population estimates 38
6.3 Sampling errors for literacy 38
6.4 Sampling errors for school enrollment 39

7 CONCLUSION 42
7.1 Conclusion 42

REFERENCES 44
APPENDIX A - Assessment of age and sex data in the population 45
APPENDIX B – Sampling errors 47
APPENDIX C – Glossary 53
APPENDIX D – Basemap 56
APPENDIX E – Questionnaire 57

FIGURES

Figure 1: Steps followed in segmenting large rural settlements 9
Figure 2: Flowchart for data edits and tabulation for PESS 17
Figure 3: Population by age in single years 22
Figure 4: Population by single years of age and sex 22
Figure 5: Population by age in single years, MICS 2011 23
Figure 6: Myer's Digit Preference 26
Figure 7: Myer's digit preference, MICS NWZ and NEZ 27
Figure 8: Population by age group 30
Figure 9: Population by broad age groups and sex 31
Figure 10: Smoothed population data by age group 35

TABLES

Table 1: Final allocation of primary sampling units by the four strata 8
Table 2: Water points substrata 10
Table 3: Survey coverage 19
Table 4: Degree of accuracy of the age reporting using the Whipple's Index 24
Table 5: Comparison of Whipple's Index for Somalia and other countries 25
Table 6: Myer's digit preference for selected African countries 26
Table 7: Summary of indices measuring the accuracy of data 29
Table 8: Comparison of UN joint score for somalia and other countries 29
Table 9: Age ratios by sex 32
Table 10: Reported and smoothed population by age – males 33
Table 11: Reported and smoothed population by age groups – females 34
Table 12: Reported and smoothed population by age groups – total 34
Table 13: Sampling errors for the population 38

Table 14:	Sampling errors for ability to read and write	38
Table 15:	Sampling errors for school enrollment	39
Table A.1:	Myer's Blended Method (Myer's Index) computation - males	45
Table A. 2:	Myer's Blended Method (Myer's Index) computation- females	46
Table B. 1:	Standard errors for population proportions	47
Table B. 2:	Standard errors for population in age groups	47
Table B. 3:	Standard errors for marital status	49
Table B. 4:	Standard errors for ability to read and write	50
Table B. 5:	Standard errors for enrolment	50
Table B. 6:	Standard errors for highest level of education attained	51
Table B. 7:	Standard errors for usual activity with the past twelve months	52

1.

INTRODUCTION

This chapter presents a brief history of the censuses undertaken in the past and a rationale for conducting the Population Estimation Survey.

1 INTRODUCTION

1.1 History of census taking

The first population and housing census for Somalia was conducted in 1975, which published limited results; the findings from a second population census conducted in 1986/87 were not published officially.

In the absence of census figures efforts have been made over the years to estimate the size, age and sex distribution of the population, in addition to population projections. The most notable attempt was made by K.E. Vaidyanathan in his 1997 consultancy report '*Population Statistics of Somalia*', and the estimates have been generally accepted as the most reliable population data available. His report provided projected population for 2005 of just over seven million. Development agencies have made several attempts to compile reliable data on the size and distribution of the population and social and economic characteristics. Such efforts however, did not receive sufficient support and recognition.

In 2005, UNDP prepared a report entitled '*Population Estimates and Projection for Somalia, 2005-2010 (draft)*', which shows the estimated population for each year by sex and region. The estimates were based on a number of sources: information available from the 1975 census; a UNDP report on Population Statistics for Somalia, 1997; estimates made by UNHCR of the number of Somali refugees; and the various settlement surveys conducted by UNDP. These estimates were examined by a group of senior Somali professionals in the pre-war Ministry of National Planning, as well as members of the UN Thematic Group on Statistics. Both groups concurred with the population estimates presented for 2005. Their best estimate was a projection of population of just over 7.5 million in mid-2005. They projected the population based on an annual population growth rate of 3.0 percent, estimating a population of 8.4 million in mid-2010. They also estimated the distribution of the population by sex and region. These estimates were intended purely for planning purposes, as they were not based on the results of a full census. Currently, the official population estimates in use are derived from the '*Population Estimation Survey, 2014*' supported by UNFPA.

1.2 Rationale for conducting the Population Estimation Survey

The PESS is the first extensive large-scale household sample survey to be carried out among the Somali population in more than three decades.

For more than three decades, Somalis and their humanitarian and development partners faced immense challenges in designing and implementing programmes because of the lack or paucity of basic demographic data. As earlier stated, not all of the census data was released or published. It is worth noting that the available limited data from the census is now obsolete due to the numerous changes that have taken place in the political, demographic and socio-economic spheres. Development agencies made numerous attempts to produce reliable population figures pertaining to size, distribution and associated socio-economic characteristics. However, such efforts were limited to producing sector specific datasets.

1.3 Filling the data gap

The absence or scarcity of information on the Somali population challenged planning and programming at all levels for years. To address this situation, Somali authorities conducted a Population Estimation Survey in 2013/2014, with support from the donor and international community. The survey was aimed at providing population and socio-economic information to policymakers; and political, economic and social actors to develop evidence-based strategies for planning and decision-making.

The PESS is the first extensive large-scale household sample survey to be carried out among the Somali population in more than three decades. The survey provided reliable and comprehensive population estimates, demographic, and socio-economic characteristics for Somalis, encompassing the demarcated 1986 pre-war regions and districts.

The Population Estimation Survey was designed to provide accurate and reliable estimates of the size and distribution of the Somali population, and its characteristics, including: population size; spatial distribution; and socio-economic attributes. In addition, the PESS serves as a first milestone towards conducting a full and comprehensive population and housing census in the future.

The main objectives of the Population Estimation Survey were:

- To establish reliable estimates of the population size by age and sex living in urban areas, camps for Internally Displaced Persons, rural areas, and of nomadic communities.
- To empower and develop the capacity and foundation of government institutions and personnel responsible for data collection, analysis and dissemination.
- To provide estimates of the number of households, their geographic distribution and structure, along with related demographic and socio-economic data for the population.
- To provide sampling frames for surveys and a potential future population census.
- To provide baseline data for socio-economic planning, policy development, facilitating the evaluation of effectiveness, outcomes and impact of development interventions.

2.

SURVEY METHODOLOGY

The methodology for the Population Estimation Survey was developed through an all-inclusive and consultative process, led by Somali experts, supported by the UNFPA Technical Unit. This section provides the PESS methodology and application.

2 SURVEY METHODOLOGY

2.1 PESS planning and organization

Driven by the urgent need to establish an accurate and reliable source of population data and indicators, the Somali authorities requested the United Nations Resident Coordinator's office to conduct a Population Estimation Survey. As the UN's agency mandated to work on population, UNFPA took the lead in bringing together all stakeholders and mobilised resources for the survey.

UNFPA provided extensive support to Somalis to work on the survey.

The PESS provided a timely opportunity for UNFPA to build the capacity of Somali institutions and individuals in the planning and implementation of large-scale surveys. At each stage, from the inception of the survey onwards, UNFPA provided extensive support to Somalis to work on the survey. Somali authorities steered all the planning processes. All quality control and ethical issues were considered carefully at each stage.

The PESS brought together Somali authorities, Somalis who contributed to the survey in various capacities, and UNFPA. Together, they engaged in structured planning, and benefited from the efforts of the departments of statistics and individuals to collect and process representative and reliable data on the demographic and socio-economic characteristics of the Somali population across the country.

The important involvement of Somalis brought ownership of the survey and its findings. An integral by-product of these processes is the enhanced capacity of Somalis to conduct similar large-scale surveys independently in the future.

The 1986 pre-war geographic regions have been used for the analyses of data in this series. It is important to note that currently there is a Federal Government, comprising Puntland, South-West, Juba-land and Galmudug states. The Federal Government is in the process of forming new states with the local governments in Hiraan and Middle Shabelle regions. Somaliland declared itself as an independent republic in May 1991. As a result of these developments, regions and districts have changed. The newly established regions are not comparable to the pre-war regional and district boundaries used in this analysis.

2.2 Sampling frames

A sampling frame is a set of source materials about a target population from which a sample is selected. It provides a means for selecting the specific members of the target population that are to be interviewed in a survey.

The sampling frames for PESS had defined clusters of enumeration areas for the urban areas, settlements for rural areas, camps for IDPs and water points for nomadic areas. These were the primary sampling units.

2.2.1 Urban sampling frame

The sampling frames for urban areas were constructed from EAs ranging between 50 and 149 households. The list of households was obtained through a household listing exercise carried out by the PESS technical team. These EAs, which formed the PSUs, were classified as the lowest statistical sub-divisions available. The survey used validated maps and satellite images to identify administrative boundaries and enhance the quality of the urban frame.

2.2.2 Rural sampling frame

Settlements with 50 to 149 households in rural settings formed the PSUs for the rural sample. In line with this, larger settlements were split into segments of approximately 50-149 households.

The survey team updated a settlement list from the Settlement Census conducted by UNDP in

Clan elders verified information on inaccessible settlements, and satellite images further affirmed the information obtained from elders.

2005/2006, to obtain a master list of settlements to form the rural frame. The team also carried out field visits to verify the settlement list prior to its use. This process entailed physically counting the number of households in accessible areas. Clan elders verified information on inaccessible settlements, and satellite images further affirmed the information obtained from elders.

2.2.3 Internally displaced persons' sampling frame

The sampling frame for internally displaced persons (IDPs) consisted of displaced persons living in households among sedentary populations, as well as within camps. UNHCR provided latest (2013) information on IDPs living in camps for this frame. The frame included 107 IDP camps of which 28 were selected.

2.2.4 Water points' frame

This frame targeted the nomadic population and comprised water points, which are used by nomads during the dry seasons. The frame was based on the list of water points from the UNDP Community Census (2005/2006) and an updated list of water points prepared in 2012 by the UN Food and Agriculture Organization/Somalia Water and Land Information Management (FAO/SWALIM, 2012).

The survey team verified the listed water points that were being used by pure nomadic households to avoid double counting of settled agro-pastoralists at water points. All non-functional water points were eliminated from the list. The sampling frame for the nomadic population was then organised by type of water point before samples were drawn.

The frame originally consisted of 4,043 water points, but after the inclusion of several *hilos* (riverbanks) and *wars* (natural reservoirs) in the south and central regions, the size of the frame reached 5,332 water points, from which a simple random sample of approximately 14 percent was drawn.

The sampling frame for nomadic communities included nine types of water points: dug wells, boreholes, springs, *berkads* (man-made traditional water basins), dams, a combination of two or more types of water points, *hilos*, *wars* and 'others'. Each type of water point was considered as a substratum. This was based on the assumption that water points of a particular type are more homogeneous than water points of another type, and would cater for the watering needs of nomads in a similar manner. For example, boreholes would not be expected to dry up during severe drought and hence would cater for a larger proportion of nomads compared to dug wells. It was logical therefore to stratify the water points to take advantage of their similarities to improve precision.

2.3 Sample stratification, design and selection

2.3.1 Stratification

Stratification is a method of sampling that involves the classification of a population into groups. Stratification, if well constituted, facilitates the production of reliable estimates of different groups/strata. Strata are subgroups within the entire population that are thought to be relatively homogeneous or have common traits but heterogenous among themselves. Thus, the differences within a stratum are relatively small compared to the variation between strata.

Stratification and sub-stratification facilitated efficient sample selection. The PESS used three levels of stratifications. The first level of stratification was formed by the geographic areas of the 18 pre-war regions. The pre-war boundaries were selected as they were clearly defined and demarcated, and had been used in the 1975 census. Within each region, a second tier of substrata comprised urban areas, rural areas, and IDP settlements. Nomadic populations were stratified by type of water point.

2.3.2 Sample design

The PESS adopted an area stratified-cluster sample design, which entailed dividing the population into groups according to the areas they lived in.

The advantages of using a stratified-cluster sample design are:

a) Clustering reduced travel and other costs pertaining to data collection, compared to the application of simple random sampling of households.

b) Enumeration of households only took place in selected primary sampling units.

c) Stratification by regions was operationally convenient and economical.

d) Stratification made it possible to have crucial information on subgroups such as rural, urban, nomadic and internally displaced population estimates.

The size of the sample is the most important parameter of the sample design, since it affects precision, cost and duration of the survey. The reliability must be further considered in terms of the requirements for national versus regional and locality (urban and rural) estimation. Moreover, the overall sample size cannot be considered independently of the number and the size of sample areas (PSUs).

2.3.3 Sample size for urban and rural population

The calculation of the sample size is based on the known formula:

$$n = d_{eff} \frac{z^2 p(1-p)}{e^2}$$

Where;

n= required sample size of the units of target population;

p= expected rate or prevalence of the key indicator to be estimated; 50 percent, the rate that gives maximum size of sample was used

d_{eff} = design effect, a measure of how much the complex sample design used differs from simple random sample; Value of 2 used for this survey

e = margin of error to be tolerated at 95 percent level of confidence, set at 0.05 for this survey

z = critical value for the standard normal distribution which is 1.96 for 95 percent confidence level.

The size of the sample is the most important parameter of the sample design, since it affects precision, cost and duration of the survey.

Adjustment due to anticipated non-response; about 10% is used in most household sample surveys.

The estimated number of primary sampling units, n, based on the above formula is 845. This was multiplied by 3 (3 stratums of urban, rural and nomads) to give 2,535.

2.3.4 Sample allocation to regions and sub-strata

The allocation of sample sizes to the rural, urban, nomadic and IDP strata was proportionate, taking into account the number of primary sampling units in the respective frames. Subsequently the following sub-strata were formed: urban enumeration areas, IDP and rural settlements, and nomadic areas. The overall samples allocated to each region was proportionate to the size of the strata (number of enumeration areas), taking into account the total observations listed in the frame. In this case, the survey teams made lists of the number of households during the field-mapping phase, before the main household survey was conducted.

The sample was initially fixed at 2,535 PSUs. The total number of PSUs in the frames was 18,708. However, sample sizes were readjusted to boost representation for regions that had few PSUs in the initial allocation. Consequently, the allocation to the urban strata was adjusted to 868, the rural sedentary substrata was adjusted to 1,104, the water points for estimating the nomadic population were adjusted to 735, and the IDPs to 28, resulting in total of sample of 2,735 PSUs.

Following sub-strata were formed: urban enumeration areas, IDP and rural settlements, and nomadic areas.

The distribution of the samples is shown in Table 1.

Table 1: Final allocation of primary sampling units by the four strata

Type of residence	Total frame	Proportional allocation	Adjusted re-allocation
Urban	6,750	969	868
Rural	6,519	936	1,104
Nomadic population	5,332	735	735
IDPs	107	15	28
Total	18,708	2,655	2,735

2.3.5 Selection of PSUs

The selection of samples in the rural, urban and IDP strata was based on the 'Probability Proportional to Size' approach using a systematic random selection procedure. For each of the PSUs in the frame, an estimated number of households was obtained. The number of households for each PSU was taken into account in a way that provided a lower probability for PSUs with fewer households to be included in the sample, and vice versa. The following formula was used in selecting PSUs:

$$P_{hi} = \frac{n_h \times M_{hi}}{M_h}$$

Where;

P_{hi} = Probability of selection for the i^{th} sample PSU in stratum h

n_h= Number of sample PSUs assigned to stratum h

M_{hi}= Total number of households for the i^{th} sample PSU in stratum h

M_h=Total number of households in the frame for stratum h

Stratum h in the case of PESS could be the region,rural or urban area

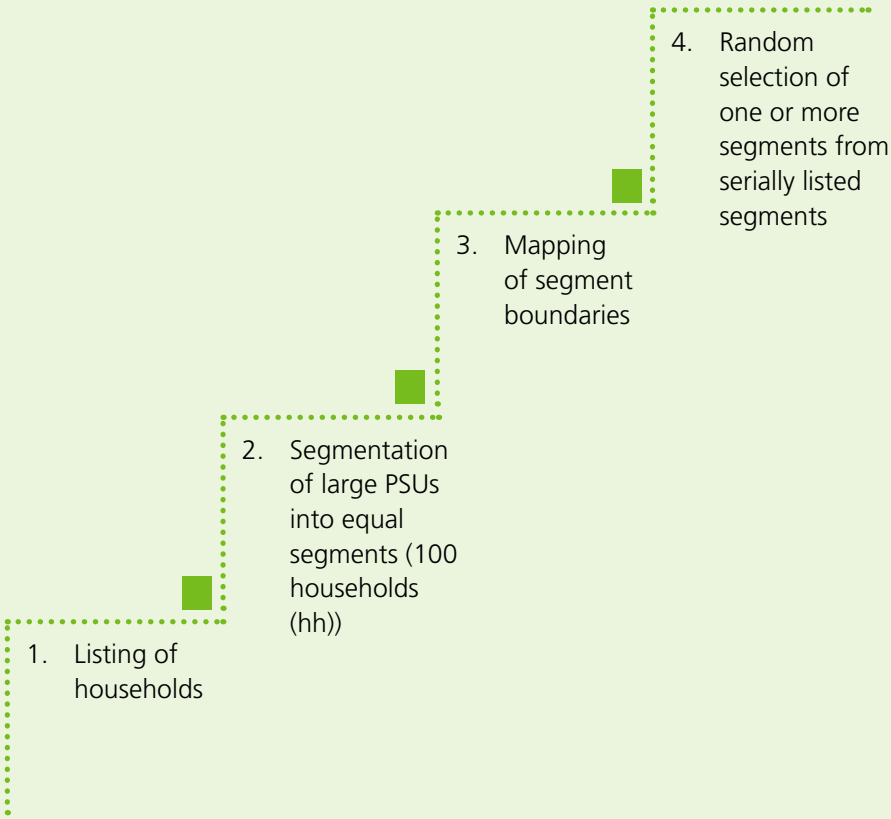
The essence of the formula was each settlement having its own probability of selection. For example, a smaller PSU had a lower probability of selection, but a larger sampling weight if selected because the weight is the reciprocal of the selection probability. This implies that the weighting, to the largest extent possible, ensured unbiased results, since the sample values in each selected settlement were multiplied by their respective weights.

2.3.6 Selection of segments from large settlements

For settlements with more than 149 households, special procedures were followed to segment them. Large settlements were segmented through proper weighting. Segmentation was only necessary in some rural areas and IDP settlements. In urban areas, towns were already divided into EAs of approximately 100 households before sample selection (Figure 1).

For settlements with more than 149 households, special procedures were followed to segment them.

Figure 1: Steps followed in segmenting large rural settlements



2.3.7 Sample selection of water points

For the nomadic samples, the allocated number of water points within each region and substratum were selected from the respective lists using a systematic random sampling approach, giving each water point an equal selection probability. Since the Single-day Model (UN, 1977) was applied in the survey, the date for coverage of each sample water point was determined randomly within an interval of 12 days. The 12-day period was determined to give nomads with the different types of animals, including camels, an equal chance of selection into the sample (it takes 12 days on average for a camel to go without drinking water).

Enumeration of nomads took place in water points which were selected based on simple random sample (SRS) probability sampling plan as opposed to probability proportional size (PPS)

While listing the water points, it was difficult to determine the estimated number of nomadic households that would be using a given water point during the dry seasons, which meant it was not possible to establish accurate measures of size. As a result, enumeration of nomads took place in water points which were selected based on simple random sample (SRS) probability sampling plan as opposed to probability proportional size (PPS). This sampling plan (SRS) gave equal probability of selection to each water point within each group (as earlier stated, water points in each region were grouped by type).

The PESS adopted a one-stage stratified cluster sample design because the probability selection was applied to primary sampling units only. At the second stage, within each selected primary sampling unit (urban enumeration areas, IDP and rural settlements and water points), all households and persons were enumerated.

Table 2: Water points substrata

Water points	Substrata
Dug wells	192
Boreholes	81
Springs	53
Berkads	121
Dams	64
Mixed (water points consisting of more than one type)	77
Hilos	33
Wars	77
Other (Sources not mentioned above e.g. roof catchment, rock catchment)	37
Total	735

2.3.8 Population count

All the individuals who usually live in a household were enumerated as members of the household during the PESS. The application of the concept of 'usual residence' at the household level gave the de jure population. The de jure population of a defined geographical area consists of all usual residents, whether or not they are present in the household at the time of the survey. "Usual residence" was defined, for the purpose of the survey, as the place at which an individual lives at the time of the survey, has been there for some time or intends to stay there for some time.

For each accessible PSU that was included in the sample, complete enumeration of the individuals in all the households was conducted, with a household response rate being 98.8 percent. The initial PESS report published in October 2014 examined population size of Somali citizens by sex and age, as well as their distribution patterns among the 1986 pre-war regions.

2.4 Data collection

Data collection covered populations living in urban, rural and nomadic areas, and in settlements for internally displaced persons. The survey was carried out in urban, rural and IDP areas between November and December 2013, as this period is believed to be the time when the non-nomadic population is least mobile. The characteristics of the population were recorded as at the time of enumeration. Due to the onset of rains, the survey for the nomadic population had to be postponed until the peak of the dry season, as the water points may not have been an adequate sampling point during the rains due to greater accessibility of water from other sources. The survey at the water points was conducted in all administrative zones for a period of 12 days in March 2014. Survey teams were assigned to each selected water point to enumerate nomadsonly. The teams moved from one water point to another.

Data collection covered populations living in urban, rural and nomadic areas, and in settlements for internally displaced persons.

The primary observation units for the PESS were households. Data was gathered on the number of persons in the household and household characteristics. Households and persons living in institutions such as barracks, hospitals, hotels, boarding school hostels and guesthouses were excluded from the survey. The principal respondent to the household questionnaire was the household head, but in his/her absence, the oldest family member available was interviewed. For the nomadic population, family members who came to draw water were interviewed as household representatives.

The survey team for each of the three Somali authorities included a PESS Director, Deputy Director, Regional Coordinators, District Coordinators and Supervisors and Enumerators. Directors of Statistics, Task Forces and experts in different fields supported the survey team. The PESS Survey Director worked closely with the PESS Deputy Survey Director and the Director of the Statistical Department in carrying out the survey operations, including advocacy conducted with each of the three Somali authorities. Regional Coordinators were appointed to manage Field Supervisors in each of the 18 regions, with Assistant Regional Coordinators providing support in the larger regions. Field Supervisors led survey teams consisting of Household listing Personnel, Field Editors and Enumerators.

In sedentary clusters, the interviewing of households was carried out soon after listing. The Supervisor assigned households to Enumerators shortly after the Listing Personnel had completed the listing of households.

2.4.1 Selection and training of field staff

Staff of the UNFPA Technical Support Unit conducted 'Training of Trainers'. Two persons were selected from each region for the training. The seven-day training sessions covered survey design, logistics, and a detailed coverage of items contained in the questionnaire, coupled with mock interviews. The Trainers then trained Supervisors and Interviewers, who were recruited locally in their respective regions. In total, 4,500 Somali men and women were trained in basic mapping skills, data collection and data entry. Manuals were produced for use in the training, and as reference materials in field data collection and editing.

In total, 4,500 Somali men and women were trained in basic mapping skills, data collection and data entry.

A refresher course of trainers was conducted before surveying the nomadic population at water points. This training focused on the listing and screening of households visiting the water points. The household questionnaire used for nomadic and non-nomadic population was similar. The questionnaire contained questions which were unique to nomads and non-nomads. The trainees who performed well in the survey in sedentary areas in November/December 2013 were selected to serve as interviewers during the nomadic survey.

2.4.2 Stages of data collection

Data for the PESS survey was collected in three stages listed below:

a) Cartographic field mapping

b) Household listing

c) Carrying out interviews in each household in each selected PSUs.

The field-mapping phase began early in 2013, led by of a Geographical Information System (GIS) Analyst and a Cartographer who trained trainers of field mapping teams.

The objectives of the field mapping exercise were to:

a) Produce the frame from which the sample would be drawn

b) Facilitate the planning for the main PESS household survey

c) Gather information on the number of households in enumeration areas, subsequently used as measures of size when selecting PSUs in urban, rural and IDPs

d) Collect information (such as the number of households, EAs etc.) that could be used in combination with data from the sampled areas to be able to produce estimates of lower-level administrative domains, such as districts

e) Collect baseline information for geo-referenced population data.

During the 'listing phase' of the survey, lists were compiled of all households from selected clusters or enumeration areas. This exercise involved (i) identifying the exact boundaries of the primary sampling units; (ii) listing all housing structures; (iii) identifying dwelling units within the housing structure; and (iv) listing households within the dwelling units. For each household, initial information was collected, such as the name of the head of household and the number of males and females who belonged to the household. The listing was undertaken shortly before the actual interviews took place.

Interviews covered all households in the selected EAs, settlements, and those reported at the water points. The interview teams gathered information using structured questionnaires that were translated from English into Somali.

The interview teams gathered information using structured questionnaires that were translated from English to Somali.

2.4.3 The household survey questionnaire

As earlier stated, the survey used one questionnaire to collect information from households in urban, rural and IDP settlements, as well as among the nomadic population. The questionnaire included questions on household listing, individual characteristics, population mobility, demographic events, socio-economic characteristics, housing characteristics and household assets.

Specifically, the questionnaire was designed to collect information on demographic and socio-economic characteristics pertaining to the survey population, including:

a) The size and geographical distribution of the population

b) Basic demographic and socio-economic characteristics of the population (including age-sex distribution, marital status, mortality, literacy, education and economic activity)

c) The number and geographical distribution of households by heads, size and composition

d) Housing and living conditions in the households (living arrangements, access to water, energy and telecommunication).

2.5 Survey limitations

The use of the survey methodology introduces sampling errors, i.e. the population estimates derived from such a sample cannot be the same as those which would have been derived from a census.

The non-responses encountered during the survey are likely to have introduced some level of bias which may not be entirely addressed through the adjustment of the base weights. For instance, the rural and nomadic strata had low coverage rates. This meant that adjustments had to be made to the weights to factor in non-response.

It was not practically possible to visit some areas to conduct physical household listing and survey due to insecurity. For these areas, the PESS team relied entirely on estimates from the digitised imagery.

The field teams interviewed households from both sampled and non-sampled IDP camps and the distinction between the two was not clear. PESS, however, uses the filled questionnaires to compute various indicators. Due to this, the UNHCR population estimates were adopted for the IDP population.

UNHCR population estimates were adopted for the IDP population.

A 'single-day model' was used to enumerate the nomadic population as opposed to a 12-day extended period model which is the estimated watering period for camels. The extended-period model is an approach in which interviewers are placed at each selected water point and households arriving to water their animals for the first time during the period of enumeration are included in the study. The single-day model, on the other hand, is one in which an interviewer attempts to enumerate all nomadic households who appear at a selected water point on one day of the 12-day enumeration period that has been randomly assigned to the water point (Kalsbeek and Cross, 1982).

Due to the magnitude of the survey and the prevailing security situation, it was not logistically possible to predict all non-response/refusals accurately and make appropriate replacements in all the EAs. This may have biased certain estimates.

3.

DATA PROCESSING

This chapter describes the stages involved in data processing such as checking for completeness, coding, capture, editing and tabulation of data.

3 DATA PROCESSING

3.1 Data centres

The completed household questionnaires and listing forms were put into folders according to survey areas and sent to processing centres in Hargeisa, Garowe and Mogadishu. These batches of documents were stored and well maintained throughout the data processing exercise. Each record had a unique identification code. Data processing included checking for completeness, coding, capture, editing and tabulation of data.

Data processing included checking for completeness, coding, capture, editing and tabulation of data.

Data processing began after the receipt of questionnaires by the various data centres. The questionnaires were manually validated to ensure that collected items had appropriate codes before data entry and data cleaning, construction of weights, and weighting of the sample data. Following these processes, tabulations were generated. The tabulations were then disseminated through reports of preliminary results and in-depth analysis (Figure 2).

3.2 Data coding

Coding is the process of assigning numerical values to responses to facilitate data capture and processing. The coding of data from the PESS survey entailed assigning numerical codes to responses from the questionnaires that had been recorded in words or in other forms that may have required modification before data entry. Numerical codes that had already been assigned and recorded were retained. Questions on 'occupation' and 'countries of origin' required the interviewers to carry and use pages of pre-coded model responses.

3.3 Data entry

Data from the PESS was keyed into computers using a data entry programme developed from the Census and Survey Processing System software. In each of the three Somali authorities one data entry team was designated and trained. They were responsible for keying in data for their respective zone. With respect to both data coding and entry, emphasis was placed on error prevention.

3.4 Data editing

To obtain reliable survey results, data must be free from errors and inconsistencies to the greatest extent possible. Data editing is the process of detecting errors made during and after data collection and entry, following which adjustments are made to improve data quality. During data collection, the Supervisors worked with Field Editors or Listers to ratify the quality of work conducted, including coverage by the Enumerators. Field office Editors would check the questionnaires for any missing information before keying in data.

Editing aims to make the data as accurate as possible, while ensuring that it remains as close as possible to the respondents' answers and reality. Predetermined rules for validation and correction were put in place, such as rules for editing syntax, which hastened the process of editing.

In summary, data was edited to ensure the validity and consistency of individual records, relationship among records in a household (known as micro-editing), and to check the plausibility of the aggregated data (known as macro-editing). At micro-level for example, unedited data may contain information that is highly unlikely or impossible, such as a one-year-old child being recorded as married or a male who reported giving birth in the last year. Therefore, the editing of the PESS data reduced distortions.

3.5 Tabulation

In preparing a tabulation plan, reference was made to the household survey questionnaire and the standard tabulations provided in the *'United Nations Principles and Recommendations on Population and Housing Censuses, Vol.2'* (UN Department of Economic and Social Affairs, 2008).

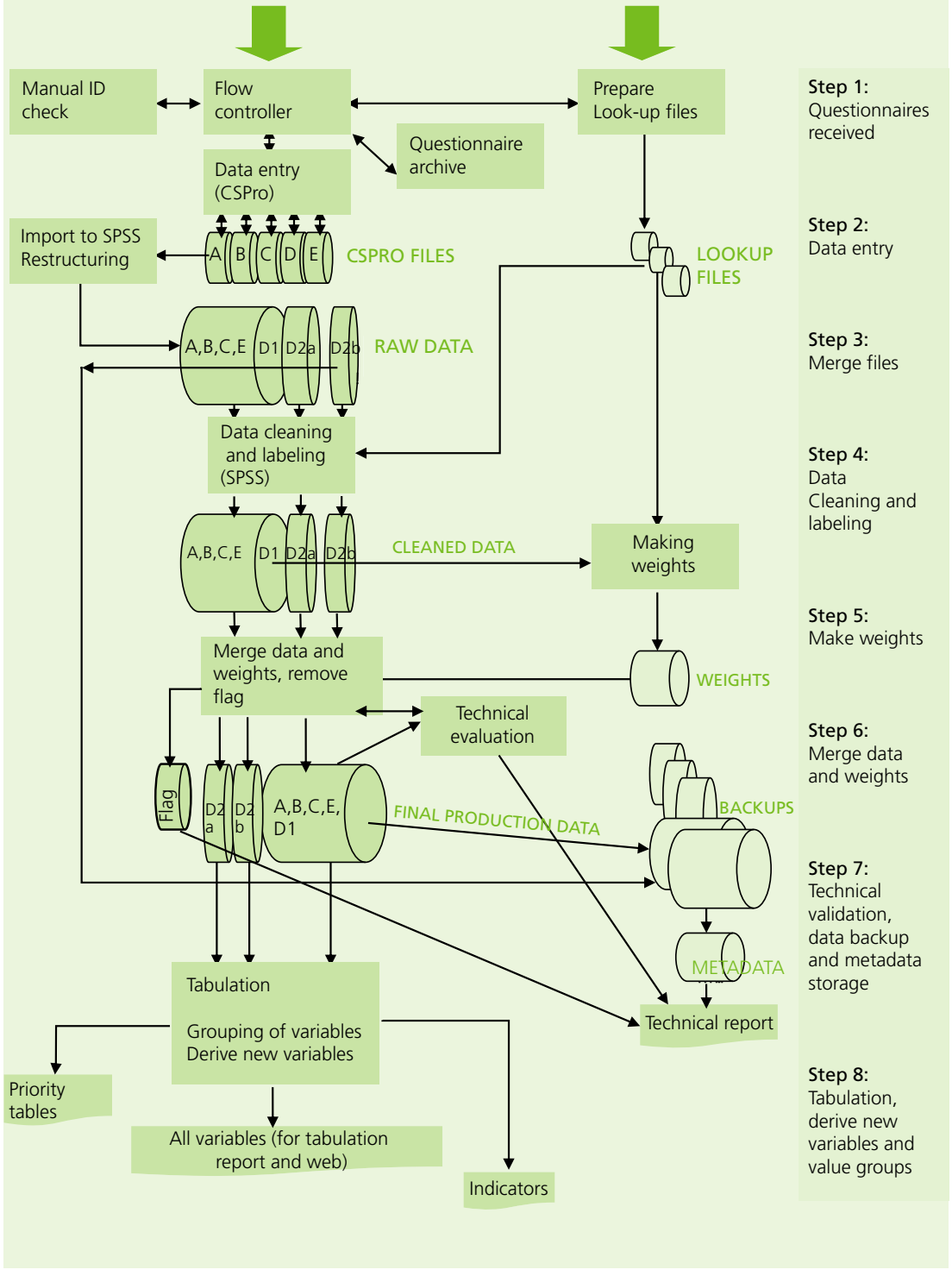
Using SPSS, dummy tables were drafted for various tabulations. The indices were computed using the Population Analysis Software (PAS) developed by the US Census Bureau. The dummy tables were subsequently filled in with relevant survey results.

3.6 Dissemination

Data dissemination entails information being published or shared with users in a format that they can understand. Preliminary data from the PESS 2014 was disseminated through the first published report, *'Population Estimation Survey'*, 2014. This report is available on the internet and in hard copies.

Preliminary data from the PESS 2014 was disseminated through, *Population Estimation Survey'*, 2014.

Figure 2: Flowchart for data edits and tabulation for PESS



4.

ESTIMATION

This chapter explains the procedures used to estimate values for the total population based on the sample results.

4 ESTIMATION

4.1 Estimation

Estimation is the means by which values are obtained for the total population based on this information gathered from a sample of this population.

In the case of the PESS, estimations were made by substratum, namely urban, rural, and nomadic strata. Information on IDPs was weighted using UNHCR data. In general, sample survey data was multiplied by the respective sampling weights to obtain estimates for the total population.

The weight is the reciprocal of the selection probability

Thus $$\frac{1}{P_{hi}} = \frac{M_h}{n_{hi} \times M_{hi}} = w_{hi}$$

Where;

P_{hi}=Probability of selection for the i^{th} sample PSU in stratum h

n_h=Number of sample PSUs selected in stratum h

M_{hi}=Total number of households for the i^{th} sample PSU in stratum h

M_h=Total number of households in the frame for stratum h

The stratum h in the case of PESS could be the region,rural or urban area

Wherever the selection was carried out in stages, such as computing the weight for a sampled segment of a large settlement, the weight was the reciprocal of the product of selection probabilities.

Where the selection was carried out in stages, such as computing the weight for a sampled segment of a large settlement, the weight was the reciprocal of the product of selection probabilities.

4.2 Adjustment for non-coverage and non-response

Non-response refers to cases where data for a sample observation unit is missing, because households refused to respond or could not be reached. Also in certain areas, entire EAs could not be covered, generally because of insecurity. In both cases, the sample design weights were adjusted to attain credible population distributions.

The coverage rate was highest in the urban strata.

Table 3 shows the PSUs coverage rate for urban, rural and nomadic populations. Overall, 69.6 percent of the sampled PSUs were covered. The coverage rate was highest in the urban strata (99.7 percent), and lowest in the rural strata (50.8 percent). The nomadic strata reached a coverage rate of 62.4 percent. Coverage for IDP camps was not computed due to difficulties in establishing which clusters some of the filled-in questionnaires were from. The field teams interviewed households from both sampled and non-sampled IDP camps. Due to this, the UNHRC population estimates were adopted for the IDP population. PESS used information collected from households in IDP camps to compute various indicators. A factor was applied to IDP PESS data to extrapolate to the UNHCR figures.

Table 3: Survey coverage

Type of residence	Sampled PSUs	Covered PSUs	Coverage rate
Urban	868	865	99.7
Rural	1,104	561	50.8
Nomadic population	735	459	62.4
Total	2,707	1,885	69.6

5.

QUALITY ASSESSMENT AND ADJUSTMENT OF AGE AND SEX DATA

This chapter presents information on the evaluation of the quality of the age and sex data, as well as methods used to determine errors made while collecting information.

5 QUALITY ASSESSMENT AND ADJUSTMENT OF AGE AND SEX DATA

5.1 Introduction

This chapter assesses the quality of age information in the PESS. In addition to the accurate presentation of information on age, the examination of age data is important to assess the plausibility of the derived estimates. A quality assessment report of the Demographic and Health Surveys showed that age misreporting and under-reporting often lead to inaccuracies in the total fertility rate and underestimation of the under-five mortality rate (Institute for Resource Development, 1990).

A number of authors have documented sources of inaccuracies in age reporting in censuses and surveys. These include: ignorance of the exact age; conscious or subconscious preference for certain ages; carelessness in reporting and recording; and tendency to round-up ages (Ewbank, 1981; Kpedekpo, 1982; Nagi, Stockwell, & Snavely, 1973). Consequently, population age distribution often displays a systematic pattern of irregularities such as: a) fewer infants and very young children, b) overstatement of seniority among the elderly, c) overstatement of certain socially significant ages, and overstatement of ages ending in certain preferred digits with a corresponding understatement of ages ending in other digits (Geralnd 2015).

Data from surveys and censuses depict significant clustering at ages ending in digits "0" and, to a lesser extent, "5", and corresponding deficiencies at ages ending in digit "1" and "9". The tendency of reporting certain ages at the expense of others is known as 'age heaping', 'age preference' or 'digit preference'. Age heaping is most noticeable among population sub-groups with low literacy rates (Bekele, 2006; Dahiru & Dikko, 2013).

The disaggregation of population figures by age and sex guides social and economic planning in a country. By analysing changes and forecasting future patterns in the age and sex distribution of a population, policymakers can identify the consequences of population dynamics on the education sector and the socio-economic situation of the population. For example, by studying labour force participation of relevant age groups by sex, educational and economic policies that take into account gender relations can be developed in order to enhance employment opportunities.

In most developing countries, it is difficult to obtain accurate data on the age of respondents and the Somali context is no exception, particularly as the civil war eroded all systems and institutions that may have stored information. Most respondents were not aware of their ages. The enumerators prompted respondents with well-known historical events and their calendar dates, with the aim of getting them to relate to events that would be of assistance in determining their ages. Additionally, the age of respondents was also imputed during data editing, using established techniques such as 'hot-deck'. Hot-deck uses the characteristics from a similar respondent whose information is available to fill in missing data.

5.2 Assessment of age heaping

Great advances have been made in techniques for evaluating the accuracy of age reporting. These techniques are used for understanding anomalies of age and sex data. The most common ones include an analysis of age ratios, sex ratios, Whipple's Index, Myer's Blended Index, Bachi's Index and the United Nations Age-Sex Accuracy Index. These indices inform about divergence from or conformity to a normal age and sex distribution (Bekele, 2006). However, for cases where the actual age and sex distribution is unusual due to social or structural reasons such as war or out-migration, the inaccuracies flagged by those evaluative indices should not be automatically considered to be data errors (Siegel & Swanson, 2004).

To evaluate the extent of inaccuracies in age and sex reporting in PESS data, the single-age and sex distribution, Whipple's Index, Myer's Blended Index and UN Age Sex Accuracy Index were applied. The indices were computed using the Population Analysis Software developed by the US Census Bureau.

To evaluate inaccuracies in age and sex reporting, the single-age and sex distribution, Whipple's Index, Myer's Blended Index and UN Age Sex Accuracy Index were applied.

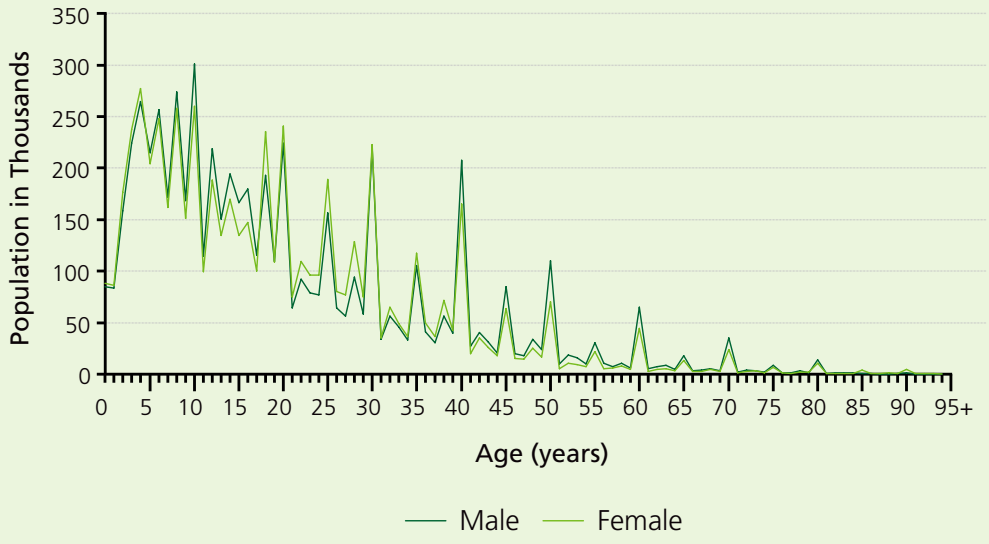
5.2.1 Single year age and sex distribution

Information on age was plotted in single years to identify age heaping or digit avoidance.

During the analysis of the PESS data, information on age was plotted in single years to identify age heaping or digit avoidance. In the absence of drastic changes of fertility and mortality and migration between ages, and assuming that age is accurately reported, it is normal to observe a systematic decline of population numbers. The age data of the Somali population shows declining numbers but peaks at ages ending in zero and five for both males and females, as is evident from Figure 3.

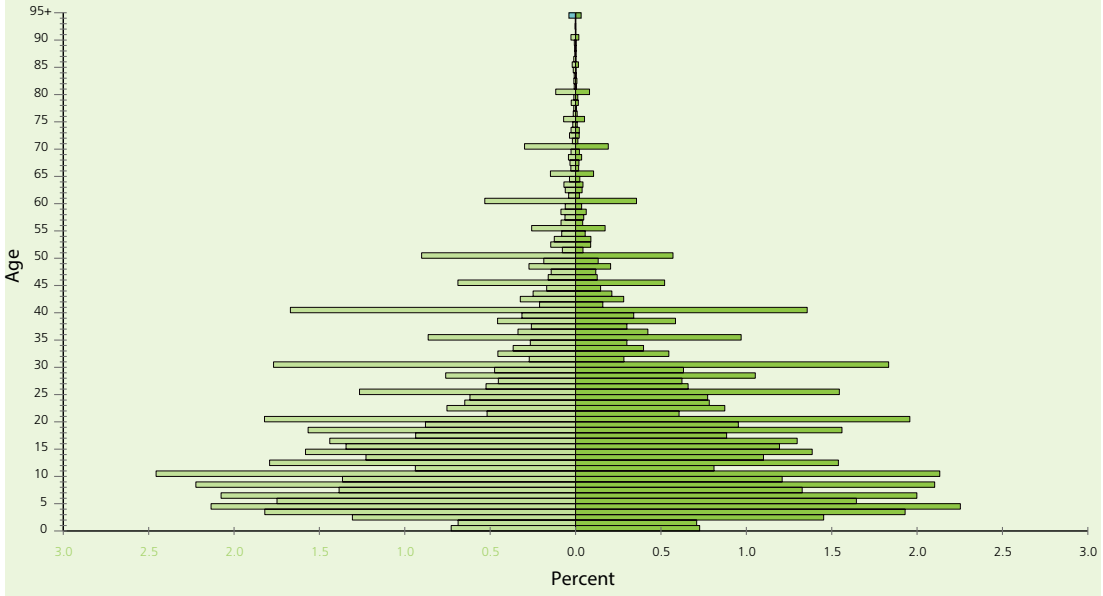
The peaks are alternated by dips, resulting from digit avoidance in favour of the preferred digit. Digit preference occurs for most ages ending with 0 and 5, thus, 10, 15, 20, 25, up to 90, except age 5 which has a dip, indicating an undercount. A few ages ending with even numbers and odd digits have been preferred, namely, 6, 12, 18 and 23 among others. Ages 0-4 have shown similar under-reporting in most censuses (US Census Bureau, 2014).

Figure 3: Population by age in single years



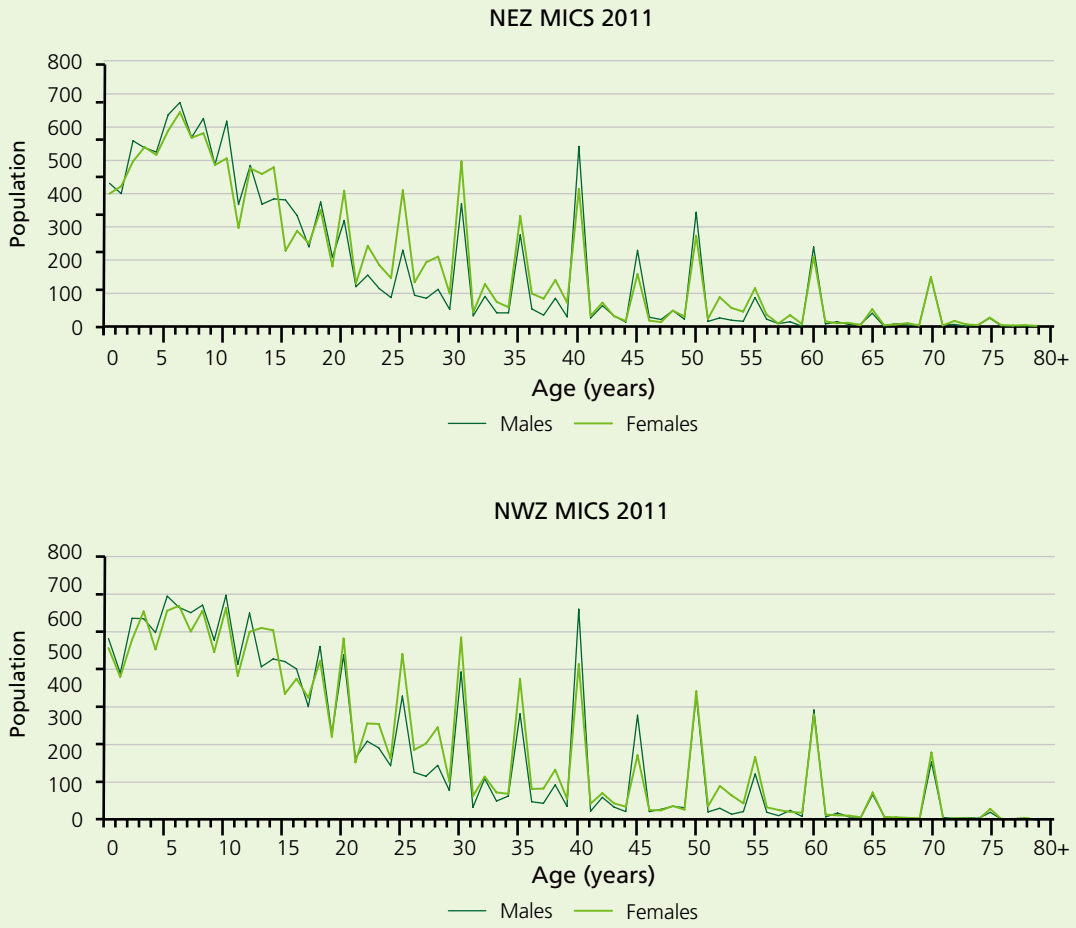
Age heaping can also be assessed by use of a population pyramid as shown in Figure 4. The longer bars show particularly pronounced age heaping in ages ending with zero, and to a relatively lesser degree in ages ending with five.

Figure 4: Population by single years of age and sex



The single years of age graphs for the *'Multiple Indicator Cluster Survey (MICS) 2011'* for North-East Zone (NEZ)[1] and North-West Zone (NWZ) display a similar pattern of preferences for zero and five-year age digits in both males and females, as revealed by the peaks in Figure 5. The peaks are seen in more ages for males than females, indicating increased rounding off of age in males compared to females. Figure 5 shows the age distribution in single years for NEZ and NWZ, computed from MICS 2011 data (UNICEF, 2014a, 2014b).

Figure 5: Population by age in single years, MICS 2011



1 The UNICEF Multi-Indicator Cluster Survey 2011 refers to zones such as the North-West Zone (NWZ) and North-East Zone (NEZ). The survey was not conducted for the South-Central Zone (SCZ).

5.2.2 Whipple's Index

The Whipple's Index is the ratio of the total number of persons between ages 23 and 62 who report ages ending in 0 and 5, to one-fifth of the total population in the same age group, multiplied by 100. To obtain the Whipple's Index, the number of persons in this age range is summed up, as the ratio of reported ages ending in 0 or 5 to one-fifth of the total population under consideration (Spoorenberg, 2009).

The following formula is used to compute Whipple's Index:

$$\frac{\sum P_{25} + P_{30} + \ldots + P_{55} + P_{60}}{1/5 \sum (P_{23} + P_{24} + \ldots + P_{60} + P_{61} + P_{62})} \times 100$$

The selection of the range 23 to 62 years is largely arbitrary. In computing indices of heaping, childhood ages and old ages are often excluded, as they are more likely to be affected by other types of errors of reporting rather than by preference for specific terminal digits. The younger ages (0-4 years) are often under-reported, while the older ages may be exaggerated and have few persons in these category due to high mortality.

A score of 100 indicates no age heaping on 0 or 5, whereas a score of 500 indicates that every age reported ends in 0 or 5. The Whipple's Index scores can be summarized through categories proposed by the United Nations as shown in Table 4 (United Nations, 1973). The Whipple Index is one of the widely used indices to measure age misreporting.

Table 4: Degree of accuracy of the age reporting using the Whipple's Index

Whipple's Index	Quality of data	Deviation from perfect
< 105	Very accurate	< 5 percent
105 –110	Fairly accurate	5 - 9.99 percent
110 –125	Approximate	10 - 24.99 percent
125 –175	Rough	25 - 74.99 percent
> 175	Very rough	> = 75 percent

The PESS age analysis illustrates pronounced age heaping of the population for the ages ending with digits 0 and 5

Table 5 compares Whipple's Indices for males, females and both sexes based on the 2014 PESS and the MICS data for NEZ and NWZ collected in 2011, as well as the Whipple's Index reported for selected African countries. The PESS age analysis illustrates pronounced age heaping of the population for the ages ending with digits 0 and 5, with a Whipple's Index of 238 and 216 for males and females respectively. This shows that the quality of single year age data is very rough for both sexes. In all surveys, there is more age heaping evident among males compared to females. However, the MICS survey for the North-East and North-West zones displays more pronounced age heaping compared to PESS. In comparison with neighbouring African countries, the quality of age data captured is among the poorest, as demonstrated by the Whipple's indices. With the exception of Sudan, other countries had much lower Whipple's indices compared to what had been reported in the PESS. This indicates lower quality of single year age and sex data from PESS compared to data from other countries.

Table 5: Comparison of Whipple's Index for Somalia and other countries

Country	Year	Whipple's Index		
		Male	Female	Both sexes
Somalia*	**2014**	**238**	**216**	**227**
Ethiopia	2007	-	-	146
Kenya	2009	148	145	147
Sudan	2008	249	229	239
South Sudan	2008	174	174	174
Tanzania	2012	153	156	154
Uganda	2002	138	131	135
Rwanda	2002	106	108	107

Source: Fajardo-Gonzalez et al, 2014; World Academy of Science, Engineering and Technology, 2015
* PESS 2014

5.2.3 Myer's blended method (Myer's Index)

Myer's Blended Method of determining errors in age-sex data is based on the principle that in the absence of age heaping, the aggregate population of each age ending in a digit from 0 to 9 should represent 10 percent of the total population. Myer's Blended Method is usually used to measure the degree of preference for each digit; it provides a 'summary index' for all terminal digits. The summary index is an estimate of the minimum proportion of persons in the population for whom an age with an incorrect final digit is reported (Hobbis, Siegel, & Swanson, 2004). It lies in the range of 0 and 90, with an index of 0 meaning absence of preference for any digit and the extreme value of 90 indicating preference of one digit.

The Myer's Blended Method avoids the biases associated with the Whipple's Index (does not cover all digits i.e. 1-4 and 6-9) that may occur when presenting the population age structure or distribution for reasons other than heaping or preference. In other words, numbers ending in "0" would normally be larger than those ending in "1" to "9" owing to natural attrition or mortality (Siegel & Swanson, 2004). The calculation of Myer's Blended Index entails determining the proportion that the population ending in a given digit ("0" through "9") is out of the total population, by varying the particular starting age for any 10-year age group.

The index is determined by calculating the number of people whose age ends with a particular digit for the population aged 10 and over, and then for the population aged 20 and over. Each series is then weighted and the results are added to obtain a blended population. Myer's Blended Index is obtained by summing the absolute deviations between the aggregate and theoretical distributions (10 percent). The process for computing the Myer's index is illustrated in the Table A.1 and A.2 (Appendix A).

Table 6 shows a comparison of Myer's Index for the PESS and UNICEF MICS 2011 for the North-East and North-West, and that of censuses conducted in other countries.

The Myer's Blended Index calculated for PESS was 22 for males, 21 for females and 22 for both sexes. This shows that around one-fifth of the population had their ages reported with an incorrect final digit. Myer's Index from PESS is slightly better than Tanzania and Sudan but falls within the same range.

The preference for and avoidance of specific digits is graphically presented in Figure 6. The positive and negative scales represent digit preference and avoidance respectively. The most preferred digit according to the results was 0 followed by 5 and 8. The most avoided digit was 1, followed by 7.

The Myer's Blended Index shows that around one-fifth of the population had their ages reported with an incorrect final digit.

Table 6: Myer's digit preference for selected African countries

Country	Year	Myer's Index		
		Male	Female	Both sexes
Somalia	**2014**	**22**	**21**	**22**
Ethiopia	2007	-	-	-
Kenya	2009	7	8	7
Sudan	2008	22	25	23
South Sudan	2008	16	16	16
Tanzania	2012	24	25	25
Uganda	2002	7	8	8
Rwanda	2002	2	2	2

Source: Fajardo-Gonzalez et al, 2014; World Academy of Science, Engineering and Technology 2015

UNICEF MICS 2011 data for North-East Zone and North-West Zone exhibits a similar age misreporting pattern, with more age heaping occurring on the terminal digits of 0 and 5, while respondents avoided virtually all other digits, particularly the digits 1 and 9.

UNICEF MICS 2011 data for North-East Zone and North-West Zone exhibits a similar age misreporting pattern, with more age heaping on 0 and 5

Figure 6: Myer's Digit Preference

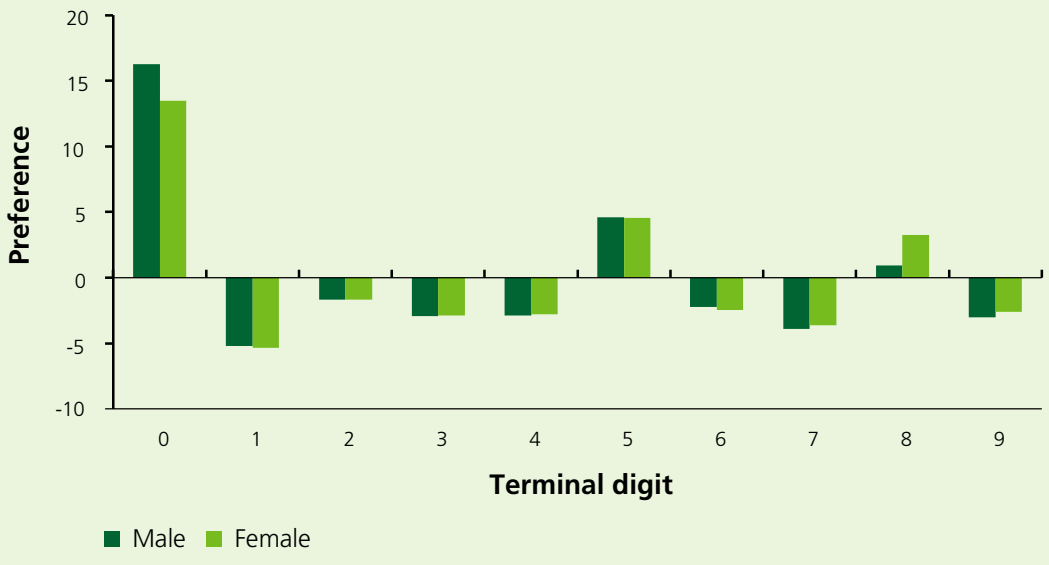
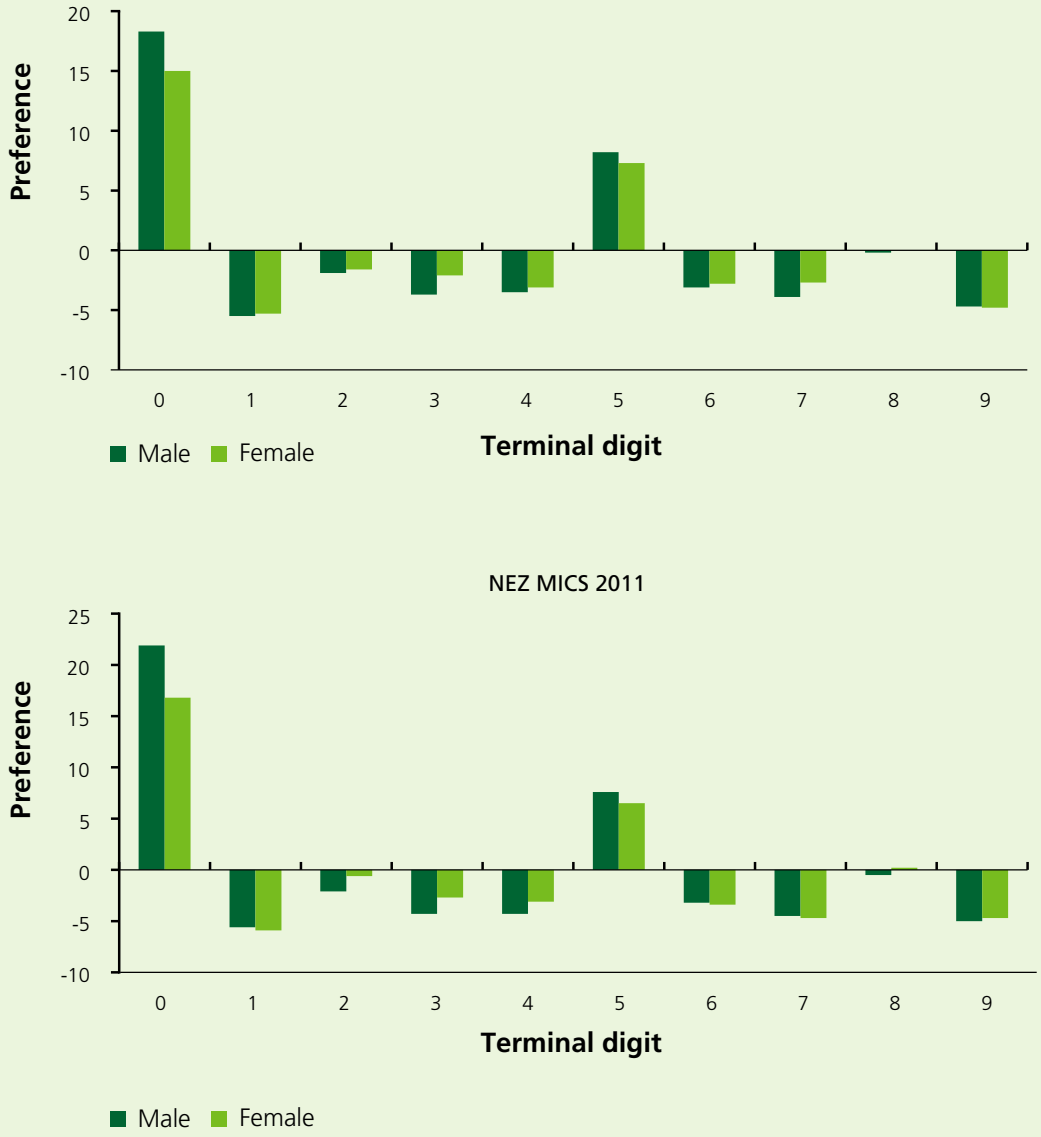


Figure 7: Myer's digit preference, MICS NWZ and NEZ



5.2.4 UN Age-Sex Accuracy Index or UN Joint Score

The UN Joint Score combines the Sex Ratio Score and the Age Ratios Score, providing an overall index on data quality. The Age Ratio Score is the average of the sum of absolute deviations of age ratios from the expected 100. The age ratio for a particular age group to the average of the counts for the adjacent cohorts should be approximately equal to 1, or 100 if multiplied by a constant of 100 (United Nations Statistics Division, 2011). The age ratio is the proportion of the population of a given age group, to the average of the number of persons in the immediately preceding and succeeding age groups.

The United Nations Statistics Division, 2011, indicates that in the absence of sharp changes in fertility over the years, mortality by age, significant levels of migration or other distorting factors, the enumerated size of a particular cohort should be approximately equal to the average size of the immediately preceding and subsequent cohorts. Significant departures from this "expected" ratio indicate either the presence of census error in the census enumeration or other factors. (United Nations Statistics Division, 2011).

The age ratio is the proportion of the population of a given age group, to the average of the number of persons in the immediately preceding and succeeding age groups.

Age ratio for the age category x to x+4

$$ {}_5AR_x = \frac{{}_5P_x}{\frac{1}{3}\left({}_5P_{x-5} + {}_5P_x + {}_5P_{x+5}\right)} \times 100 $$

$_5AR_x$ = Age ratio for the age group x to x+4

$_5P_x$ = Enumerated population in the age category x to x+4

$_5P_{x-5}$ = Enumerated population in the adjacent lower age category

$_5P_{x+5}$ =Enumerated population in the adjacent higher age category

The sex ratio is the number of males for every 100 females. The sex ratio score is defined as the average of the sum of absolute deviations of consecutive sex ratios. For most sub-Saharan African populations, the sex ratio fluctuates between 101 and 105 at birth, thereafter declining gradually with age below 100. Any major deviations from this pattern can be attributed to age and sex misreporting, and/or to age-sex selective migration or mortality.

For most sub-Saharan African populations, the sex ratio fluctuates between 101 and 105 at birth declining gradually with age below 100

The sex ratio is computed as:

Where: $\frac{{}_5M_x}{{}_5F_x} \times 100$

$_5M_x$ = Number of males enumerated in a specific age group

$_5F_x$ = Number of females enumerated in the same age group

The UN Age-Sex Accuracy Index (UN Joint Score) combines the age ratio score and the sex ratio score to result in a single index measuring the degree of accuracy of the age-sex distribution, i.e.

Age-Sex Accuracy Index (UN Joint Score) = 3*(SRS) + ARSM + ARSF

Where:

SRS is the Sex Ratio Score

ARSM is the Age Ratio Score for Males

ARSF is the Age Ratio Score for Females.

The census data is considered "accurate" if the index is under 20, "inaccurate" if the index is 20 to 40 and "highly inaccurate" if the index is over 40 (United Nations, 1952b). Table 7 provides summary measures of the accuracy of age and sex reporting for PESS 2014 data, computed from the reported age and sex distribution in addition to different smoothing/adjustment methods applied to the reported ages.

The male age ratio score (28.8) was higher than the female age ratio score (19.6), indicating that age misreporting was higher among males compared to females. This is consistent with the Whipple's and Myer's indices computed earlier. The United Nations Joint Score Index was 43.2, indicating that reporting on age and sex data was very rough. This is consistent with Whipple's and Myer's Indices shown earlier.

The male age ratio score (28.8) was higher than the female age ratio score (19.6), indicating that age misreporting was higher among males compared to females.

Table 7: Summary of indices measuring the accuracy of data

Index	Reported	Smoothed				
		Carrier Farrag	K-King Newton	Arriaga	United Nations	Strong
Sex Ratio Score	15.6	11.3	13.3	12.1	10.8	6.5
Male Age Ratio Score	28.8	7.1	7.8	7.3	5.5	2.8
Female Age Ratio Score	19.6	4.5	5.7	5.0	5.4	3.3
Accuracy Index	95.2	45.6	53.5	48.6	43.2	25.7

Note: The accuracy index is the sum of the male and female Age Ratio Scores plus three times the Sex Ratio Score, all calculated using data for ages 10-14 through 65-69.

A comparison of the index with other countries in Table 8 indicates that the data quality for Somalia is second highest and in the same range as Sudan, South Sudan and Uganda.

Table 8: Comparison of UN joint score for somalia and other countries

Country	Year	UN Age Sex Accuracy Index
Somalia *	**2014**	**43.2**
Ethiopia	2007	-
Kenya	2009	21.7
Sudan	2008	45.3
South Sudan	2008	42.6
Tanzania	2012	30.5
Uganda	2002	41.1
Rwanda	2002	27.3

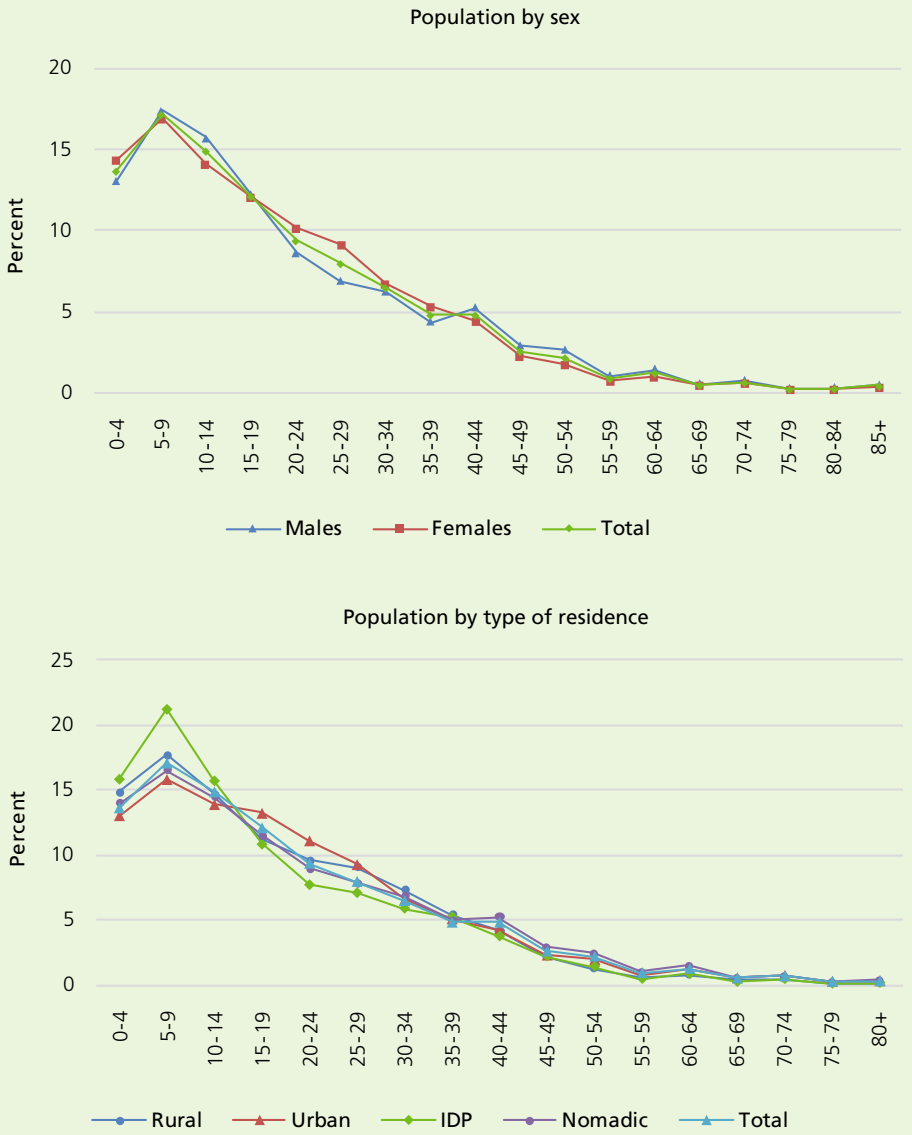
Source: Minesota Population Centre, 2014

* PESS 2014

5.3 Five-year age and sex distribution

The distribution of the total male and female population by five-year age groups; sex; and urban and rural settings, nomadic populations and IDPs is shown in Figure 8. It may be observed that although there is an erratic nature of age distribution as revealed in the single year of age as discussed earlier, this is somewhat minimised when age data is classified into five-year age groups. It is still possible to trace the evidence of under-enumeration (undercount) or age shifting from the five-year age distribution.

Figure 8: Population by age group

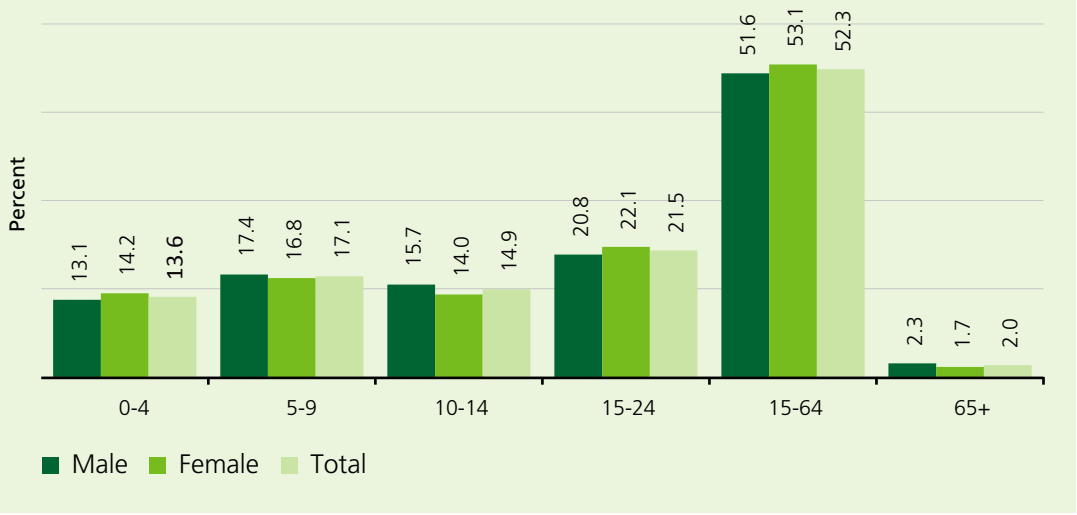


A comparison of population size between males and females shows minimal differences, except for the age group 0-4. This was also seen in the age group 15-64 years.

5.3.1 Population by broad age groups

A comparison of population size between males and females shows minimal differences, except for the age group 0-4 years, which had more females at 14.2 percent than males at 13.1 percent. This was also seen in the age group 15-64 years.

Figure 9: Population by broad age groups and sex



5.3.2 Age ratios

To assess the magnitude of error in age reporting for PESS 2014 data, age ratios were calculated. The age ratio analysis using population data for five-year age groups is useful in detecting age misreporting in populations where fertility has not fluctuated greatly during the past years, and where international migration has not been significant, with the ideal age ratios being closer to 100. The larger the fluctuations of these ratios, or their departure from 100, the greater the probability of errors in the data (United Nations, 1952a). Table 9 shows the age ratios by sex. The overall age ratios fluctuated from 50.1 to 180.8 denoting the presence of age misreporting in the age data. There are high uncertainities in age reporting for males as compared to females since the age ratios fluctuate by large margins.

There are high uncertainities in age reporting for males as compared to females since the age ratios fluctuate by large margins.

Table 9: Age ratios by sex

Age (years)	Male			Female			Both sexes		
	Population	Age ratio	Deviation	Population	Age ratio	Deviation	Population	Age ratio	Deviation
0-4	815,629			864,734			1,680,363		
5-9	1,085,531	120.9	20.9	1,022,833	119.1	19.1	2,108,364	120.0	20.0
10-14	980,123	106.0	6.0	852,642	97.5	-2.5	1,832,765	101.9	1.9
15-19	763,831	100.7	0.7	726,378	98.9	-1.1	1,490,209	99.8	-0.2
20-24	536,505	89.9	-10.1	616,758	96.7	-3.3	1,153,263	93.4	-6.6
25-29	429,989	93.0	-7.0	549,729	107.2	7.2	979,718	100.5	0.5
30-34	388,496	110.6	10.6	408,504	94.1	-5.9	797,000	101.5	1.5
35-39	272,814	76.2	-23.8	318,224	94.7	-5.3	591,038	85.2	-14.8
40-44	327,507	144.5	44.5	263,568	116.2	16.2	591,075	130.3	30.3
45-49	180,461	73.4	-26.6	135,471	73.9	-26.1	315,932	73.6	-26.4
50-54	164,062	133.5	33.5	102,952	114.3	14.3	267,014	125.4	25.4
55-59	65,249	51.3	-48.7	44,681	54.8	-45.2	109,930	52.6	-47.4
60-64	90,511	182.5	82.5	60,167	171.5	71.5	150,678	178.0	78.0
65-69	33,922	49.5	-50.5	25,467	55.1	-44.9	59,389	51.8	-48.2
70-74	46,486	186.6	86.6	32,328	173.1	73.1	78,814	180.8	80.8
75-79	15,892	48.4	-51.6	11,889	52.5	-47.5	27,781	50.1	-49.9
80-84	19,162			12,930			32,092		
85+	28,594			22,876			51,470		
All Ages	6,244,764			6,072,131			12,316,895		

Smoothing of reported data means the elimination or minimisation of irregularities present in data.

5.3.3 Smoothing of age and sex data

Smoothing is an approach used to correct data for age heaping. Smoothing the population includes adjusting of the undercount 0-4 and 5-9 age groups, since the undercount of the two age groups does affect the population count (Arriaga et al, 1994).

In general, the smoothing of reported data means the elimination or minimisation of irregularities present in data. The smoothing techniques used in this report refer to a variety of procedures, ranging from the fitting of simple models to simple averaging. Smoothing techniques involve redistributing the enumerated population numbers in the adjacent ages to produce new results, based on the assumption that these would have been the outcome if distortions had not occurred. The technique used depends on the perceived severity of errors in the age and sex distributions. Since it is necessary to maintain original estimated totals, the techniques of smoothing which either preserved the original totals or minimally altered the totals, were used.

Smoothing of the PESS data was done separately for each sex using the following techniques: Carrier-Farrag, Karup-King Newton, Arriaga, United Nations, and a Strong moving average. All of these methods do not consider adjusting for the youngest age groups, 0-4, 5-9 and the uppermost age groups, except Arriaga smoothing and Strong smoothing methods (also known as Arriaga Strong smoothing). Carrier Farrag adds a quarter of the succeeding and preceding ten-year age band to the age group under consideration. Karup King Newton method also adds a proportion of the shifted numbers into the succeeding and preceding ten-year age band to half of the average of the ten year age band. This is considered as the difference of the subsequent and preceding ten-year age band averaged across 16 intervals.

The United Nations method of smoothing is done by five-year age groups. It subtracts the second intermediate succeeding and preceding five-year age groups from the inflated immediate succeeding five-year age group by a factor of four and that of the five-year age group by a factor of 10. The resulting figure is averaged into all the 16 age intervals. Arriaga's methods incorporate the first and last year age groups in smoothing. The approach used is linear averages. The Strong method (also Arriaga Strong Smoothing) uses an additional formula on the Arriaga smoothing, also known as Arriaga light smoothing, on the ten-year age bands and makes sure the totals add up to the original totals.

The United Nations method of smoothing is done by five-year age groups.

As Arriaga and Associates (1994) noted, differences in results across procedures are very small. The results of the smoothing of males and females, as well as total population, are presented in Tables 10, 11 and 12.

Table 10: Reported and smoothed population by age – males

		Smoothed				
Age (years)	Reported	Carrier Farrag	K-King Newton	Arriaga	United Nations	Strong
Total, 0-79	6,197,008			6,197,008		6,197,008
Total, 10-69	4,233,470	4,233,470	4,233,470	4,233,470	4,257,016	4,233,470
0-4	815,629			944,387		983,208
5-9	1,085,531			956,773		917,952
10-14	980,123	945,543	930,394	943,316	990,409	845,179
15-19	763,831	798,411	813,560	800,638	761,831	734,833
20-24	536,505	541,537	550,912	541,073	548,232	592,076
25-29	429,989	424,957	415,582	425,421	435,203	486,427
30-34	388,496	357,185	359,313	356,150	364,510	382,094
35-39	272,814	304,125	301,997	305,161	311,356	313,269
40-44	327,507	287,415	280,984	283,595	283,476	263,108
45-49	180,461	220,553	226,984	224,373	214,551	210,868
50-54	164,062	134,610	138,626	135,006	137,840	155,468
55-59	65,249	94,701	90,685	94,305	91,025	115,764
60-64	90,511	72,253	72,650	71,758	68,203	78,895
65-69	33,922	52,180	51,783	52,675	50,379	55,489
70-74	46,486			37,162		37,488
75-79	15,892			25,216		24,890
80+	47,756					

Table 11: Reported and smoothed population by age groups – females

		Smoothed				
Age (years)	Reported	Carrier Farrag	K-King Newton	Arriaga	United Nations	Strong
Total, 0-79	6,036,325			6,036,325		6,036,325
Total, 10-69	4,104,541	4,104,541	4,104,541	4,104,541	4,132,108	4,104,541
0-4	864,734			978,019		987,558
5-9	1,022,833			909,548		900,009
10-14	852,642	836,951	834,578	836,744	877,611	811,309
15-19	726,378	742,069	744,443	742,276	723,051	716,851
20-24	616,758	639,642	636,512	637,079	625,679	619,888
25-29	549,729	526,845	529,975	529,408	534,609	521,226
30-34	408,504	411,796	411,330	408,995	417,283	415,537
35-39	318,224	314,932	315,399	317,733	324,083	327,125
40-44	263,568	238,752	235,713	234,124	246,188	237,806
45-49	135,471	160,287	163,326	164,915	153,618	173,610
50-54	102,952	87,844	93,404	89,458	89,150	113,624
55-59	44,681	59,789	54,229	58,175	58,647	78,239
60-64	60,167	49,221	49,281	48,852	46,686	52,694
65-69	25,467	36,413	36,354	36,782	35,505	36,632
70-74	32,328			26,428		25,354
75-79	11,889			17,789		18,863
80+	35,806					

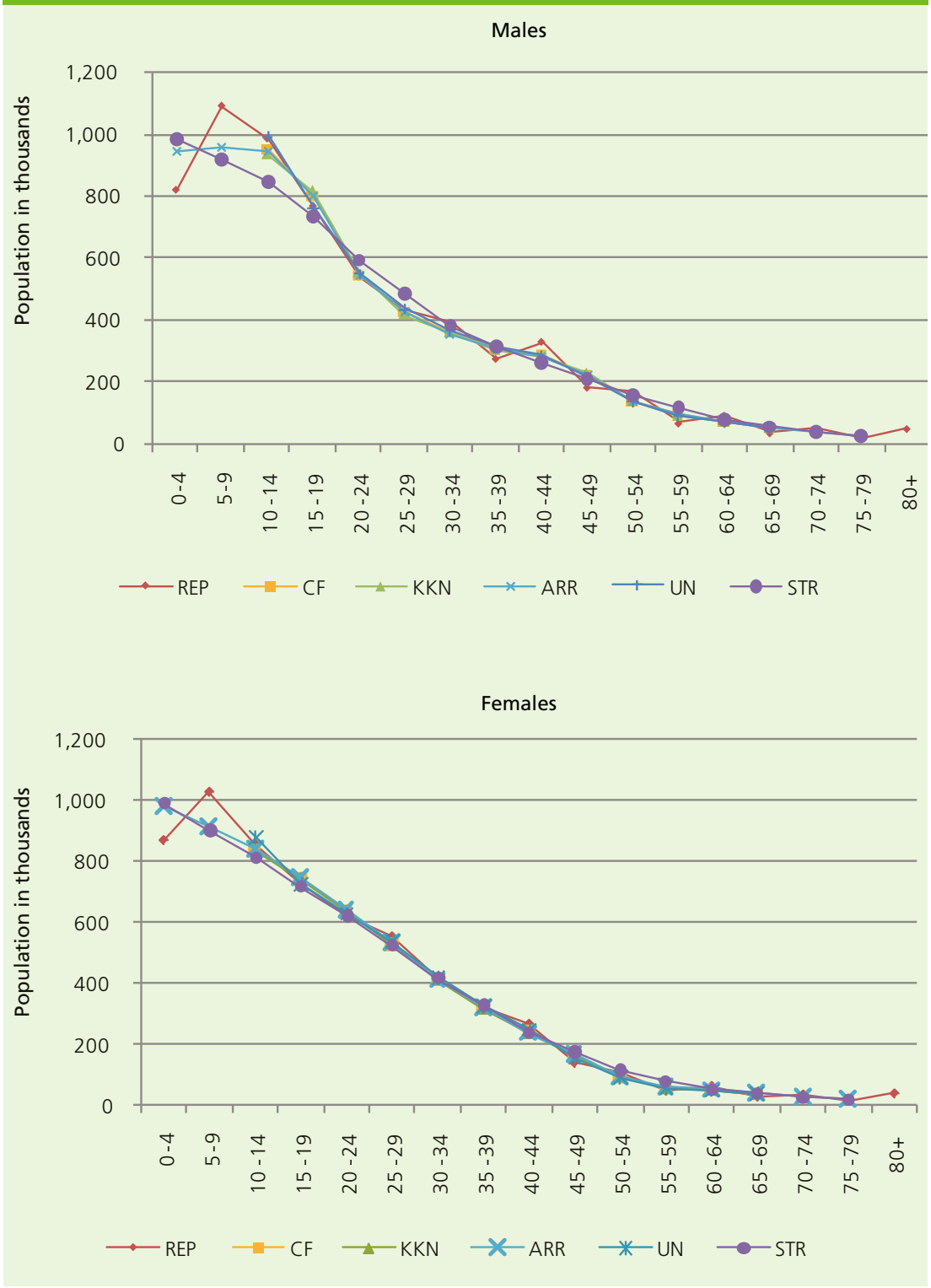
Table 12: Reported and smoothed population by age groups – total

		Smoothed				
Age (years)	Reported	Carrier Farrag	K-King Newton	Arriaga	United Nations	Strong
Total, 0-79	12,233,333			12,233,333		12,233,333
Total, 10-69	8,338,011	8,338,011	8,338,011	8,338,011	8,389,124	8,338,011
0-4	1,680,363			1,922,406		1,970,766
5-9	2,108,364			1,866,321		1,817,961
10-14	1,832,765	1,782,494	1,764,972	1,780,060	1,868,020	1,656,488
15-19	1,490,209	1,540,480	1,558,003	1,542,914	1,484,882	1,451,684
20-24	1,153,263	1,181,179	1,187,424	1,178,152	1,173,911	1,211,964
25-29	979,718	951,802	945,557	954,829	969,812	1,007,653
30-34	797,000	768,981	770,643	765,145	781,793	797,631
35-39	591,038	619,057	617,396	622,894	635,439	640,394
40-44	591,075	526,167	516,697	517,719	529,664	500,914
45-49	315,932	380,840	390,310	389,288	368,169	384,478
50-54	267,014	222,454	232,030	224,464	226,990	269,092
55-59	109,930	154,490	144,914	152,480	149,672	194,003
60-64	150,678	121,474	121,931	120,610	114,889	131,589
65-69	59,389	88,593	88,137	89,457	85,884	92,121
70-74	78,814			63,590		62,842
75-79	27,781			43,005		43,753
80+	83,562					

The reported and smoothed age distributions for males and females – calculated using different procedures – are presented in Figure 10. The figure shows the difference between the reported and smoothed age distributions up to about age 14, indicating that the reported age distributions are relatively inaccurate at younger ages. The variance, however, is more pronounced for males than females.

The reported age distributions are relatively inaccurate at younger ages. The variance, however, is more pronounced for males than females.

Figure 10: Smoothed population data by age group



6.

SAMPLING ERRORS

The quality and reliability of the information gathered by the Population Estimation Survey is discussed in this chapter.

6 SAMPLING ERRORS

6.1 Introduction to sampling errors

Estimates derived from sample surveys are subject to two types of errors: sampling errors and non-sampling errors. Sampling errors are present only in sample surveys and not in census surveys. Non-sampling errors are present in both sample surveys and censuses, and may arise for a number of reasons: the frame may be incomplete, some respondents may not accurately report data, and data may be missing for some respondents. Sampling error, measured in terms of the standard error, occurs in sample surveys since only a portion of the population is enumerated and the sampled units do not have exactly the same characteristics as all of the population units that they represent (Statistics Canada, 2010). Although numerous efforts were made during the implementation of PESS to minimise these types of errors, non-sampling errors are impossible to avoid and difficult to evaluate statistically.

Estimates derived from sample surveys are subject to two types of errors: sampling errors and non-sampling errors. Sampling errors are present only in sample surveys and not in census surveys.

Sampling errors, however, can be evaluated statistically. Sampling errors are usually measured by estimating the extent to which sample estimates, based upon all possible samples of the same size and using the same method of sampling (sample design), differ from one another. The magnitude of the sampling error can be controlled by the sample size (it decreases as the sample size increases), the sample design and the method of estimation. The standard error is used to calculate confidence intervals within which the true value for the population is assumed to fall (Statistics Canada, 2010). Several methods exist to estimate variance for complex sample survey data. The two most commonly used approaches are Taylor Series Linearization and Replication techniques. Software packages that analyse complex sample survey data implement only one of these two methods. For estimators that are smooth functions of the sample data (for example totals, means, proportions, differences between means/proportions), both methods give comparable variance estimates and neither is clearly preferred (United Nations, 2005).

Analysis of complex sample survey data from Burundi using five software packages (Stata, SAS, SUDAAN, WesVar and Epi-Info) gave equivalent variance estimation results using either Taylor Series Linearization (TSL) or Balanced Repeated Replication (BRR) (United Nations, 2005).

This section presents the sampling errors of key variables and how they were calculated. A software package called WesVar was used to calculate the sampling/standard errors, based on the replication method of the Jack-knife Technique (Westat, 2007). A set of replicate weights were computed for each selected replicate so that each one represented the same population of the full sample.

The calculation of standard errors took into account the complexity of the sample design that generated the data, which, in this case, was a cluster-stratified sample design. The sampling error information is presented in three different forms:

i. Absolute value standard errors

ii. Confidence intervals

iii. Relative standard errors (coefficient of variations (CV)).

In this report, standard errors relate to the PESS proportion estimates of several variables. With regard to confidence intervals, the confidence level of 95 percent implies a margin error of five percent. It is worth mentioning that this is a common level used in interpreting the reliability of the results from large-scale household sample surveys. The CV measures the relative amount of variability associated with a sample estimate. Low CV values indicate more reliable estimates. The Census American Community Survey (ACS) Compass products suggest that users should be cautious about using an estimate if the CV is greater than 15 percent; however, the actual cut-off value really depends upon the context of the research at hand. Some users may find that an estimate with a CV value of 35 meets their needs, especially when better quality data are not available, while others may find they need a more precise estimate (ACS, 2010).

Confidence level of 95 percent implies a margin error of five percent. It is worth mentioning that this is a common level used in interpreting the reliability of the results from large-scale household sample surveys.

In the case of the American Community Survey, conducted by the United States Bureau of the Census, the following classification scheme for assessing the reliability of an estimate is used: good (CV < = 15%), fair (15% > CV < =30%), or use with caution (CV > 30%) (ACS, 2010). This assessment scheme is also used to evaluate the reliability of PESS data.

Tables 13, 14 and 15 present the sampling errors for the total population, literacy, and highest level of education attained respectively. Details for these three tables, including five additional variables, are found in the Appendix. The estimates for IDPs were based on a complete count; therefore, there was no need to compute sampling errors.

6.2 Sampling errors for population estimates

An estimated 49.24 percent of the population comprises females, with a standard error of 0.111 and a confidence interval ranging between 49.02 and 49.46. The overall coefficient of variation is relatively low, at 0.23 percent. The estimated proportion of the population that is male is 50.76 percent, with a standard error of 0.113 and a confidence interval ranging between 50.54 and 50.98. The overall coefficient of variation, just like for females, is low, at 0.22 percent. In general, the population proportion estimates are reliable, considering the low levels of dispersion found in standard errors shown in the coefficient of variation. The coefficient of variation values according to types of residence vary from the lowest value in urban areas, at 3.1 percent, and the highest value in the nomadic context, at 6.2 percent. The reliability of population estimates are high since their CVs are below 15 percent.

The population proportion estimates are reliable, considering the low levels of dispersion found in standard errors shown in the coefficient of variation.

Table 13: Sampling errors for the population

Type of residence	Sex	Estimate	Standard error	Lower 95%	Upper 95%	CV (%)
Rural	Both sexes	25.04	1.250	22.59	27.49	4.994
Urban	Both sexes	46.53	1.452	43.69	49.38	3.120
Nomadic	Both sexes	28.43	1.759	24.98	31.88	6.186
Total	Male	50.76	0.113	50.54	50.98	0.223
Total	Female	49.24	0.113	49.02	49.46	0.230
Total	Both sexes	100	0	.	.	0

6.3 Sampling errors for literacy

The reliability of estimates for those able to read and write are high, since they are well below 15 percent, the upper limit of CV for estimates assessed to be good. The CV for literacy of the nomadic population (13.8 percent) is, however, closer to the upper margin of 15 percent than those for rural and urban populations. A comparison of reliability of estimates shows that the estimate for urban is higher than that for rural and nomadic populations.

Table 14: Sampling errors for ability to read and write

Type of residence	Estimate	Standard error	Lower 95%	Upper 95%	CV (%)
Rural	6.90	0.446	6.03	7.78	6.460
Urban	29.65	0.899	27.89	31.42	3.032
Nomadic	3.45	0.475	2.52	4.38	13.769
Total	40.01	0.767	38.51	41.51	1.916
Total	100	0	.	.	0

6.4 Sampling errors for school enrolment

The CVs for most estimates of the current level of school enrolment fall below 15 percent and are assessed to be good. The CVs for some estimates, such as university students from rural areas, fall between 15 and 30 and are assessed to be fairly reliable. The CVs for estimates of college students in nomadic areas is above 30 percent and should be used with caution.

The CVs for most estimates of the current level of school enrolment fall below 15 percent and are assessed to be good.

The CVs for estimates of college students in nomadic areas is above 30 percent and should be used with caution.

Table 15: Sampling errors for school enrolment

Type of residence	Current level of schooling	Estimate	Standard error	Lower 95%	Upper 95%	CV (%)
Rural	Informal school	2.48	0.215	2.06	2.90	8.685
Rural	Koranic school	14.81	0.976	12.90	16.72	6.587
Rural	Primary school	4.97	0.376	4.23	5.71	7.573
Rural	Secondary school	0.93	0.105	0.72	1.13	11.366
Rural	College	0.07	0.011	0.05	0.09	16.655
Rural	University	0.22	0.043	0.14	0.31	19.255
Rural	**Total**	**23.48**	**1.218**	**21.09**	**25.87**	**5.187**
Urban	Informal school	6.17	0.232	5.71	6.62	3.764
Urban	Koranic school	18.48	0.623	17.26	19.70	3.372
Urban	Primary school	24.98	0.652	23.70	26.26	2.612
Urban	Secondary school	9.86	0.273	9.32	10.39	2.769
Urban	College	0.99	0.050	0.90	1.09	5.022
Urban	University	4.21	0.150	3.92	4.51	3.569
Urban	**Total**	**64.69**	**1.442**	**61.87**	**67.52**	**2.229**
Nomadic	Informal school	2.98	0.823	1.37	4.60	27.582
Nomadic	Koranic school	7.88	0.851	6.21	9.55	10.805
Nomadic	Primary school	0.84	0.136	0.58	1.11	16.118
Nomadic	Secondary school	0.09	0.019	0.06	0.13	20.155
Nomadic	College	0.01	0.003	0.00	0.02	33.006
Nomadic	University	0.02	0.005	0.01	0.03	21.101
Nomadic	**Total**	**11.83**	**1.236**	**9.40**	**14.25**	**10.451**
Total	Informal school	11.63	0.788	10.09	13.18	6.775
Total	Koranic school	41.16	0.930	39.34	42.99	2.260
Total	Primary school	30.79	0.661	29.50	32.09	2.146
Total	Secondary school	10.88	0.272	10.34	11.41	2.504
Total	College	1.07	0.050	0.97	1.17	4.690
Total	University	4.46	0.152	4.16	4.76	3.400
Total	**Total**	**100**	**0**	**.**	**.**	**0**

The bulk of the indicators have high reliability with the CVs well below the upper limit of the range assessed to be good.

The findings show high reliability of estimates for indicators whose standard errors were computed. One indicator is recommended to be interpreted with caution. The bulk of the indicators have high reliability with the CVs well below the upper limit of the range assessed to be good according to ACS.

7.

CONCLUSION

This chapter sums up the findings about the methodology and quality of PESS data.

7 CONCLUSION

7.1 Conclusion

Information from about 190,000 households was compiled to produce the results viewed in the series of volumes that follow this report.

The PESS was based on a standardised survey methodology following the internationally agreed criteria. Information from about 190,000 households was compiled to produce the results viewed in the series of volumes that follow this report.

Challenges specific to the country were encountered while developing the sampling frame and gathering survey data. Decades of civil war led to the collapse of governance, including statistical institutions and systems. This was compounded by difficulties faced in enumerating the nomadic population, a challenge in any stable society. It was difficult to access areas experiencing conflict during frame development and enumeration.

The quality of data obtained from a survey depends on the quality of the sampling frame used. A good quality frame should be complete, updated and accurate. The construction of the sampling frame distinguished between urban and rural populations, IDPs and a nomadic stratum. Due to insecurity, it was not possible to visit certain areas to conduct preparatory household listings and interviews. For such areas, the PESS team relied on estimates based on digitised satellite imagery. Given the challenges faced, multiple approaches may be explored over time. To determine accurate population estimates, the units within the stratum should be segmented to standard PSUs. The complete coverage of sampled PSUs is critical for the accurate estimation of the population and socio-economic indicators.

Nomadic populations were covered at water points, using the single-day model. Using this model, one or more interviewers were stationed at each sample water points for one day only (randomly chosen from the 12 days of interviewing). All households drawing water during the designated day of interviewing were enumerated after a screening process. The rural and nomadic strata had relatively low coverage rates. Several rural areas were not covered were due to insecurity. For the nomadic populations, an appropriate methodology should be derived and tested ahead of a survey/census to estimate the population. All these constraints had implications on the estimation of the population of the different strata. Adjustments had to be made to the weights for estimation to factor in differential coverage and non-response.

After the enumeration of the sedentary population, in November and December 2013, the PESS team had to wait for the peak of the drought period, in March 2014, to estimate the nomadic population. This is generally not ideal because the nomadic population cannot always be clearly differentiated from settled inhabitants who visit water points. However, the screening process facilitated the enumeration of nomads.

The assessment of the quality of age and sex data using the Whipple's Index and Myer's Index indicates that the reported single-year age and sex structure is strongly biased towards the terminal digits "0" and "5". Due to the important role played by age and sex data in planning for the population and computing of important demographic indicators (such as fertility and mortality rates), it is vital that in any census or survey, the extent of misreporting is highlighted to facilitate the correct data analysis based on age and sex. Efforts should also be made to determine the causes of misreporting to address issues in age and sex reporting in the future, which in turn would improve the quality of information gathered. Enumeration teams would need to be trained to detect such challenges, and address them to obtain accurate ages from respondents in future surveys and censuses.

Age data was 'smoothed' to eliminate distortions in the age and sex structure and provide an ideal denominator for health, fertility and other socio-economic indicators, as well as a basis for population projections. Based on a comparative assessment of various smoothing methods, the

Arriaga Smoothing Method was used to produce ideal five-year age and sex structure for the Somali population.

Standard errors show that most of PESS findings are reliable, except for a few indicators that may need to be used with caution. The estimates can therefore be used to direct programming, address Somalis' needs, as well as monitor development results.

There is a dire need for additional data sources for the country, which should gather information regularly, and be timely and accessible. Information gathered should be able to verify age and sex data, particularly for under-fives. Policymakers and their partners should also consider collecting information through sources such as Civil Registration and Vital Statistics (CVRS), hospital births' records, religious records, school registers and voters' registers. All these sources would be useful in validating the data collected on the population.

Standard errors show that most of PESS findings are reliable, except for a few indicators that may need to be used with caution. The estimates can therefore be used to direct programming, address Somalis' needs, as well as monitor development results.

REFERENCES

Bekele, S. (2006). Analysis on the Quality of Age and Sex Data Collected in the Two Population and Housing Censuses of Ethiopia. Ethiopia Journal of Science, 29(2), 123–132.

Dahiru, T., & Dikko, H. G. (2013). Digit Preference in Nigerian Censuses Data of 1991 and 2006. Epidemiology, Biostatistics and Public Health, 10(2).

Ewbank, D. (1981). Age Misreporting and Age Selective Under-enumeration: Patterns and Consequences for Demographic Analysis (Report of Committee on Population and Demography). Washington DC: National Academy Press.

FAO-SWALIM. (2012). Hydrogeological Survey and Assessment of Selected Areas in Somaliland and Puntland (Technical Report No. W-20,FAO-SWALIM (GCP/SOM/049/EC)). Nairobi, Kenya: FAO-SWALIM.

Hobbis, F., Siegel, J. S., & Swanson, D. A. (2004). Age and sex composition (2nd ed.). California, USA: Elsvier Academic Press.

Institute for Resource Development. (1990). An Assessment of DHS-I Data Quality. Columbia, Maryland: Institute for Resource Development/Macro System, Inc.

Kpedekpo, G. M. K. (1982). Essentials of Demographic Analysis for Africa. London: Heinemann.

Nagi, M. N., Stockwell, E. G., & Snavely, L. M. (1973). Digit Preference and Avoidance in the Age Statistics of Some African Censuses: Some Patterns and Correlates. Statistical Review, 2, 165–174.

Siegel, J. S., & Swanson, D. A. (2004). The Methods and Materials of Demography, (2nd ed.). 525 B Street, Suite 1900, San Diego, California. USA: Elservier Academic Press.

Spoorenberg, T. (2009). Is the Whipple's index Really a Fair and Reliable Measure of the Quality of Age Reporting?: An Analysis of 234 Censuses from 145 Countries. United Nations: Population Division.

Statistics Canada. (2010). Survey Methods and Practices.

Tanzania National Bureau of Statistics. (2013). Population Distribution by Age and Sex. Dar es Salaam.

UN Department of Economic and Social Affairs. (2008). Principles and Recommendations for Population and Housing Censuses Revision 2.

UNICEF. (2014a). NEZ Multiple Indicator Cluster Survey 2011.

UNICEF. (2014b). NWZ Multiple Indicator Cluster Survey 2011.

United Nations. (1952a). Accuracy Tests for Census Age Distributions Tabulated in Five-year and Ten-year Groups. Population Bulletin, 2(ST/SOA/SER.N/2), 59–79.

United Nations. (1952b). Methods of Appraisal of Quality of Basic Data for Population Estimates (Manual II, Population Studies No. 23). New York: , Series A, No. 23, New York.

United Nations. (1973). Demographic Yearbook 1973, 25th Issue, Special Topic Population Census Statistics III.

United Nations Statistics Division. (2011, November). Evaluation of Age and Sex Distribution. 2010 World Population and Housing Census Programme presented at the Sub-regional Workshop on Census Data Evaluation, Phnom Penh, Cambodia.

United Nations (UN). (2005). Sampling Error Estimation for Survey. In Household Sample Surveys in Developing and Transition Countries.

US Census Bureau. (2014). The Undercount of Young Children.

Westat. (2007). WesVar® 4.3 User's guide (Version 2007). Westat. Retrieved from www.westat.com

APPENDIX A - Assessment of age and sex data in the population

Table A.1: Myer's Blended Method (Myer's Index) computation - males

Terminal digit, a	Population with terminal digit, a		Weights for--		Blended population		Deviation of percentage from 10.00 (6) – 10.00 = (7)
	Starting at age 10+a (1)	Starting at age 20+a (2)	Column 1 (3)	Column 2 (4)	Number (1) x (2) + (3) x (4) =(5)	Percent distribu-tion (6)	
0	1,176,830	875,776	1	9	9,058,812	26.3	16.3
1	257,767	142,811	2	8	1,658,020	4.8	-5.2
2	439,918	221,184	3	7	2,868,040	8.3	-1.7
3	334,953	184,111	4	6	2,444,477	7.1	-2.9
4	343,384	148,848	5	5	2,461,158	7.1	-2.9
5	570,998	404,761	6	4	5,045,031	14.6	4.6
6	321,578	142,001	7	3	2,677,053	7.8	-2.2
7	233,231	117,553	8	2	2,100,956	6.1	-3.9
8	397,113	203,890	9	1	3,777,904	10.9	0.9
9	241,344	132,229	10	0	2,413,437	7.0	-3.0
Total	(x)	(x)	(x)	(x)	34,504,889	100	43.6
Summary index of age preference = Total ÷ 2	(x)	(x)	(x)	(x)	(x)	(x)	21.8

Table A. 2: Myer's Blended Method (Myer's Index) computation- females

Terminal digit, a	Population with terminal digit, a		Weights for--		Blended population		Deviation of percentage from 10.00 (6) – 10.00 = (7)
	Starting at age 10+a (1)	Starting at age 20+a (2)	Column 1 (3)	Column 2 (4)	Number (1) x (2) + (3) x (4) =(5)	Percent distribu-tion (6)	
0	1,036,894	776,655	1	9	8,026,789	23.5	13.5
1	239,643	140,060	2	8	1,599,768	4.7	5.3
2	417,395	228,766	3	7	2,853,546	8.4	1.6
3	324,601	189,855	4	6	2,437,535	7.1	2.9
4	331,315	161,871	5	5	2,465,933	7.2	2.8
5	550,742	415,835	6	4	4,967,791	14.5	4.5
6	302,150	154,904	7	3	2,579,763	7.6	2.4
7	237,979	137,706	8	2	2,179,244	6.4	3.6
8	476,653	241,198	9	1	4,531,078	13.3	3.3
9	252,180	143,683	10	0	2,521,800	7.4	2.6
9	(x)	(x)	(x)	(x)	34,163,247	100	42.6
Summary index of age preference = Total ÷ 2	(x)	(x)	(x)	(x)	(x)	(x)	21.3

APPENDIX B – Sampling errors

Table B. 1: Standard errors for population proportions

Type of residence	Sex	Estimate	Standard error	Lower 95%	Upper 95%	CV (%)
Rural	Male	12.82	0.634	11.57	14.06	4.945
Rural	Female	12.22	0.624	11.00	13.45	5.106
Rural	Total	25.04	1.250	22.59	27.49	4.994
Urban	Male	23.16	0.729	21.73	24.59	3.148
Urban	Female	23.38	0.726	21.95	24.80	3.106
Urban	Total	46.53	1.452	43.69	49.38	3.120
Nomadic	Male	14.79	0.884	13.05	16.52	5.978
Nomadic	Female	13.64	0.881	11.91	15.37	6.461
Nomadic	Total	28.43	1.759	24.98	31.88	6.186
Total	Male	50.76	0.113	50.54	50.98	0.223
Total	Female	49.24	0.113	49.02	49.46	0.230
Total	Total	100	0	.	.	0

Table B. 2: Standard errors for population in age groups

Type of residence	Age group (years)	Estimate	Standard error	Lower 95%	Upper 95%	CV (%)
Rural	0-4	3.45	0.207	3.04	3.85	5.998
Rural	5-9	4.55	0.257	4.05	5.06	5.639
Rural	10-14	3.98	0.204	3.59	4.38	5.110
Rural	15-19	2.93	0.153	2.63	3.23	5.227
Rural	20-24	2.21	0.122	1.97	2.45	5.545
Rural	25-29	1.92	0.114	1.69	2.14	5.928
Rural	30-34	1.69	0.094	1.51	1.88	5.531
Rural	35-39	1.22	0.065	1.09	1.35	5.349
Rural	40-44	1.21	0.065	1.08	1.34	5.404
Rural	45-49	0.63	0.036	0.56	0.70	5.757
Rural	50-54	0.44	0.026	0.39	0.49	5.839
Rural	55-59	0.19	0.015	0.17	0.22	7.745
Rural	60-64	0.24	0.015	0.21	0.27	6.106
Rural	65-69	0.11	0.009	0.10	0.13	7.805
Rural	70-74	0.13	0.009	0.11	0.15	6.644
Rural	75-79	0.05	0.004	0.04	0.06	8.622
Rural	80-84	0.04	0.004	0.04	0.05	9.223
Rural	85+	0.04	0.004	0.03	0.05	9.798
Rural	**Total**	**25.04**	**1.250**	**22.59**	**27.49**	**4.994**
Urban	0-4	6.00	0.204	5.60	6.40	3.393
Urban	5-9	7.47	0.247	6.99	7.96	3.311
Urban	10-14	6.82	0.229	6.37	7.27	3.364
Urban	15-19	6.10	0.203	5.70	6.50	3.331
Urban	20-24	4.76	0.159	4.45	5.07	3.334

Contd...

Table B. 2: Standard errors for population in age groups (Contd...)

Type of residence	Age group (years)	Estimate	Standard error	Lower 95%	Upper 95%	CV (%)
Urban	25-29	3.89	0.128	3.64	4.15	3.294
Urban	30-34	2.91	0.096	2.72	3.10	3.308
Urban	35-39	2.16	0.073	2.02	2.31	3.362
Urban	40-44	2.14	0.074	2.00	2.29	3.456
Urban	45-49	1.16	0.039	1.08	1.24	3.361
Urban	50-54	1.08	0.037	1.01	1.15	3.404
Urban	55-59	0.44	0.016	0.41	0.47	3.706
Urban	60-64	0.61	0.021	0.56	0.65	3.535
Urban	65-69	0.25	0.011	0.23	0.28	4.266
Urban	70-74	0.35	0.014	0.32	0.38	3.895
Urban	75-79	0.12	0.006	0.11	0.14	5.259
Urban	80-84	0.14	0.007	0.12	0.15	4.791
Urban	85+	0.12	0.008	0.10	0.13	7.140
Urban	**Total**	**46.53**	**1.452**	**43.69**	**49.38**	**3.120**
Nomadic	0-4	4.06	0.364	3.34	4.77	8.968
Nomadic	5-9	4.64	0.336	3.98	5.30	7.259
Nomadic	10-14	4.06	0.242	3.59	4.54	5.963
Nomadic	15-19	3.23	0.187	2.86	3.60	5.797
Nomadic	20-24	2.60	0.186	2.24	2.96	7.148
Nomadic	25-29	2.35	0.175	2.01	2.70	7.440
Nomadic	30-34	1.96	0.125	1.72	2.21	6.345
Nomadic	35-39	1.45	0.079	1.29	1.60	5.471
Nomadic	40-44	1.47	0.099	1.28	1.67	6.709
Nomadic	45-49	0.80	0.053	0.69	0.90	6.687
Nomadic	50-54	0.66	0.045	0.57	0.75	6.779
Nomadic	55-59	0.27	0.032	0.21	0.34	11.723
Nomadic	60-64	0.37	0.030	0.32	0.43	7.896
Nomadic	65-69	0.13	0.012	0.11	0.15	9.026
Nomadic	70-74	0.17	0.016	0.14	0.21	9.497
Nomadic	75-79	0.06	0.010	0.04	0.08	16.033
Nomadic	80-84	0.08	0.012	0.06	0.11	15.291
Nomadic	85+	0.05	0.010	0.04	0.07	17.460
Nomadic	**Total**	**28.43**	**1.759**	**24.98**	**31.88**	**6.186**
Total	0-4	13.50	0.224	13.06	13.94	1.661
Total	5-9	16.66	0.159	16.35	16.97	0.955
Total	10-14	14.87	0.143	14.59	15.15	0.965
Total	15-19	12.26	0.158	11.95	12.57	1.285
Total	20-24	9.57	0.104	9.36	9.77	1.082
Total	25-29	8.16	0.093	7.98	8.35	1.139
Total	30-34	6.56	0.067	6.43	6.70	1.019
Total	35-39	4.83	0.058	4.72	4.95	1.199
Total	40-44	4.83	0.077	4.68	4.98	1.600
Total	45-49	2.59	0.041	2.50	2.67	1.589
Total	50-54	2.18	0.034	2.11	2.24	1.561

Contd...

Table B. 2: Standard errors for population in age groups (Contd...)

Type of residence	Age group (years)	Estimate	Standard error	Lower 95%	Upper 95%	CV (%)
Total	55-59	0.91	0.032	0.84	0.97	3.586
Total	60-64	1.22	0.025	1.17	1.27	2.008
Total	65-69	0.50	0.015	0.47	0.53	3.023
Total	70-74	0.65	0.018	0.62	0.69	2.825
Total	75-79	0.23	0.010	0.21	0.25	4.214
Total	80-84	0.26	0.013	0.24	0.29	5.005
Total	85+	0.21	0.012	0.19	0.24	5.396
Total	**Total**	**100**	**0**	**.**	**.**	**0**

Table B. 3: Standard errors for marital status

Type of residence	Marital status	Estimate	Standard error	Lower 95%	Upper 95%	CV (%)
Rural	Never married	6.84	0.384	6.09	7.60	5.607
Rural	Married	14.60	0.768	13.10	16.11	5.260
Rural	Abandoned	0.76	0.056	0.65	0.87	7.336
Rural	Divorced	0.82	0.051	0.72	0.92	6.299
Rural	Widowed	0.84	0.052	0.73	0.94	6.251
Rural	**Total**	**23.86**	**1.186**	**21.53**	**26.18**	**4.971**
Urban	Never married	18.20	0.612	17.00	19.40	3.362
Urban	Married	25.43	0.773	23.92	26.95	3.038
Urban	Abandoned	1.03	0.045	0.95	1.12	4.336
Urban	Divorced	1.40	0.056	1.28	1.51	4.040
Urban	Widowed	1.48	0.060	1.37	1.60	4.032
Urban	**Total**	**47.55**	**1.447**	**44.71**	**50.39**	**3.043**
Nomadic	Never married	8.97	0.545	7.90	10.04	6.081
Nomadic	Married	17.78	1.197	15.43	20.13	6.734
Nomadic	Abandoned	0.55	0.067	0.42	0.68	12.213
Nomadic	Divorced	0.63	0.085	0.47	0.80	13.383
Nomadic	Widowed	0.66	0.085	0.50	0.83	12.825
Nomadic	**Total**	**28.59**	**1.704**	**25.25**	**31.93**	**5.960**
Total	Never married	34.02	0.415	33.20	34.83	1.219
Total	Married	57.82	0.502	56.83	58.80	0.868
Total	Abandoned	2.34	0.080	2.18	2.49	3.426
Total	Divorced	2.85	0.093	2.67	3.03	3.251
Total	Widowed	2.98	0.084	2.82	3.15	2.802
Total	**Total**	**100**	**0**	**.**	**.**	**0**

Table B. 4: Standard errors for ability to read and write

Type of residence	Literacy	Estimate	Standard error	Lower 95%	Upper 95%	CV (%)
Rural	Cannot read and write	15.53	0.824	13.92	17.15	5.302
Rural	Can read only	2.52	0.176	2.17	2.86	7.008
Rural	Can read and write	6.90	0.446	6.03	7.78	6.460
Rural	**Total**	**24.95**	**1.218**	**22.56**	**27.34**	**4.883**
Urban	Cannot read and write	13.17	0.503	12.18	14.15	3.819
Urban	Can read only	4.05	0.180	3.69	4.40	4.448
Urban	Can read and write	29.65	0.899	27.89	31.42	3.032
Urban	**Total**	**46.87**	**1.428**	**44.07**	**49.67**	**3.046**
Nomadic	Cannot read and write	21.59	1.257	19.12	24.05	5.824
Nomadic	Can read only	3.14	0.270	2.61	3.67	8.603
Nomadic	Can read and write	3.45	0.475	2.52	4.38	13.769
Nomadic	**Total**	**28.18**	**1.697**	**24.85**	**31.51**	**6.022**
Total	Cannot read and write	50.29	0.758	48.80	51.77	1.508
Total	Can read only	9.71	0.263	9.19	10.22	2.712
Total	Can read and write	40.01	0.767	38.51	41.51	1.916
Total	**Total**	**100**	**0**	**.**	**.**	**0**

Table B. 5: Standard errors for enrolment

Type of residence	Enrolment	Estimate	Standard error	Lower 95%	Upper 95%	CV (%)
Rural	Enrolled	8.54	0.468	7.62	9.45	5.478
Rural	Not enrolled	16.45	0.852	14.78	18.12	5.176
Rural	**Total**	**24.99**	**1.226**	**22.58**	**27.39**	**4.907**
Urban	Enrolled	24.75	0.770	23.24	26.26	3.112
Urban	Not enrolled	22.14	0.735	20.70	23.58	3.320
Urban	**Total**	**46.89**	**1.430**	**44.09**	**49.69**	**3.049**
Nomadic	Enrolled	4.42	0.418	3.60	5.24	9.455
Nomadic	Not enrolled	23.70	1.453	20.85	26.55	6.128
Nomadic	**Total**	**28.12**	**1.699**	**24.79**	**31.45**	**6.041**
Total	Enrolled	37.70	0.673	36.39	39.02	1.785
Total	Not enrolled	62.30	0.673	60.98	63.61	1.080
Total	**Total**	**100**	**0**	**.**	**.**	**0**

Table B. 6: Standard errors for highest level of education attained

Type of residence	Highest level of education attained	Estimate	Standard error	Lower 95%	Upper 95%	CV (%)
Rural	None	19.85	1.006	17.87	21.82	5.07
Rural	Primary	2.71	0.204	2.31	3.11	7.521
Rural	Secondary	0.77	0.075	0.62	0.91	9.762
Rural	College	0.10	0.015	0.07	0.13	14.203
Rural	University	0.14	0.024	0.09	0.19	16.964
Rural	Unknown	0.54	0.070	0.40	0.68	12.875
Rural	**Total**	**24.11**	**1.203**	**21.75**	**26.47**	**4.990**
Urban	None	24.75	0.852	23.08	26.42	3.442
Urban	Primary	11.57	0.385	10.81	12.32	3.327
Urban	Secondary	6.18	0.202	5.78	6.58	3.263
Urban	College	1.00	0.045	0.91	1.08	4.550
Urban	University	2.25	0.086	2.08	2.42	3.834
Urban	Unknown	1.30	0.060	1.18	1.42	4.615
Urban	**Total**	**47.04**	**1.418**	**44.26**	**49.82**	**3.015**
Nomadic	None	26.94	1.573	23.85	30.02	5.840
Nomadic	Primary	1.04	0.118	0.81	1.27	11.284
Nomadic	Secondary	0.10	0.015	0.07	0.13	14.914
Nomadic	College	0.04	0.006	0.02	0.05	16.455
Nomadic	University	0.03	0.007	0.01	0.04	29.650
Nomadic	Unknown	0.71	0.101	0.52	0.91	14.182
Nomadic	**Total**	**28.86**	**1.661**	**25.60**	**32.11**	**5.755**
Total	None	71.53	0.661	70.23	72.83	0.924
Total	Primary	15.31	0.364	14.60	16.03	2.376
Total	Secondary	7.05	0.208	6.64	7.46	2.952
Total	College	1.13	0.046	1.04	1.23	4.083
Total	University	2.41	0.088	2.24	2.59	3.656
Total	Unknown	2.56	0.131	2.30	2.81	5.130
Total	**Total**	**100**	**0**	**.**	**.**	**0**

Table B. 7: Standard errors for usual activity with the past twelve months

Type of residence	Activity status	Estimate	Standard error	Lower 95%	Upper 95%	CV (%)
Rural	Working (including housewives having activities)	8.48	0.470	7.56	9.4	5.545
Rural	Not working and looking for work	3.3	0.235	2.84	3.76	7.133
Rural	Housewife	6.74	0.374	6	7.47	5.548
Rural	Student	5.25	0.315	4.63	5.86	6.008
Rural	Retired	0.49	0.032	0.43	0.55	6.472
Rural	Disabled	0.21	0.018	0.17	0.24	8.554
Rural	Other not working	0.5	0.063	0.38	0.62	12.552
Rural	**Total**	**24.96**	**1.202**	**22.6**	**27.32**	**4.817**
Urban	Working (including housewives having activities)	10.13	0.334	9.47	10.78	3.298
Urban	Not working and looking for work	5.53	0.202	5.13	5.92	3.662
Urban	Housewife	12.49	0.403	11.7	13.28	3.229
Urban	Student	17.27	0.545	16.2	18.33	3.156
Urban	Retired	1.09	0.044	1.01	1.18	4.01
Urban	Disabled	0.41	0.024	0.36	0.45	5.785
Urban	Other not working	0.51	0.044	0.42	0.59	8.654
Urban	**Total**	**47.41**	**1.417**	**44.63**	**50.19**	**2.989**
Nomadic	Working (including housewives having activities)	15.63	1.120	13.44	17.83	7.164
Nomadic	Not working and looking for work	1.99	0.183	1.63	2.35	9.217
Nomadic	Housewife	6.17	0.508	5.17	7.16	8.235
Nomadic	Student	1.37	0.152	1.07	1.67	11.094
Nomadic	Retired	0.51	0.059	0.39	0.62	11.559
Nomadic	Disabled	0.27	0.033	0.2	0.33	12.195
Nomadic	Other not working	1.7	0.306	1.1	2.3	17.993
Nomadic	**Total**	**27.63**	**1.653**	**24.39**	**30.87**	**5.981**
Total	Working (including housewives having activities)	34.24	0.819	32.63	35.85	2.392
Total	Not working and looking for work	10.81	0.287	10.25	11.38	2.657
Total	Housewife	25.39	0.422	24.56	26.22	1.663
Total	Student	23.88	0.533	22.84	24.93	2.233
Total	Retired	2.09	0.057	1.98	2.2	2.711
Total	Disabled	0.88	0.037	0.81	0.95	4.201
Total	Other not working	2.7	0.307	2.1	3.31	11.335
Total	**Total**	**100**	**0**	**.**	**.**	**0**

APPENDIX C – GLOSSARY

Age: Number of years one had lived as at last birthday i.e. in reference to the survey date.

Age Dependency Ratio: The ratio of people in the "dependent" ages (those under age 15 and age 65 and older) to those in the "working age population" (15-64 years).

Age Ratio Score: The mean deviation of the age ratios for males or females from 100 percent, irrespective of sign.

Age Sex Accuracy Index (UN Joint Score): Defined as "ASAI = 3 (Sex Ratio Score) + Age-Ratio Score for Male + Age-Ratio Score for Females".

Bachi Index: used to measure digit preference at ages ending in zero to none. The range of Bachi's Index is 0-90. An index of zero means no digit preference while an index of 90 means a preference for a single digit.

Baseline: A clearly defined starting point (point of departure) from where implementation begins, improvement is judged, or comparison is made.

Base-weight: The inverse of the probability of selection.

Berkad: A man-made cistern sunk into the ground to store run-off water.

Borehole: A drilled hole in the ground to extract underground water.

Brain drain: the emigration of highly skilled individuals, where a highly skilled emigrant is a foreign-born individual, aged 25 or more, with an academic or professional degree beyond high school (i.e. 'post-secondary' or 'tertiary educated') at the census or the survey date.

Census: A complete enumeration of the population within a defined territory such as a country or a well-delineated part of a country with the purpose of estimating the households and population together with associated variables including age, sex, marriage, education, mobility etc.

Cluster: A (usually geographically defined) group of individuals.

Cohort: A group of people with a common statistical characteristic.

Data: Information in raw or unorganized form (such as alphabets, numbers, or symbols) that refer to, or represent, conditions, ideas, or objects.

Design weights: Computed as the inverse of the inclusion probabilities and then scaled such that their sum equals the net sample size.

Domain hierarchy: A system of nested areas units designed for administrative or data collection purposes.

Dug well: A hand-excavated well.

Dwelling unit: A place of abode (residence), a room or a group of rooms with a private entrance normally intended as a residence for one household (for example, a single house, an apartment, a group of rooms in a house). A dwelling unit can also have more than one household.

Editing: The application of checks to identify missing, invalid or inconsistent data entries that point to records that are potentially in error.

Enumeration area: A designated area with 50 to 149 households, with an average of around 100 households.

Enumerator: The person responsible for collecting information from the sampled household.

Geographical Information System (GIS): A system designed to capture, store, manipulate, analyse, manage and present all types of geographical data.

High-resolution Satellite imagery: An imagery collected by a satellite instrument with a ground resolution of less than one metre.

Hot-deck: A method for handling missing data in which each missing value is replaced with an observed response from a "similar" unit.

Household: A group of persons who normally live together, take their meals from a common kitchen and report to a common household head.

Jack-knife technique: A sampling technique that allows subgroups/replicates to overlap.

Myer's Index: Calculated for the age above ten years and shows the excess or deficit of people in ages in any of the ten digits expressed as percentage. It is based on the assumption that, the population is equally distributed among all ages. The theoretical range of Myer's Index is from 0 to 90, where "0" indicates no preference, while "90" indicates absolute preference.

Nomadic population: Part of the population that normally stays in a temporary nomadic settlement, or that has not got a permanent place of residence anywhere.

Non-response: cases where data for a sample observation unit is missing, because households refused to respond or could not be reached.

Population distribution: The spread of surveyed people with respect to a particular characteristic e.g. age.

Population pyramid: A graphical illustration that displays a population's age and sex composition. Horizontal bars present the numbers or proportions of males and females in each age group.

Primary sampling unit: The first-stage area cluster included in a sampling frame.

Sample: A group of people or things that is chosen out of a larger number and is asked questions or tested in order to get information about the larger group.

Sampling frame: A collection of all relevant units e.g. settlements from which a sample is selected.

Satellite image: A picture of the earth taken from an earth-orbital satellite.

Sedentary: Staying or living in one place, instead of moving to different places.

Segmentation: The process of dividing a primary sampling unit into several area segments according to a measure of size

Sex ratio: The ratio of males to females in a given population usually expressed as the number of males for every 100 females.

Sex Ratio Score (SRS): The mean difference between sex ratios for the successive age groups, averaged irrespective of sign.

Strata: A collection of seemingly similar/homogeneous units.

Stratification: A system of dividing an area into homogeneous units.

Structure: A building used for purposes of residential, business or any other activity.

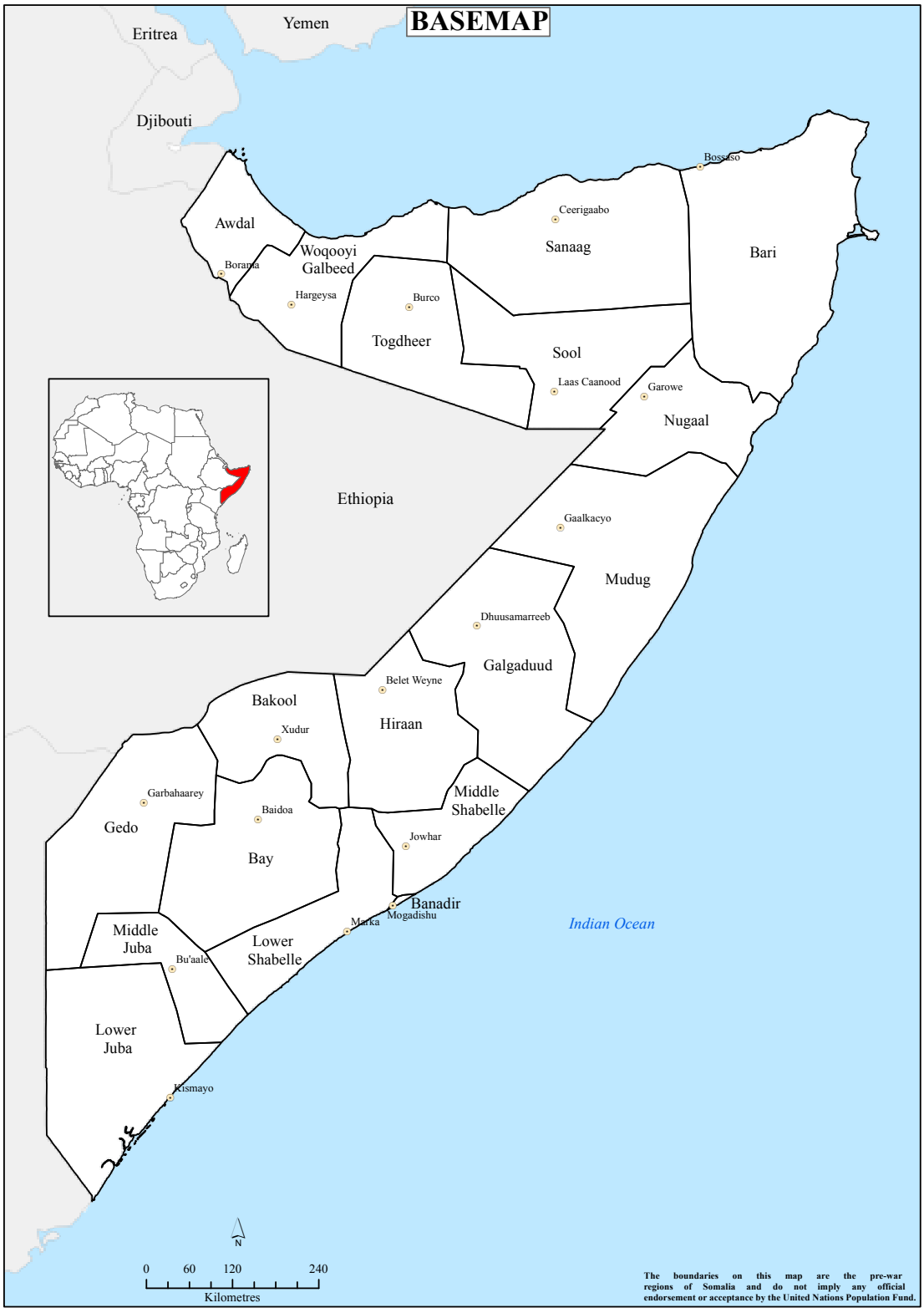
Substrata: A subdivision within a strata

Syntax: The set of rules that define the combinations of symbols that are considered to be correctly structured programs in a programming language

WESVAR: A software programme used to compute estimates and variance estimates from potentially complicated survey data

Whipple's Index: A summary measure of age heaping on ages ending in "0" or "5" used to determine variability in the quality of age reporting.

APPENDIX D – Basemap



APPENDIX E – Questionnaire

Serial Number.....

Population Estimation Survey 2013 (PESS)
Household Questionnaire

Sheet		of	

The information in this questionnaire is strictly confidential and will only be used for statistical purposes

A	Identification Information				
A1	Region	|___|___|	A6	Settlement/Water Point name: _____	|__|__|__|
A2	District	|___|___|	*Enumerator: For nomadic households → go to B1*		
A3	Enumeration Area Type	|___|	A7	Housing Structure Serial No. in the EA	|__|__|__|
A4	Enumeration Area Code	|___|___|___|	A8	Number of dwelling units in the structure	|__|__|
A5	HH Serial No. in the EA	|___|___|___|	A9	Number of households in dwelling unit	|__|__|

AFTER COMPLETION OF INTERVIEW								
PESS role	**Name**	**ID Code**	**Date (DD/MM**	**Signature**	**Interview Status**		**Summary** *(From question B4)*	
					1 Completed	|___|		
Enumerator:		|__|__|__|	|__|__|__|__|		2 Partially Completed		Males	|___|___|
Editor:		| _|__|__|	|__|__|__|__|		3 Refused		Females	|___|___|
Supervisor:		| _|__|__|	|__|__|__|__|		4 Not able to visit		Total	|___|___|

B Individual Characteristics

List of Household Members			Demographic Characteristics				Educational Characteristics (6+years old)				Labour Force (10+years old)		Migration		Voter Registration (15+ years)
SERIAL NUMBER (Circle Respondent's Ser No)	"Please tell me the full name of each person who usually lives here, starting with the head of the household" when all are listed probe: "Are there any others who live here, even if they are not at home now?" Arrange the members as follows - Head of the HH - Spouse - Unmarried children from old to young -Married Children with their families -Relatives -Non-relatives	What is the relation of <name> to the head of household? 1=Head of HH 2=Spouse 3=Son/daughter 4=Brother/ sister 5= Parent or grandparent 6=Grandson/ granddaughter 7=Other relatives 8 = Unrelated	Is <name> male or female? 1= Male 2=Female	How old is <name> in comple-ted years? *If less than one year, write "00"; if 95 years or older, write '95'*	What is the year of birth <name>? *If respondent does not know age or date of birth, please probe by using calendar of events given to you or by relating age to ages of relatives* |Y|Y|Y|Y|	*(for 15 years old)* What is <name's> marital status? 1=Never Married 2=Married 3=Separated 4=Divorced 5=Widowed	Can <name> read & write? 1=Cannot read & write 2= Can read only 3= Can read and write	Is <name> curren-tly enrol-led in school? 1= Yes; 2=No →B11	What is the level of school <name> is enrolled in? 1=Informal School 2= Koranic School 3= Primary School 4=Secondary School 5= College 6=University	What is the highest level of formal education completed by *(name)*? 1=None 2=Primary 3=Secondary 4=College 5=University 6= Unknown	What has <name> mostly been doing in the last 12 months? 1= Working (including housewives having activities 2= Not working, but looking for work → B14 3= Housewife not working → B14 4= Student → B14 5= Retired → B14 6= Disabled → B14 7= Other not working → B14	*If working in B12* What is <name's> occupation? *(Write the name of the occupation and the appropriate code from the list of occupations given to you)*	Did <name> come to live in the country after December 2012? 1-Yes 2=No → B16	*If yes in C14* Which country did <name> come from? *(Write the name of the country and its code from list of countries given to you)*	*Currently applicable for Somaliland only* "Was <name> registered as a voter for the last election?" 1=Yes; 2=No
B1	B2	B3	B4	B5	B6	B7	B8	B9	B10	B11	B12	B13	B14	B15	B16
01		|____|	|____|	|____|	|____|____|____|____|	|____|	|____|	|____|	|____|	|____|	|____|	|____|	|____|____|:|____|____|	|____|	|____|____|:|____|____|	|____|
02		|____|	|____|	|____|	|____|____|____|____|	|____|	|____|	|____|	|____|	|____|	|____|	|____|	|____|____|:|____|____|	|____|	|____|____|:|____|____|	|____|
03		|____|	|____|	|____|	|____|____|____|____|	|____|	|____|	|____|	|____|	|____|	|____|	|____|____|:|____|____|	|____|	|____|____|:|____|____|	|____|
04		|____|	|____|	|____|	|____|____|____|____|	|____|	|____|	|____|	|____|	|____|	|____|	|____|____|:|____|____|	|____|	|____|____|:|____|____|	|____|
05		|____|	|____|	|____|	|____|____|____|____|	|____|	|____|	|____|	|____|	|____|	|____|	|____|____|:|____|____|	|____|	|____|____|:|____|____|	|____|
06		|____|	|____|	|____|	|____|____|____|____|	|____|	|____|	|____|	|____|	|____|	|____|	|____|____|:|____|____|	|____|	|____|____|:|____|____|	|____|
07		|____|	|____|	|____|	|____|____|____|____|	|____|	|____|	|____|	|____|	|____|	|____|	|____|____|:|____|____|	|____|	|____|____|:|____|____|	|____|
08		|____|	|____|	|____|	|____|____|____|____|	|____|	|____|	|____|	|____|	|____|	|____|	|____|____|:|____|____|	|____|	|____|____|:|____|____|	|____|
09		|____|	|____|	|____|	|____|____|____|____|	|____|	|____|	|____|	|____|	|____|	|____|	|____|____|:|____|____|	|____|	|____|____|:|____|____|	|____|
10		|____|	|____|	|____|	|____|____|____|____|	|____|	|____|	|____|	|____|	|____|	|____|	|____|____|:|____|____|	|____|	|____|____|:|____|____|	|____|
11		|____|	|____|	|____|	|____|____|____|____|	|____|	|____|	|____|	|____|	|____|	|____|	|____|____|:|____|____|	|____|	|____|____|:|____|____|	|____|
12		|____|	|____|	|____|	|____|____|____|____|	|____|	|____|	|____|	|____|	|____|	|____|	|____|____|:|____|____|	|____|	|____|____|:|____|____|	|____|
13		|____|	|____|	|____|	|____|____|____|____|	|____|	|____|	|____|	|____|	|____|	|____|	|____|____|:|____|____|	|____|	|____|____|:|____|____|	|____|
14		|____|	|____|	|____|	|____|____|____|____|	|____|	|____|	|____|	|____|	|____|	|____|	|____|____|:|____|____|	|____|	|____|____|:|____|____|	|____|
15		|____|	|____|	|____|	|____|____|____|____|	|____|	|____|	|____|	|____|	|____|	|____|	|____|____|:|____|____|	|____|	|____|____|:|____|____|	|____|
16		|____|	|____|	|____|	|____|____|____|____|	|____|	|____|	|____|	|____|	|____|	|____|	|____|____|:|____|____|	|____|	|____|____|:|____|____|	|____|
17		|____|	|____|	|____|	|____|____|____|____|	|____|	|____|	|____|	|____|	|____|	|____|	|____|____|:|____|____|	|____|	|____|____|:|____|____|	|____|
18		|____|	|____|	|____|	|____|____|____|____|	|____|	|____|	|____|	|____|	|____|	|____|	|____|____|:|____|____|	|____|	|____|____|:|____|____|	|____|
19		|____|	|____|	|____|	|____|____|____|____|	|____|	|____|	|____|	|____|	|____|	|____|	|____|____|:|____|____|	|____|	|____|____|:|____|____|	|____|
20		|____|	|____|	|____|	|____|____|____|____|	|____|	|____|	|____|	|____|	|____|	|____|	|____|____|:|____|____|	|____|	|____|____|:|____|____|	|____|

C. Population Mobility	
ENUMERATOR: If household in urban/rural/IDP area, ask questions C1 to C5; if household is at water point, skip C1-C5, ask C6-C8	
C1 For how many years has this household lived in this place?	
(If less than one year, write "00") *(If always lived here, write "95") ==> go to D1*	|____|____|
C2 What is the name of the district you lived in before you moved to this place?	
If mentioned district is not in list, write |0|0|0|0| *If household came from abroad, write name of country and code*	
Name of district:	|____|____|____|____|
OR Name of country :	|____|____|
C3 What was the type of your previous place of residence?	
1- Rural 2- Urban 3 IDP Settlement 4 Nomadic area 5- Refugee camp	|___|
C4 What was the main reason why this household left its place of origin?	
1- Insecurity 2- Drought 3- Floods 4- Total loss of livelihood (destitution) 5- Better economic opportunities elsewhere → **go to D1** 6- Lack of access to services → **go to D1** 7- Other (*specify*): _____	|___|
C5 What are your plans regarding the place of origin for the coming year?	
1- Stay here in this current location → **go to D1** 2- Return to place of origin → **go to D1** 3- Temporarily return to place of origin → **go to D1** 4- Relocate to another place permanently → **go to D1** 5- Have no plan → **go to D1**	|____|
Question C6 to C8 for nomadic households only	
C6 Where do you usually stay during the rainy season? *(Specify name and code of district or country below)*	
1. Same district as district in which waterpoint is located 2. Different district within the same region: 3- District in different region: _____ 4. Neighbouring country: _____	|___|
Enter code of district from the district coding list provided	|____|____|____|____|
Enter code of country from the country coding list provided	|____|____|
C7 How many months in a normal year do you stay in <*the place mentioned in C6***>?**	|____|
C8: How many times during the last 12 days (including today) has your household been watering any kind of your livestock at this waterpoint?	|____|____|

D. Demographic Events

D1 How many previous members of this household are currently living abroad ? *Record total number of males and females in the boxes on the right.* *If no female or male members living abroad write | 0 | 0 |* *Specify numbers for males and females by country in the male and female colums;* *Record code of the country on the left-hand side in the row*	Males	Females
|___|___| **Country 1:** _____	|___|___|	|___|___|
|___|___| **Country 2:** _____	|___|___|	|___|___|
|___|___| **Country 3:** _____	|___|___|	|___|___|
|___|___| **Country 4:** _____	|___|___|	|___|___|
D2 How many of the above members left the country to live abroad after December 2012?	|___|___|	|___|___|
D3 Did this household have any live birth during the last 24 month (including those who died)? **1 =yes, 2 = No → Go to D10**	|___|	

Ser no	Sex of live birth 1= Male 2= Female	Age of Mother at the time of birth	Is the child still alive? 1= Yes 2 = No	*If mother is member of the household* **Line number mother** *(from B1)*	*If child is member of the household* **Line number child** *(from B1)*
D4	**D5**	**D6**	**D7**	**D8**	**D9**
1	|___|	|___|___|	|___|	|___|___|	|___|___|
2	|___|	|___|___|	|___|	|___|___|	|___|___|
3	|___|	|___|___|	|___|	|___|___|	|___|___|
4	|___|	|___|___|	|___|	|___|___|	|___|___|
5	|___|	|___|___|	|___|	|___|___|	|___|___|

D10 Did this household have any death among its members during the last 24 month? 1 = Yes, 2 = No *(If household is sedentary → go to E1; if household is nomadic (at water point), → End Interview*	|___|

Ser no	Sex of deceased person 1= Male 2= Female	Age at death in years *(Write "00" if less than one year)*	***Enumerator skipping instructions (starting after D13)*:** 1. If the deceased is a male → go to next line 2. If the deceased is a female not aged 15-49 → go to next line 3. If the deceased is a female aged 15 -49 → continue with D14. 4. If all deaths are recorded and household is sedentary → go to E1 5. If all deaths are recorded and household is nomadic → end interview			
			Was she ever married? 1= Yes → go to D15 2 = No → next line	**Was she pregnant when she died?** 1 = Yes → next line 2 = No → go to D16	**Did she die during delivery** 1 = Yes → next line 2 = No → go to D17	**Did she die during the 2 months following delivery?** 1 =Yes → nextline 2 = No → nextline
D11	**D12**	**D13**	**D14**	**D15**	**D16**	**D17**
1	|___|	|___|___|	|___|	|___|	|___|	|___|
2	|___|	|___|___|	|___|	|___|	|___|	|___|
3	|___|	|___|___|	|___|	|___|	|___|	|___|
4	|___|	|___|___|	|___|	|___|	|___|	|___|
5	|___|	|___|___|	|___|	|___|	|___|	|___|

E Housing Characteristics and Household Assets

E1 What is the main source of energy for lighting?

1-Electricity 2- Solar Energy 3- Kerosene 4- Firewood 5- Torch 6- Other *(specify)* ……	|___|

E2 What is the main source of energy for cooking?

1- Electricity 2- Liquefied Petroleum Gas 3- Kerosene 4- Charcoal 5- Firewood 6- Agricultural crop residue 7- Livestock dung 8- Other *(specify)* ……	|___|

E3 What is the main source of drinking water?

1- Piped water/tap inside dwelling 2- Piped water outside -public tap; 3- Mineral water 4- Borehole/dug well 5- River, dam, stream etc. (Open water) *6- Other (specify)* ……	|___|

E4 How does the household dispose human waste?

1- Flush toilet 2- Pit latrine 3- Bush 4- Other (*specify*) ……	|___|

E5 Main Material of walls *(By observation)*

1-Stone/Brick/Block 2- Mud and wood 3- Wood only 4- Iron Sheets 5- Grass / Dirt 6- Other	|___|

E6 Main roofing Material *(By observation)*

1-Concrete 2- Roof Tiles 3- Iron Sheet 4- Wood 5- Palm leaf/sod 6- Other *(specify)* ……	|___|

E7 Main floor Material *(By observation, if possible)*

2- Floor tiles 3- Wood 4- Earth 5- Other *(specify)*	|___|

E8 Does your household own any land?

1 Yes 2 No, → **go to E10**	|___|

E9 What is the main use of the land owned?

1 Crop growing 2 Keeping livestock 3 Commercial use (e.g. Rental. 4 Vacant	|___|

E10- Does your household own any of the following items? 1=Yes or 2= No

Radio	|___|
Television	|___|
Refrigerator	|___|
Washing Machine	|___|
Computer	|___|
Car	|___|
Donkey cart	|___|
Boat / Canoe	|___|

E11: Does this household own any of the following livestock? 1=Yes; 2=No → **End interview**

E12: If Yes, how many? *If more than 995, write 995*

Type of	**E11**: 1- Yes,	**E12:** Number
Camels	|___|	|___|___|___|
Cattle	|___|	|___|___|___|
Shoats	|___|	|___|___|___|
Donkeys	|___|	|___|___|___|
Horses	|___|	|___|___|___|


UNFPA


UKaid
from the British people


SWEDEN

United Nations Population Fund Somalia
Block P, Level 2, UN Complex,
UN Avenue, Gigiri.
P.O. Box 28832 – 00200 Nairobi, Kenya.
Tel: (+254) (0) 207 625 742
www.unfpa.org