STUDY ON AGING IN SUB-SAHARAN AFRICA:

SAMPLING MANUAL

UNITED NATIONS

Department of Economic and Social Affaires

FINAL VERSION PREPARED AFTER THE WORKSHOP IN MALAWI, JULY 18-21 2016

PREFACE

The sampling manual familiarizes the implementing agency with the principal directions required to conduct the Study on Aging in sub-Saharan Africa. The main objective of this sampling manual is to outline the sampling approach for the Survey on Aging in sub-Saharan Africa and to provide guidance to the country's survey implementing agency [IA] how to draw the sample for the survey. The sampling manual follows the official guidelines developed by the United Nations for designing household survey samples. In addition to the general procedures for drawing the sample, the sampling manual provides detailed steps for the implementation of the survey in Malawi, the first sub-Saharan African country that will pilot and conduct the survey.

Table of Contents

1. Introduction 5
2. Study Objectives of the Survey on Aging in sub-Saharan Africa 6
3. General Principles for the Design of the Survey on Aging in sub-Saharan Africa 7
4. Summary of the Key Principles for the Implementation of the Survey on Aging in sub-Saharan Africa 9
5. Target Population and Sampling Procedures for the Survey on Aging in sub-Saharan Africa 11
5.1. Target Population 11
5.2. Sampling Frame 12
5.3. Stratification 13
5.4. Two-stage cluster sampling procedure 13
5.4.1 Possible design modification by interviewing non-coresiding spouses 14
5.5. Sample "take" per cluster 15
5.6. Sample size, sample weights and statistical considerations 16
5.6.1. Sample size 16
5.6.2. Sample weights 17
5.6.2. Sampling error and precision 18
5.7. Household listing 18
5.7.1. General principles 18
5.7.2. Specific procedures for selection of the clusters (EAs) 19
5.7.3. Preparation of household listing data for household selection 20
5.8. Sample Selection of Age-Eligible Households 21
5.8.1. General principles 21
5.8.2. . Equal probability systematic sampling procedures for selection of households 21
6. Survey Data Collection 22
7. Sample Documentation 22
8. Sampling Procedures for the Implementation of the Survey on Aging in sub-Saharan Africa in Malawi 23
8.1. Sampling Frame and Selection of Districts 23
8.2. Determination of Sample Size in Malawi ..... **Error! Bookmark not defined.**
8.3. Household Listing in Malawi 26
8.4. Selection of the household sample 28
8.5. Next Steps for the Implementation of the Survey on Aging in sub-Saharan Africa 28

8.6. Special Instructions for the Administration of the Household Listing, Household Questionnaire and the Individual Questionnaire 31

References 32

1. Introduction

The objective of this sampling manual is to familiarize the implementing agency [IA] with the principal directions required to conduct the Study on Aging in sub-Saharan Africa. The manual outlines the sampling approach and provides guidance to the country's implementing agency [IA] how to draw the sample for the survey. In addition to the general procedures for drawing the sample, the sampling manual provides detailed steps for the implementation of the survey in Malawi, the first sub-Saharan African country that will pilot and conduct the survey.

The Survey on Aging in sub-Saharan Africa is implemented as a nationally-representative stratified random sample of households that includes household members aged 60 years or older. The implementation of the survey follows standard principles of sample selection and documentation, as for instance outlined by the United Nations guidelines "*Designing Household Survey Samples: Practical Guidelines*" (United Nations Statistics Division 2005, hereafter UNSD 2005) and in the Demographic and Health Survey's (DHS) Toolkit "*Sampling and Household Listing Manual*" (ICF International 2012).

The rationale for this study design is as follows: Household surveys are among three major sources of social and demographic statistics in many countries. They are used for collection of demographic characteristics and detailed and varied socio-demographic data pertaining to conditions under which people live, their well-being, activities in which they engage, and cultural factors which influence behavior, as well as social and economic change. Often these surveys are also complemented by data from other sources such as censuses and administrative records.

Household sample surveys have become a key source of data on social phenomena in the last 60-70 years. They are among the most flexible methods of data collection. In theory almost any population-based subject can be investigated through household surveys. It is common for households to be used as second-stage sampling units in most area-based sampling strategies (see chapters 3 and 4 of UNSD 2005). In household sample surveys, part of the population is selected and inferences are made to the whole population. In contrast to a census, household sample surveys are characterized by a greater variety of topics covered, smaller workloads for interviewers and longer time periods assigned to data collection. As household surveys require fewer field staff, more qualified individuals can be recruited and trained more intensively. Moreover, not all the data needs of a country can be met through census-taking; therefore, household surveys provide a mechanism for meeting the additional and emerging needs on a continuous basis. The flexibility of household surveys makes them excellent choices for meeting data users' needs for statistical information which otherwise would not be available.

In the specific case of the Survey on Aging in sub-Saharan Africa, the focus on older individuals also allows the survey to provide extensive training of interviewers to address the specific challenges related to the collection of survey data from older individuals in sub-Saharan Africa who have often limited schooling or no formal education at all and might be affected by declining

physical health, mental health and cognitive abilities. From a probability sample of older individuals, properly selected and implemented, accurate and reliable results can provide a strong basis for making inferences on the total population of older adults in a sub-Saharan African country.

2. Study Objectives of the Survey on Aging in sub-Saharan Africa

The number of older persons in Africa is growing rapidly: between 2015 and 2030 the number of people aged 60 years or over in the region is projected to increase by more than 63 per cent, from 64.4 million in 2015 to 105.4 million by 2030 (United Nations, 2015). Rapid population ageing and growing absolute numbers of older persons globally and particularly in Africa demand the full and comprehensive inclusion of ageing matters into the formulation of many sustainable development goals and targets. More than ten years ago, African governments formally adopted the United Nations Madrid International Plan of Action on Ageing (MIPAA) and the African Union Policy Framework and Plan of Action on Ageing (AUPFAA). Following the adoption of MIPAA, successive sessions of the General Assembly have called for the international community and the United Nations system to "support national efforts to provide funding for research and data collection initiative on ageing" (A/RES/69/146).

The objective of the Survey on Aging in sub-Saharan Africa is therefore to develop a standard methodology to produce, analyze and deliver harmonized indicators on the situation of older persons in Africa, particularly in the framework of the MIPAA and the 2030 Agenda for Sustainable Development. The survey instruments are tailored to the situation of older adults in sub-Saharan African countries. The survey will be first implemented in Malawi in 2017, with the goal of subsequently implementing it in other African countries. The survey addresses data gaps in the three priority areas of the Madrid Plan of Action on Ageing (MIPAA): (1) older persons and development; (2) advancing health and well-being into old age; (3) ensuring enabling and supportive environments. The questionnaire is designed by taking into the consideration and reviewing several leading ageing studies already underway in other parts of the world, especially in Europe (e.g. Survey of Health and Retirement in Europe (SHARE), the English Longitudinal Study of Ageing (ELSA), and the Irish Longitudinal Study of Ageing (TILDA)), in North America (e.g. the Health and Retirement Survey (HRS)), in Latin America (e.g. the Mexican Health and Aging Study (MHAS) and the Brazilian Longitudinal Study of Health, Ageing and Well Being (ELSI-Brazil)), in Asia (e.g. the China Health, Ageing, and Retirement Longitudinal Survey (CHARLS), the Longitudinal Ageing Study in India (LASI)) as well as other harmonized ageing studies such as the WHO Study on Global Ageing and Adult Health (SAGE) and the Longitudinal Studies of INDEPTH communities in Africa (HAALSI). Many other questionnaires such as the Malawi Longitudinal Study of Families and Health (MLSFH), Botswana Core Welfare Indicators Surveys 2009/10, the Living Standards Measurement Study (LSMS) and survey guidelines developed by DSPD's Technical Cooperation Unit were also taken into the consideration.

The Survey on Aging in sub-Saharan Africa consists of two parts:

1) A household questionnaire, with one collected per sampled household by interviewing a selected key informant (ideally the household head) about the members of the household, household finances and living conditions for the members of the sampled household. The goal of the household questionnaire is also to identify age-eligible respondents who will participate in the second part of the study.

2) the individual questionnaire that is to be collected for each household member age 60 and over.

The household questionnaire consists of 8 sections: 1) in the first section, information on the individual responding to the household questionnaire is collected; 2) Section 2 covers the household listing; 3) section 3 focuses on the housing environment; 4) Section 4 collects information on household income; 5) section 5 asks about the household's agricultural income and assets, 6) section 6 asks questions on the household's financial and non-financial assets; 7) the questions in section 7 focus on access to social programs and benefits and 8) the last section 8 evaluates the overall economic situation of the household.

The individual questionnaire is more comprehensive and consists of 15 sections that cover demographic characteristics, information on children ever born, step-children and adopted children, different aspects of physical and mental health, intergenerational relationships and transfers among the respondent and family and friends networks, employment history and current income sources, access to pensions and other social benefits and programs. The individual questionnaire also collects information on abuse of older people, perception of aging, loneliness, and day-to-day experience of older people. In addition, the individual questionnaire includes 3 optional modules with a focus on migration history, health behaviours and risk factors such as smoking and alcohol consumption and physical activity. The country's implementing agency [IA] will determine if these modules are to be implemented depending on the needs and data gaps on older adults identified in each country-specific context.

3. General Principles for the Design of the Survey on Aging in sub-Saharan Africa

In order for surveys to provide cost-effective and reliable means for making inferences for the total population of older adults in a country, several general principles need to be followed (see UNSD 2005 for detailed discussion of these principles):

1. For a survey to yield desired results there is a need to pay particular attention to the preparations that precede the fieldwork. In this regard all surveys require careful and judicious preparations if they are to be successful.

2. It is imperative that the objectives of a survey be clearly spelled out from the start of the project. There should be a clear statistical statement on the desired information, giving a clear description of the population and geographical coverage. It is also necessary at this stage to

stipulate how the results are going to be used. The given budget of the survey should guide the survey statistician in tailoring the objectives. Taking due cognizance of the budgetary constraints will facilitate successful planning and execution of the survey.

3. It is very important that stakeholders, such as various users and producers of statistics, be involved in defining the objectives of the survey as well as its scope and coverage. The consultations help to come up with consensus or compromises on what data are needed, the form in which data are required, levels of disaggregation, dissemination strategies and frequency of data collection.

4. Only probability samples following well-established sampling procedures are suitable for making inferences from the sample population to the larger overall population that it is designed to represent. Snow-ball or convenience samples are generally not suitable for this purpose and should therefore not be implemented. Instead, virtually all sample designs for household surveys, both in developing and developed countries, are implemented as multi-stage, stratified and/or clustered probability samples.

Probability sampling in the context of a household survey refers to the means by which the elements of the target population -- geographic units, households and persons -- are selected for inclusion in the survey. The requirements for probability sampling are:

- that each element must have a known mathematical chance of being selected,
- that chance must be greater than zero and
- it must be numerically calculable.

In order to be cost-effective, most household surveys are not implemented as simple random samples. Instead, the sampling procedure usually includes stratification to increase the efficiency of the sample design and ensure that the sample actually selected is spread over geographic sub-areas and population sub-groups properly. In addition, the sampling design usually makes use of clusters of households in order to keep costs to a manageable level.

5. The size of the sample must take into account competing needs so that costs and precision are optimally balanced. The sample size must also address the needs of users who desire data for sub-populations or sub-areas – domains.

6. A two-stage sample design is a well-established procedure for implementing household surveys. It is the sample design of choice for the Multiple Indicator Cluster Surveys (MICS) that UNICEF has carried out in over 100 countries since the mid-1990s (http://www.unicef.org/statistics/index_24302.html). It is also used predominantly in the Demographic and Health Surveys (DHS). Typically the two-stage design involves stratifying the sampling frame (generally based on the most recent Census of Population and Housing) by regions (such as provinces) and by urban/rural classification. At the first sampling stage several hundred relatively small geographical areas (referred to as primary sampling units or PSUs), are selected systematically with probability proportional to size (PPS), separately within each stratum. The

PSUs are generally based on enumeration areas (EAs) defined for the census operations.. For the second sampling stage it is necessary to have a list of all the households in each sample PSU that can be used for selecting the sample households. Generally the list of households from the last census is out-dated, so an updated listing of households needs to be implemented in the sample PSUs.. During this listing it is also necessary to identify the households with any special subpopulation of interest for the survey. This is of particular relevance in the case of the Survey on Aging in sub-Saharan Africa since few countries will have reliable recent household listing information that identifies the target population for this survey (i.e., the population age 60+).

The household listing is followed by a random systematic sample of a fixed number of eligible households at the second stage. The PSUs are based on small geographical units, commonly referred to as the "clusters", usually defined as villages or census enumeration areas (EAs) in rural areas and city blocks in urban areas. This two-stage design is appealing because of its simplicity and reliability, thereby reducing non-sampling errors in the survey implementation, and because it is approximately self-weighting within stratum (all the households in the stratum are selected with a similar probability). The sample is almost never completely self-weighting within stratum, because the number of households listed in the sample PSUs is different from the number of households in the census frame, especially when we are identifying households with at least one person age 60 years or older during the listing. Also, the sample allocation by stratum generally results in different weights by stratum. As a result, it is always necessary to calculate sampling weights for the tabulations and analysis, where the weight is equal to the inverse of the overall probability of selection.

7. It is important in designing and executing a household survey to develop good survey procedures aimed at maximizing the response rate. We emphasize the importance of having procedures in place to reduce the number of refusals, such as arranging to return to conduct an interview at the convenience of the respondent. Also, the objectives and uses of the surveys should be carefully explained to reluctant respondents to help win their cooperation.

8. An alternative to implementing a stand-alone Survey on Aging in sub-Saharan Africa in a specific country is to include relevant modules of the questionnaire for this survey in other suitable national surveys that include individuals aged 60 and over as respondents. This approach limits the information that is gained on the older population, but it may nevertheless represent a useful option for countries without funds for a stand-alone Survey on Aging. Details of such an approach are very country specific and need to be evaluated for each specific context where such an option is considered.

4. Summary of the Key Principles for the Implementation of the Survey on Aging in sub-Saharan Africa

The Survey on Aging in sub-Saharan Africa is implemented as a nationally-representative stratified two-stage clustered-random sample of households including individuals aged 60 and older based on established survey procedures. This is a cost-effective and reliable way to collect population-

level information on the well-being, socioeconomic situation and demographic context of older individuals in sub-Saharan African countries. Key principles of the Survey on Aging in Sub-Saharan Africa include:

Target population: The target population for the survey is defined as non-institutionalized individuals age 60 years and older living in individual households in the selected countries.

Household sample: The survey is implemented as a nationally-representative stratified two-stage clustered-random sample of households that include household members aged 60 years and older.

Selection of household members: All regular household members aged 60 years and older in the sampled households are selected for participation in the survey. This includes the spouses who are age 60+ years old and who reside in the same household. However, age–eligible spouses not co-residing in the same village will not be followed due to cost and logistical considerations.

Use of an existing sampling frame: A clustered random sample of households can only be obtained from an existing sampling frame which is a complete list of statistical units covering the target population. In order to be cost-efficient, the Survey on Aging in sub-Saharan Africa should rely on an adequate pre-existing sampling frame which is officially recognized (e.g., census frames, complete list of villages/communities, sampling-list from other nationally-representative surveys).

Full coverage of the target population: The Survey on Aging in sub-Saharan Africa should be nationally representative and cover 100% of the target population (i.e., no household-based persons age 60 and older are systematically excluded).

Probability sampling: The household sample for the Survey on Aging is obtained as a probabilistic sample based on an existing sampling frame using established sampling procedures (e.g., see ICF International 2012; United Nations Statistics Division 2005). This is the only way to obtain unbiased estimation and to be able to calculate the sampling errors and other measures of precision. A probability sample obtained through a two-stage household-based sample design will ensure cost-efficient, transparent and reliable implementation (for a detailed discussion of two-stage sampling, see Chapter 3.7 in United Nations Statistics Division 2005). The term **"probability sampling"** excludes purposive sampling, quota sampling, and other uncontrolled non-probability methods because they cannot provide the evaluation of precision of survey findings.

Use of a suitable sample size: A suitable sample size is determined given the trade-offs between survey precision, data quality, organizational capacities and survey budget; considerations for determining sample size are outlined in this manual. ***Comparisons with related multi-purpose surveys in SSA, and considerations of the budget and the initial need of an overall population***

perspective on aging in SSA, suggest a target national sample size of approximately 2,000 respondents (men and women age 60 and over) for the Survey on Aging in sub-Saharan Africa.

Conducting a household listing and pre-selection of households: Data quality is enhanced if eligible households are pre-selected prior to the start of fieldwork rather than by teams in the field. In many sub-Saharan African countries, a recent and reliable household listing in sample enumeration areas (EAs) that carefully enumerates older individuals is not available. The sampling procedure for the Survey on Aging in sub-Saharan Africa therefore includes conducting a specific household listing in selected EAs that provides an effective basis for selecting respondents for this survey. This household listing operation will be conducted before the main survey. This complete household listing of all residential households with individuals aged 60 years and older in each of the selected clusters will then be used for the random selection of households that will participate in the survey. To ensure the highest data quality, the random selection of the households will be performed by the [IA] with the assistance of personnel from the country's national statistical office.[1]

The interviewers are then asked to interview only the pre-selected households with age-eligible household members aged 60 and older, where all age-eligible household members will be surveyed. If spouses age 60 and older are not co-residing in the same household they are excluded from the sample because of higher costs associated with finding them in different locations, and the fact that the spouse would also have a chance of being selected in the household where they currently reside. To assure accuracy of the sample, no changes or replacements of pre-selected households are allowed in the field.[2]

Maintaining confidentiality of individual's information: Adequate procedures are implemented during data collection and data management to ensure the privacy of the survey data collection and confidentiality of the collected data. The protection of a respondents' privacy and confidentiality of data are key concerns for the Survey on Aging in sub-Saharan Africa.

5. Target Population and Sampling Procedures for the Survey on Aging in sub-Saharan Africa

5.1. Target Population

The Survey on Aging in sub-Saharan Africa is implemented as a nationally-representative clustered random sample of households that include household members aged 60 years and older. All

[1] This approach implies that the interviewers and/or supervisors will not be in charge of selecting the age-eligible households.

[2] An exception is if the IA has a preselected set of replacement households from the sampling frame that can be used to replace non-respondent households.

regular household members aged 60 and older in the sampled households are selected for the survey. Age-eligible spouses not co-residing in the same household will not be followed due to cost and logistics considerations. Each spouse 60 years or older residing in a different household has an individual probability of selection, so this procedure also avoids a household having a multiple chance of selection.

The sampling frame, stratification, and sample selection for this two-stage household-based survey follow established procedures (e.g., see ICF International 2012; United Nations Statistics Division 2005).

5.2. Sampling Frame

A first stage sampling frame is a complete list of all primary sampling units (PSUs) that entirely covers the target population. The second stage sampling frame is based on a new listing of households in sample PSUs.The existence of a sampling frame allows a probability selection of sampling units (for a general discussion of sampling frames, see Chapter 4 in United Nations Statistics Division 2005).

The conventional sampling frame for a two-stage household-based survey such as the Survey on Aging in sub-Saharan Africa consists of a ***list of Enumeration Areas (EAs) from a recently completed population census***. This is available in most sub-Saharan African countries where the survey is to be implemented. An EA is usually a geographic area which groups a number of households together for practical enumeration purposes for the census, and for which the number of households and population at the time of the census are known, as well as possibly other population characteristics (e.g., socioeconomic status). A complete list of EAs covers the entire geographic area of the country and is the most ideal frame for the Survey on Aging in sub-Saharan Africa. Since the EAs are delineated for the purpose of census enumeration, they generally have a relatively small number of households (for example, between 80 and 120 households, which is a practical size for the listing operation. It is important that the EAs have well-defined boundaries, which are generally defined on census maps.

If the country has a pre-existing and well-documented master sample based on the census frame, it may be used for selecting the EAs for the Survey on Aging in sub-Saharan Africa. When a list of EAs from a recent census is not available, alternative sampling frames can be considered, including for instance a list of administrative units such as villages with estimated number of households or population for each unit, a list of electoral zones with estimated number of qualified voters for each zone, or a gridded high resolution satellite map with estimated number of structures for each grid. The specific aspects of selecting PSUs based on these sampling frames would need to be evaluated in each specific context.

A careful evaluation of the sampling frame with respect to coverage, distribution, identification and coding, and consistency is required (for a detailed discussion of sampling frame evaluation, see Chapter 4 in UNSD 2005).

5.3. Stratification

Stratification is an important aspect of the sample design for the Survey on Aging in sub-Saharan Africa. Stratification is the process by which the survey population is divided into subgroups or strata that are as homogeneous as possible using certain criteria. Within each stratum the sample is selected independently. Stratification is desirable to improve the efficiency of the sample design. The stratification is generally consistent with the geographic domains defined for the survey. It is important to ensure that subpopulations of interest are adequately represented in the survey, and thus to reduce sampling errors for subgroup-specific estimates.

Stratification for the Survey on Aging in sub-Saharan Africa should take into account the geographic domains as well as the urban and rural populations. In the sampling frame the EAs are stratified for the first stage of selection.. For example, a typical two-level stratification involves first stratifying the EAs by region at the first level and then by urban-rural within each region.

Stratification by region also allows the possibility of obtaining some results at the regional level, depending on the sample size; it would be necessary to ensure a sufficient number of sample households in each region. If it is not possible to conduct a national-level survey, it would also be possible to limit the survey to selected districts/areas (for a variety of reasons, such as minimizing financial costs or safety reasons if applicable). However, in this case the report would have to state that the results are only representative of the geographic domains included in the survey.

5.4. Two-stage cluster sampling procedure

The Survey on Aging in sub-Saharan Africa is implemented using a stratified two-stage cluster sampling procedure. A cluster is a group of adjacent households which serves as the primary sampling unit (PSU) for field work efficiency; the PSU should have well-defined boundaries identified on a map. Interviewing a certain number of households in the same cluster can reduce greatly the amount of travel and time needed during data collection and the corresponding travel costs. A cluster is typically an EA with a measure of size equal to the number of households or the population in the EA, provided by the population census.

At the first stage, a sample of EAs is selected systematically with probability proportional to size (PPS): in each stratum, a sample of a predetermined number of EAs is selected independently with probability proportional to the EA's measure of size. Ideally, the size should be measured in terms of the number of individuals aged 60 and older per EA; but such detail may not be available. Since age information is available in the most recent census data, it is important to check with the Statistical Office of the country to determine the possibility of obtaining a special EA summary

database with the population age 60 years and older in each EA. However, if variations in age structures across EAs are relatively modest, sampling proportional to overall population size or total adult population size in each EA will provide a reasonable alternate approach. In the selected EAs, a listing procedure is performed such that all dwellings/households are listed. This procedure is important for updating the information in the sampling frame, and it provides a sampling frame for the household selection; this procedure is particularly important for the Survey on Aging in sub-Saharan Africa as few countries in the region will have reliable household listings that provide accurate enumerations of households with older individuals in the EAs.

At the second stage, after a complete household listing is conducted in each of the selected EAs, a fixed number of households containing (regular) household members aged 60 and older is selected using random systematic sampling (with equal probability) in the selected EAs.

In *each* selected household, the following questionnaires are completed:

1) A household questionnaire with the household head or other person with a knowledge about the household's circumstances
2) An individual questionnaire with *all* regular household members aged 60 years and older

The above design that surveys *all* household members aged 60 years has the advantage of maximizing the number of respondents for a given sample of households, and ismore cost-effective for achieving a specific sample size.

Some disadvantages of this approach however need to be considered. Given within-household correlation of observations, this study design has lower statistical power than an equally-sized sample of respondents aged 60 and over drawn all from separate households. There are also certain logistical and fieldwork challenges, such as ensuring the feasibility of interviewing multiple persons per household either simultaneously (probably not feasible) or in temporal proximity. However, the alternative of selecting more households and then randomly selecting one eligible elderly person per household would be more costly and result in differential weights per household (depending on the number of persons age 60 or more in each household).

5.4.1 Possible design modification by interviewing non-co-residing spouses

A possible modification of the above study design is to consider the inclusion of age-eligible spouses aged 60 years and older who are not co-residing in the same household, but live in nearby proximity (i.e., live in the same village). Following this approach is advantageous since it will provide the opportunity of analysis to investigate the interactions among spouses, within and between household variation of health and other characteristics of older individuals, and the interdependence of health and wellbeing between spouses. Some downsides of this approach however need to be considered. Given within household correlation of observations, this study design has lower statistical power than an equally sized sample of respondents aged 60 and over drawn from separate households. It would also be necessary to take into account multiple probabilities of selection in the calculation of the weights, since an eligible elderly person in a

household with a non-co-residing spouse would also be selected if the household of the spouse is selected. However, the questionnaire could include questions regarding any non-co-residing spouse, in order to study this phenomenon further.

5.5. Number of sample households per cluster

A key design decision for this two-stage sampling procedure is the *sample "take" per cluster*, that is, the number of households to be selected in each cluster. The DHS recommends for its surveys a sample take of approximately 25—30 households per EA. For the MICS, some countries select 20 households per cluster, in order to provide a more disperse sample and lower the design effects and sampling errors, as discussed in the next section. A lower number of sample households per EA also results in higher costs, since additional sample EAs would need to be selected and listed. Another important consideration is that since only a portion of the households in each EA will have an eligible member age 60 years or older, it will be necessary to determine the average number of households per PSU with eligible elderly that would be available for selection. Because in the Survey on Aging in sub-Saharan Africa more than one eligible individual aged 60+ may reside in a household, it would still be effective to select less than 25 households per EA.

Estimates of the number of 60+ individuals per household with at least one regular household member aged 60+ can be obtained from recent census or household survey data. For example, if each eligible household contains on average 1.5 individuals aged 60+, then a sample take of 25—30 respondents per cluster corresponds to 17—20 (25/1.5 to 30/1.5) households per cluster. If the overall target sample size of the Survey on Aging in sub-Saharan Africa, for example, is 2,000 individuals, approximately 67—80 clusters (EAs) should be selected. In each of these clusters, 17—20 households with at least one household member above age 60 are selected, resulting in a sample take of approximately 25—30 individuals aged 60+ (=primary respondents) per cluster who are selected for interviews.

If one of the objectives of the survey is to obtain reliable estimates for subnational domains, such as regions, a target sample size would have to be determined for each domain.

A fixed number of sample households per cluster is easy for survey management and implementation, but it requires sampling weights that vary by sample PSU within a stratum, although the first stage selection of the PSUs with PPS will reduce this variability in the weights. This approach also has the advantage of easier survey implementation, and a variable sample take in which the number of households per cluster varies across clusters is only recommended in exceptional cases for the Survey on Aging in sub-Saharan Africa.

For further reference, additional discussions regarding optimal sample take and variable sample take for obtaining self-weighting samples are contained in (United Nations Statistics Division 2005) and (ICF International 2012). Since the weighting of the survey data is done

automatically by the tabulation software, the need for a self-weighting sample is not so important, although the design should try to avoid any extreme variability in the weights.

5.6. Sample size, sample weights and statistical considerations

5.6.1. Sample size

The factors and parameters that must be considered in determining the sample size are many but they revolve primarily around the measurement objectives of the survey. Detailed discussions of sample size considerations are for instance included in (UNSD 2005). Typically the sample size is influenced by the desired level of precision for the key indicators, the estimation domains, whether measuring level or change, clustering effect, allowance for non-response and available budget. Clearly, sample size is the pivotal feature that governs the overall design of the sample.

Guidelines for determining sample sizes are discussed for instance in the United Nations "*Designing Household Survey Samples: Practical Guidelines*" (UNSD 2005) and in the Demographic and Health Survey's (DHS) Toolkit "*Sampling and Household Listing Manual*" (ICF International 2012). Key inputs into the sample size calculations are the predicted value of the key proportion indicator being estimated (or standard deviation of the variable); the desired level of precision of the key indicators for the target population (or relevant subsets thereof, depending on the national context); as well as the number of respondents per cluster and the extent of correlation of measures within clusters (both of which influence the design effects or "deff"). It is also necessary to adjust the sample size based on the expected overall level of nonresponse based on past experience. It may also be necessary to oversample certain subpopulations because of their singular interest for the study.

Because of the scarcity of data on older individuals in SSA, uncertainty about which indicators in the data will be most important, lack of data to estimate design effects, and the lack of knowledge about the level of disaggregation for which data will be analyzed, many inputs for the performance of detailed sample size calculations are missing. Moreover, for a multi-purpose survey such as the Survey on Aging in sub-Saharan Africa, there are many indicators which would require different sample sizes. As a result, conventional sample size calculations can only provide limited guidance about the target sample size of this study.

An alternative for determining sample size is therefore to decide the sample size based on comparisons with other surveys that have broadly similar design and similar purpose. It should to be noted however that most surveys on aging conducted in other regions sample individuals who are younger than the sample targeted for the Survey on Aging in sub-Saharan Africa. For instance the national samples in the first wave of the WHO SAGE studies range from 2,742 individuals in the Mexican study to 14,811 in the Chinese study (men and women combined, with an age range starting at age 18). The sample size of individuals age 60 and over in the first wave of SAGE varies from 2,142 older individuals in the South African Study to 7,474 in the Chinese study (Kowal et al. 2012). DHS sample sizes vary depending on whether the standard DHS survey or the Interim DHS survey is administered, but they vary from 5,000 to about 30,000 households

(http://dhsprogram.com/What-We-Do/Survey-Types/DHS.cfm). The Malawi DHS 2015-16 had a sample size of 26,362 households (http://dhsprogram.com/what-we-do/survey/survey-display-483.cfm). The Longitudinal Study of Aging in India (LASI) currently has a pilot sample size of approximately 1,547 sampled individuals (men and women), however beginning at age 45. Among aging-related studies in SSA, the older adult sample of the Malawi Longitudinal Study of Families and Health includes approximately 1,300 individuals, and the goal of the Health and Aging in Africa: Longitudinal Studies of INDEPTH Communities (HAALSI) is to establish a cohort of about 4,000 men and women age 50 and over in Agincourt, South Africa (https://www.hsph.harvard.edu/population-development/research-focal-areas/major-projects/haalsi/).

Given the limited budget for the implementation of the Survey on Aging in Sub-Saharan Africa, the initial need of an overall population perspective on aging in SSA, and a weighing of the various trade-offs affecting sample size and precision of estimates, the comparison with other related data collection efforts suggest that a target sample size of 2,000 individuals aged 60 and older will make it possible to achieve the aims of this survey, allowing disaggregation of estimates by broad age groups, gender and rural/urban and/or major ethnic groups.

The sample will contain slightly more women than men, given differential survival to older ages, and this gender gap will increase with age. The age-structure in the sample will reflect the age structure of the older population of the respective countries, that is, including a relatively large proportion of relatively young older individuals (around ages 60-75 years), and relatively few individuals who will be 80+ years old. This age structure of the sample population is appropriate and desirable given the overall purpose of the UN Study of Aging.

5.6.2. Sample weights

Weighting for household surveys generally involves up to three operations – calculation of the base or design weights, adjustments for non-response, and sometimes adjustments for post-stratification when accurate population projections are available. In many applications only the design weights are used, adjusted for nonresponse. In comparatively few applications the weights may be adjusted based on the distribution from an independent source of data such as a recent census or population projections. The latter is often referred to as post-stratified weighting. In some applications no weighting is done at all; this would occur when two conditions are met: the sample is completely self-weighting and the data generated are restricted to percentage distributions, proportions and ratios, as opposed to estimated totals or absolutes. However, national household surveys are almost never self-weighting, but this is not a problem because appropriate sampling weights can be attached to the data.

In practice, household survey samples are rarely self-weighting at the national level for several reasons. First, sampling units may be selected, by design or in deviation from the protocol, with unequal probabilities of selection. Indeed, even though the PSUs are often selected with

probability proportional to size, and households selected at an appropriate rate within PSUs to yield an approximately self- weighting design, this may be nullified by the selection of one person for interview in each sampled household, or in the specific case, by variation in the age structure and family compositions across clusters. Second, the selected sample often has deficiencies including non-response and non-coverage that vary by age and/or gender. Third, the need for precise estimates for domains and special subpopulations often requires over-sampling these domains to obtain sample sizes large enough to meet pre-specified precision requirements. Fourth, when the sample design entails preparing a current listing of households in the selected clusters (PSUs or SSUs) and a pre-determined fixed number of households is to be selected in each cluster, the actual probability of selection of the household is somewhat different than its design probability, the latter of which was based on frame counts rather than current counts of households; consequently, unequal probabilities of selection arise even though a self-weighting design may have been targeted.

Discussion of how weights are computed are provided for instance in the DHS manual (ICF International 2012) and UN Sampling Manual (UNSD 2005).

5.6.2. Sampling error and precision

The sampling error, a measure of the precision of estimated indicators, is affected by the size of sample (number of respondents age 60 and older), as well as various design decisions such as number of sample clusters and households selected per cluster, and the correlation of observations within clusters. Additional complications arise in the specific study design as the selection of multiple individuals per household results in within-household correlation of observations. None of these aspects cause any problems in the analyses of the data, as standard statistical software allows to adjust for such complex survey designs. Although accurate a priori sampling error calculations may not be feasible as the relevant inputs (including correlations of observations within clusters and households, and number of respondents per household) are not known prior to the data being collected, it may be possible to use the results from the Census and other surveys to estimate approximate values. General principles of the determinants of sampling error, and the trade-off between various study design parameters and precision of estimates apply to this study, and these principles are for instance discussed in the DHS manual (ICF International 2012) and UN Sampling Manual (UNSD 2005).

5.7. Household listing

5.7.1. General principles

The household listing is an essential operation for the Survey on Aging in sub-Saharan Africa as it identifies eligible households with individuals aged 60 and older as regular household members. The complete household listing is the first step required before the implementation of the survey.

The listing operation is an important procedure for the Survey on Aging in sub-Saharan Africa, especially when the sampling frame is outdated or concerns exist that it might have been subject to underreporting of older individuals. The listing operation provides an updated complete list of occupied residential households in the EA, and identifies households that include regular household members aged 60 and older. This information is necessary for an equal probability random selection of households at the second sampling stage.

Data quality is ensured by conducting the household listing prior to the main survey, as it is possible to pre-select the sample households in advance and requires interviewers to only survey the pre-selected households without replacement of non-responding households. The collected GPS coordinates for each household, along with other household characteristics that are collected as part of the household listing, will ensure that the sampled households can be easily found by survey interviewers later.

It is the decision of the country's [IA] how the household listing will be implemented in practice. If possible to implement, it is recommended that the household listing is conducted using tablet computers or related devices that minimize data entry needs and that automatically assign unique IDs to households and screens for households with at least one member age 60 years or older.

For the listing it is important to have a census map for each sample EA that clearly identifiesthe EA boundaries. All the structures within the boundaries of the sample EA should be listed to identity households in occupied dwelling units. The listing sheet or application should include a question on whether the household has at least one member who is age 60 years or older.

5.7.2. Specific procedures for selection of the clusters (EAs)

The sample of EAs for the Survey on Aging in sub-Saharan Africa is chosen from the sampling frame using systematic probability proportional to size (PPS) sampling (UNSD 2005). There are many ways for drawing a pps sample of EAs, and the following guidelines are adapted from the DHS (see chapter 3.3 and example 3.3.1 in ICF International 2012).

1. List the sampling units (EAs) available from the sampling frame with their measure of size M_i. Ideally, size is measured in terms of the number of individuals aged 60 and older per EA; although such detail may not be available, the National Statistical Office can be requested to produce such a listing from the most recent Census data. If variations in age structures across EAs are relatively modest, sampling proportional to overall population size or total adult population size in each EA will provide a reasonable alternate approach.
2. Calculate the cumulative measure of size $C_k = \sum_{i=1}^{k} M_i$ for each unit k, and check that the last entry C_N equals the total measure of size $\sum_{i=1}^{N} M_i$.
3. Let n be the number of primary sampling units (PSUs) to be selected. Compute the sampling interval $I = \frac{\sum_{i=1}^{N} M_i}{n} = \frac{C_N}{n}$.
4. Generate a random number R between 0 and I.

5. Compute the sampling numbers *R*, *R +I*, *R +2*I*, ..., *R +(n-1)*I*., where these values are rounded up to the next integer.
6. For each sampling number *R +(j-1)*I*, the j^{th} sampled unit is unit *k* if C_k is the first cumulative size bigger than or equal to the sampling number *R +(j-1)*I*.
7. Calculate the selection probability of each selected unit *j*: $\frac{n*M_j}{\sum_{i=1}^{N} M_i}$

5.7.3. Preparation of household listing data for household selection

After the household listing is completed, the next step implies the random selection of the age-eligible households based on the household listing. The procedures for the preparation of the household listing data for the household random selection include the following:

Household IDs:

All households in each cluster are assigned a serial number that sequentially identifies each household within each cluster; the household IDs are uniquely assigned to each household.

Identification of eligible households for Survey on Aging in sub-Saharan Africa:

All households containing at least one regular household member aged 60 and older are identified

Household listing for eligible households and Individual IDs:

For each eligible household, i.e., a household containing at least one regular member aged 60+, all household members are listed along with basic socioeconomic and demographic information. Unique Individual IDs for each individual within each household are created by combining Household IDs with a sequential ID for each listed person. The combining of household and individual serial numbers should be done by the computer application, not by the enumerator. We should keep the instructions as simple as possible for the enumerators.

Households per cluster:

The total number of households and eligible households (those containing at least one regular member aged 60+) is reported for each survey cluster.

5.8. Sample Selection of Age-Eligible Households

5.8.1. General principles

The target-sample of households is selected after the household listing by the [IA]'s Project Directors. For the Survey on Aging in sub-Saharan Africa, within each selected cluster (EA), a random subset of households containing regular household members aged 60 and older will be selected for the survey.

The procedure for selecting is as follows, where N is the overall sample size (e.g., N = 2, 000 individuals aged 60+), x_c is the sample take in terms of the number of eligible households per cluster (eligible household = household with at least one regular household member aged 60+), n_{HH} is the number of eligible individuals (individuals aged 60+) per eligible household, and $n_c = \frac{N}{x_c * n_{HH}}$ is the total number of primary sampling units (EAs or clusters):

Selection of primary sampling units (EAs or clusters):

A total of $n_c = \frac{N}{x_c * n_{HH}}$ primary sampling units (EAs or clusters) are randomly chosen from the existing sampling frame using systematic sampling with probability proportional to size (for a detailed discussion of two-stage sampling, see Chapter 3.7 in United Nations Statistics Division 2005).

Sampling of eligible households:

A total of x_c eligible households with at least one member aged 60+ are selected from the household listing in each cluster using equal probability systematic sampling.

Selection of eligible individuals in sampled households:

All eligible individuals, i.e., all individuals aged 60+ in the sample of eligible households are selected for survey interviews.

Since the sample will be stratified (for example, by region, urban/rural), it will first be necessary to allocate the sample clusters by stratum (for example, using proportional allocation). Then the selection procedures described above would be implemented separately for each stratum.

5.8.2. . Random systematic sampling procedures for selection of households

Within each selected EA, households containing individuals aged 60 and older are selected for the Survey on Aging in sub-Saharan Africa using random systematic sampling.. The procedures for drawing an equal probability systematic sample of size *n* are for instance outlined in chapter 3.2 and example 3.2.1 in ICF International 2012. A template for the random systematic selection of households from the listing for each sample cluster was developed for the MICS (MICS Systematic Random Selection of Households Template 20171208.xlsx), and can be downloaded from the website http://mics.unicef.org/tools. This template can easily be adapted for other surveys such as the Survey on Aging in Sub-Saharan Africa.

6. Survey Data Collection

The implementation of the survey data collection is described in detail in the "Interviewers' Manual for the Survey on Aging in Sub-Saharan Africa".

7. Sample Documentation

Sample documentation is an essential part of the Survey on Aging in sub-Saharan Africa, and includes all useful information about the study design that is relevant for data analysis and data quality assessment. Sample documentation should be produced separately by the Implementing Agency [IA] in each implementing country and describe all aspects of how the study population was selected. It is expected that the implementing countries and IA will follow these guidelines for drawing the sample population for the Survey on Aging in sub-Saharan Africa, but in specific cases it is possible that the IA may adopt some modifications because of special circumstances. Hence, all procedures have to be described in detailed and the sample documentation should include the following aspects (ICF International, 2012):

- Target population
- Expected sample size
- Sampling frame
- Stratification
- Sampling procedures
- Household listing results
- Sampling weights
- Results of survey implementation and response rates
- Sampling errors and limitations of survey implementation

8. Sampling Procedures for the Implementation of the Survey on Aging in sub-Saharan Africa in Malawi

Malawi was the first sub-Saharan African country in which the survey on aging was implemented and piloted. The survey will provide important information about the situation of older individuals aged 60 years and over in the country andwill provide important information about the overall feasibility of the study, test the validity of the survey instruments and help identify potential problems that can be only identified during a data collection process and feedback from respondents and survey team members.

Below is a detailed description of the sampling procedures for the implementation of the survey in Malawi. This sampling procedure has been discussed in detail and agreed upon at the meeting of the Technical Group during the regional workshop in Malawi, July 18-21, 2016. The Technical Group was comprised of members from the National Statistical Offices of Malawi and Kenya, representatives from the Chancellor College (Malawi), the UN Population Division and the consultant on the development of the survey instruments and sampling procedure.

8.1. Sampling Frame and Selection of Districts

- The primary objective of the sample design is to produce statistically reliable estimates of indicators for four selected districts located in three regions in Malawi, namely, Northern,Central and Southern regions. The selection of the districts were based on a number of considerations, including the proportion of persons age 60 years and older, available budget and time constraints. Initially a sample of 8 districts was considered, but then this was reduced to 4 districts, and the number of sample PSUs per district was increased to provide more reliable results at the district level.The final selected districts are: Mzimba
- Lilongwe rural
- Mangochi
- Nsanje

These districts cover different geographic regions, ethnic groups and other characteristics. They also have the highest proportion of older individuals aged 52 years or older in the 2008 Census data. The number of households per district with household members aged 52 years and older in the 2008 Census was determined, since these individuals would be aged 60 or older in 2016.

The survey will only provide indicators for the four selected districts, so it will not be possible to obtain national-level indicators from the survey data. The sampling frame was based on the data and cartographic materials from the 2008 Malawi Census.

Table 1 shows the distribution of the EAs, households and population for the four selected districts from the 2008 Census data. This table also shows the number of households with at least one

person age 52 years or older, and the corresponding population in this age group. Given that the Census was conducted at least 8 years earlier, surviving persons in that age group would be at least 60 years old in 2016.

Table 1. Distribution of EAs, households and population age 52+ by district, 2008 Malawi Census

District	No. EAs	Total no. of households	Average no. households per EA	Total population	No. of households with 52+	Households with 52+ per EA	Total population 52+	Persons 52+ per household with 52+
Mzimba	829	135,507	163	714,211	36,369	44	65,302	1.8
Lilongwe	1,158	262,378	227	1,213,923	66,440	57	105,165	1.6
Mangochi	613	175,642	287	787,155	47,202	77	70,778	1.5
Nsanje	240	47,734	199	228,256	14,868	62	23,014	1.5
Total	2,840	621,261	219	2,943,545	164,879	58	264,259	1.6

It can be seen in Table 1 that the districts vary considerably in size. The average number of households per EA also varies by district, from 163 in Mzimba to 287 in Mangochi. The average number of households with at least one person age 52+ per EA also varies by district, from 44 for Mzimba to 77 for Mangochi, with an overall average of 58 households with elderly persons per EA. Some households have two or more persons age 52+. Table 1 also shows that the overall average number of persons age 52+ in the households with at least one person in this age group is 1.6, and varies by district from 1.5 in Mangochi and Nsanje to 1.8 in Mzimba.

Within each district the EAs are also variable in size, and some EAs are much smaller or larger than the average. Table 2 shows the distribution of the EAs by the number of households with persons age 52+.

Table 2. Distribution of EAs by district and number of households with persons 52+, 2008 Malawi Census

No. households with persons 52+	Number of EAs by district and size group				
	Mzimba	Lilongwe	Mangochi	Nsanje	Total
1-9 hhs. with 52+	19	5	4	0	28
10-19 hhs. with 52+	93	32	7	4	136
20-29 hhs. with 52+	165	109	27	13	314
30-49 hhs. with 52+	267	341	84	65	757
50-74 hhs. with 52+	198	394	180	89	861
75-99 hhs. with 52+	54	218	169	53	494
100-149 hhs. with 52+	27	57	133	15	232
150+ hhs. with 52+	6	2	9	1	18
Total	829	1,158	613	240	2,840

Most of the EAs are in the range of 20 to 99 households with persons age 52+, but there are a few EAs with less than 10 households with older persons, and a few with more than 150 households with older persons. Given this variability in the size of the EAs, it is recommended to select the EAs with probability proportional to size (PPS); the measure of size would be based on the number of households in the EA with at least one person age 52 years or more from the 2008 Malawi Census frame. This number should be highly correlated with the number of households that currently have at least one person age 60 years or more. In this case the EAs with less than 12 households with at least one elderly person would have a very small chance of being selected in the sample.

8.2. Sample Size and Allocation for Malawi Survey on Aging

A two-stage sample design was used for each of the four districts. A sample of 40 EAs was selected within each district at the first stage, using systematic sampling with PPS. A listing was conducted in each sample EA to cover all households within the EA boundaries, with screening questions to identify the households with at least one person age 60+. These households with at least one elderly person would be eligible for selection. At the second sampling stage 12 households with persons age 60+ were selected in each sample EA, and all the household members in this age group were interviewed for the survey. In this case the total sample size for each district was 480 households with persons aged 60+, and 1,920 sample households for all 4 districts. This sample allocation is shown in Table 3. This table also shows the 2008 Census average number of persons age 52+ per household with at least one person in this age group. The total number of sample persons age 52+ was estimated by multiplying the number of eligible households selected by the average number of elderly persons per household. Finally the effective sample size for elderly persons age 60+ was estimated by multiplying the total number of sample individuals aged 60+ by 0.9, based on a conservative estimate of the response rate and the smaller percentage of persons age 60+ compared to 52+.

Table 3. Allocation of sample EAs and households with population age 60+ by district for Malawi Survey on Aging

District	No. sample EAs	No. sample households with person 52+ per EA	Total no. sample households	Average persons 52+ per household with at least one person 52+	Estimated no. sample individuals 52+	Estimated completed elderly interviews
Mzimba	40	12	480	1.8	819	737
Lilongwe	40	12	480	1.6	722	650
Mangochi	40	12	480	1.5	684	615
Nsanje	40	12	480	1.5	706	635
Total	160	12	1,920	1.6	2,930	2,637

The decision to select 12 households with persons 60 years or older per sample EA was based on considerations for ensuring a moderate design effect based on the experience of similar surveys, as well as operational considerations and the expected proportion of households with persons age 60 or older in the EAs. The design effect is mostly related to the clustering effect due to the intraclass correlation of households and persons within a cluster for key characteristics.

In order to estimate the level of precision for a key indicator that can be expected from this sample size and the corresponding effective number of completed interviews, we used the estimated ratio corresponding to the survival rate of the population from age 60 to 70 for each district. These results are shown in Table 4.

Table 4. Approximate level of precision for estimates of the survival rate from age 60 to 70 years from the 2017 Malawi Survey of Aging based on initial proposed sample allocation

District	r	deff	Estimated effective sample of individuals age 60+	Estimated standard error	95% confidence interval	
					Lower	Upper
Mzimba	0.575	1.5	737	0.022	0.531	0.619
Lilongwe Rural	0.625	1.5	650	0.023	0.579	0.670
Mangochi	0.609	1.5	615	0.024	0.562	0.657
Nsanje	0.603	1.5	635	0.024	0.556	0.650

The EAs were selected using systematic sampling with probabilities proportional to size. This procedure involved selecting the EAs with the following selection numbers: R*I, R*I+I, R*I+2*I.....R*I(n-1)*I.
where,
R=random number between (0,1)
I= Selection Interval

8.3. Household Listing in Malawi

As discussed in the general guidelines for developing the study sample, an important step before the actual implementation of the survey is to update the household listing in the selected EAs. Since the sampling frame was from the 2008 Census, a complete household listing in the selected EAs was conducted, where the following information was collected for each enumerated household:

- Region
- District
- Cluster number
- Household Number
- First name of each listed individual
- Last name of each listed individual
- Age of each listed individual
- Household Head
- Additional question that is: "Is there anyone in the household aged 60 and older that you provide food to, and he/she is not residing in this dwelling?", with responses "1. Yes", and "2. No".

The household listing starts with the oldest person in the household, followed by the second oldest, etc. This is of particular importance and has the purpose to list first the members of the household who are the population of interest in the present study.

The design of the household listing form is captured in Table 5, including the introductory text to be read by the interviewer. [Interviewer's introductory text: I would like to ask you information about all members of your household. Please start with the oldest person in your household, followed by the second oldest, etc. until we list all members of your household.]

Table 5. Listing form

Line Number	District	Cluster Number	Household Number	First Name	Last Name	Age _____ Years Old Or	Household Head 1. Yes 2. No	Is there anyone in the household aged 60 and older that you provide food to, and he/she is not residing in

						Year born:_____		this dwelling?", 1. Yes 2. No
1.								
2.								
Etc.								

Age of each listed individual should be recorded as age in completed years (age at last birthday). If this is not known, then the interviewer should record the year of birth.

During household listing, all households in the selected cluster/EA will be listed and assigned a number which will be recorded on the listing form. Only households which will be selected for interviews will be visited.

8.4. Selection of the household sample

Once all households in the selected EAs are listed, the information will be computerized; ideally, the household listing will be conducted on tablets; and in this scenario, unique household IDs will be generated automatically at the time of the data collection; alternatively, unique household IDs can be assigned before the beginning of the household listing fieldwork.

The next step will be based on the complete household listing to select the households that have at least 1 age-eligible respondent age 60 years and over. In addition, if there is anyone in the household aged 60 and older to whom the household provides food to, and he/she is not residing in this dwelling of the household (see table with Household Listing), then this household will be also selected as eligible for the household survey. As a result of these selection criteria, a full list with all households with at least one age-eligible household member aged 60 + years will the compiled.

The eligible households with at least one member age 60 years or older were then sequentially numbered from 1 to n (the total number of households in each enumeration area) at the National Statistical Office, where the selection of 12 eligible households in each enumeration area was carried out using random systematic selection procedures.8.5. Weighting Procedures for Malawi Survey on Aging

In order for the sample estimates from the Malawi Survey on Aging to be representative of the household-based population age 60 years and older in each district, it is necessary to multiply the data by a sampling weight, or expansion factor. The basic weight for each sample household is

equal to the inverse of its probability of selection (calculated by multiplying the probabilities at each sampling stage).

As indicated previously, the sample EAs for this survey were selected within each district with PPS from the 2008 Malawi Census frame. The measure of size for each EA was the number of households with at least one person aged 52 years or older in the 2008 Census frame. Following a listing that identified all the eligible households with at least one person aged 60 or older in each sample EA, at the second stage 12 eligible sample households were selected with equal probability from the listing. Therefore the overall probability of selection for the sample eligible households can be expressed as follows:

$$p_{hi} = \frac{n_h \times M_{hi}}{M_h} \times \frac{m_{hi}}{M'_{hi}},$$

where:

p_{hi} = overall sampling probability for eligible households selected for the survey in the i-th sample EA in district h

n_h = 40 = number of sample EAs selected in district h

M_{hi} = total number of households with at least one person aged 52 years or older in the 2008 Malawi Census data for the i-th sample EA in district h

M_h = total number of households with at least one person aged 52 years or older in the 2008 Malawi Census data for district h

m_{hi} = 12 = number of sample eligible households with at least on person aged 60 years or older selected in the i-th sample EA in district h

M'_{hi} = total number of households with at least one person aged 60 years or older listed in the i-th sample EA in district h

The basic weight for the sample households is the inverse of this probability of selection, expressed as follows:

$$W_{hi} = \frac{1}{p_{hi}} = \frac{M_h \times M'_{hi}}{n_h \times M_{hi} \times m_{hi}},$$

where:

W_{hi} = basic weight for the sample eligible households in the i-th sample EA in district h

Following the survey data collection, it was necessary to adjust the basic weights to account for non-interviews, as follows:

$$W'_{hi} = W_{hi} \times \frac{m_{hi}}{m'_{hi}},$$

where:

W'_{hi} = adjusted weight for the sample households in the i-th sample EA in district h

m'_{hi} = number of sample eligible households with completed interviews in the i-th sample EA in district h

A spreadsheet with the sampling frame information for the sample EAs was used to enter the information on the number of eliglble households listed and the number of households with completed questionnaires in each sample EA, as well as the sampling parameters for each stratum. Then the specified formula was used for calculating the weight of the sample households in each sample EA.

8.6. Tabulation of Sampling Errors for Malawi Survey on Aging

In the publication of the results for the Malawi Survey on Aging it is important to include a statement on the accuracy of the survey data. In addition to presenting tables with calculated sampling errors and confidence intervals for the most important survey estimates, the different sources of nonsampling error should be described.

The standard error, or square root of the variance, is used to measure the sampling error, although it may also include a small variable part of the nonsampling error. The variance estimator should take into account the different aspects of the sample design, such as the stratification and clustering. Software packages such as the SPSS Complex Samples module and Stata use a variance estimator that takes into account the sample design. These programs use a linearized Taylor series variance estimator.

In order to measure the level of precision for a key indicator from the final data of the Malawi Survey on Aging, we used the survey estimates of the ratio of the population age 70+ years to the population 60+ years by district. The SPSS Complex Samples module was used to tabulate the standard errors, coefficients of variation, 95% confidence intervals, design effects and number of observations for these estimates. The results of this study are presented in Table 6.

Table 8. Estimated ratio of persons aged 70+ to persons aged 60+ by domain from Malawi Survey of Aging data, with standard errors, coefficients of variation, 95% confidence intervals, design effects and number of observations

Domain	Estimate	SE	CV	95% confidence interval	DEFF	

				Lower	Upper		No. sample individuals aged 60+
Combined 4 districts	0.483	0.015	0.031	0.453	0.513	2.02	2,236
Sex (combined 4 districts)							
Male	0.449	0.019	0.042	0.411	0.486	1.31	902
Female	0.506	0.018	0.035	0.471	0.540	1.67	1,334
District							
Mzimba	0.468	0.024	0.052	0.420	0.515	1.22	583
Lilongwe	0.462	0.025	0.053	0.413	0.510	2.25	563
Mangochi	0.512	0.035	0.069	0.443	0.581	2.99	562
Nsanje	0.533	0.029	0.054	0.477	0.590	0.66	528

8.7. Next Steps for the Implementation of the Survey on Aging in sub-Saharan Africa

After the final random sample of age-eligible households is selected, the teams of interviewers will be sent to the selected households to administer first the "Screening Form for Eligibility" followed by the "Household Questionnaire". If based on the "Screening Form for Eligibility" it is determined that the household does not have persons aged 60+ years, then the household will not be further surveyed with the "Household Questionnaire". In this case, the interviewer proceeds with the next household on his/her Assignment Sheet. No household will be substituted by the interviewers. The likelihood of this scenario happening is probably low, however, it is possible that the person who was interviewed for the household listing misreported some of the ages of the household members. Hence, to avoid that households that do not have age-eligible respondents are interviewed in the "Household Survey" (i.e., with the implication of saving costs), the "Screening Form for Eligibility" will be first administered.

If the "Screening Form for Eligibility" confirms that the household has at least one age-eligible person, then the household questionnaire will be administered to the household head or other most knowledgeable person in the household, followed by the "Individual Questionnaire" administered to all age-eligible members of the household. For details on how these 2 questionnaires will be administered, consult the "Interviewer's Manual".

8.8. Special Instructions for the Administration of the Household Listing, Household Questionnaire and the Individual Questionnaire

Although not part of the sampling procedure, this section provides some additional instructions that are important for the administration of the household listing and the subsequent surveys.

Specifically, it is advised to consider that before the beginning of the fieldwork, the [IA] needs to follow well-established procedures for conducting surveys in Malawi such as obtaining permission from the District's Office, TA and the village headmen to work and conduct surveys in the sampled EAs. These permissions should be obtained in advance, at least 2-3 days before the interviewer teams are sent to the EAs.

References

ICF International (2012). Demographic and Health Survey sampling and household listing manual. Measure DHS, Calverton, Maryland, USA: ICF International, https://dhsprogram.com/pubs/pdf/DHSM4/DHS6_Sampling_Manual_Sept2012_DHSM4.pdf.

Kowal, Paul, et al. "Data resource profile: the World Health Organization Study on Global Ageing and Adult Health (SAGE)." International journal of epidemiology 41.6 (2012): 1639-1649.

Unicef (2016). MICS 5 Sample Size Calculation. Available at http://mics.unicef.org/tools

United Nations Statistics Division (2005). Designing household survey samples: Practical guidelines. New York: United Nations, Department of Economic and Social Affairs, Statistics Division, Studies in Methods, Series F No. 98, http://unstats.un.org/unsd/demographic/sources/surveys/Handbook23June05.pdf.