Statistical Journal of the IAOS 30 (2014) 21–27
DOI 10.3233/SJI-140795
IOS Press

Building a master sampling frame by linking the population and housing census with the agricultural census

Naman Keita and Pietro Gennari*
Food and Agriculture Organization of the United Nations, Rome, Italy

Abstract. The objective of the Global Strategy to Improve Agricultural and Rural Statistics is to enable developing countries to build a sound and comprehensive agricultural statistical system, that is sustainable, well integrated in the overall national statistical system and that is capable of responding to the information needs of the 21st century.
One of the three pillars of the Global Strategy is the integration of agriculture into the national statistical system which is obtained through three methodological tools: (i) an integrated survey framework, that defines the mix and frequency of data collection initiatives needed to meet the predefined data requirements; (ii) an integrated statistical data base, that provides data management system for different data sources (censuses, surveys, administrative data), using common definitions and classifications and ensuring that only one number is agreed as official statistics; (iii) a master sampling frame, that is the basis for selecting the sample of all agriculture-related surveys across the national statistical system, thus allowing coordinated data collections.
Little guidance is currently available on buiding a Master Sample Frame for agricultural surveys in different country contexts. The FAO and UNFPA *Guidelines for Linking Population and Housing Censuses with Agricultura Censuses* aim to fill this gap. This paper draws on the Guidelines and country reports to examine how the agricultural data collected through the Population Census can contribute to building a Master Sample Frame for agriculture censuses and surveys, with illustrations from the recent experience of Mozambique.

Keywords: Master sampling frames, population and housing census, agricultural census, sample design

1. Introduction

The objective of the Global Strategy to Improve Agricultural and Rural Statistics (GS) is to enable developing countries to build a sound and comprehensive agricultural statistical system, that is sustainable, well integrated in the overall national statistical system and that is capable of responding to the information needs of the 21st century.

One of the pillars of the GS is the integration of agriculture into the national statistical system which is obtained through three methodological tools: (i) An integrated survey framework, that defines the mix and frequency of data collection initiatives needed to meet the predefined data requirements; (ii) An integrated statistical data base, that provides data management system for different data sources (censuses, surveys, administrative data), using common definitions and classifications and ensuring that only one number is agreed as official statistics; (iii) A Master Sample Frame (MSF), that is the basis for selecting the sample of all agriculture-related surveys across the national statistical system, thus allowing coordinated data collections.

The MSF is indeed a powerful tool for data integration. It is one of the main tools for establishing closer links between results from different statistical processes and statistical units. In the context of the GS, the MSF ensures that information on three basic statistical units – land parcel, household and farm –

*Corresponding author: Pietro Gennari, Food and Agriculture Organization of the United Nations (FAO), Rome, Italy. E-mail: pietro.gennari@fao.org.

are interlinked, thus allowing to simultaneously provide consistent and integrated statistics on the environmental, social and economic dimensions of agriculture. The MSF provides the basis for selection of probability based samples of farms and households. In addition to its technical role for integration and simplification of sampling processes, a MSF can be cost efficient since the cost of its construction can be spread over several surveys instead of building ad-hoc frames for each new survey.

Little guidance is currently available on buiding a MSF for agricultural surveys in different country contexts. The GS aims at filling this gap with the development of guidelines and tools that are tailored to the specific situation of each country, taking into account both the structural characteristics of the agricultural sector and the level of development of the national statistical system.

Depending on country capacity and circumstances, the GS, proposes five different approaches for establishing a MSF:

a) List frame based on the population census;
b) List frame based on the agricultural census;
c) List frame based on the business register of farms;
d) Area frame (based on remote sensing; aerial photos; etc.);
e) Mixed list and area frame (Multiple frame approach).

Detailed instructions and practical examples that help countries to adopt the first approach listed above are contained in the *Guidelines for Linking Population and Housing Censuses with Agricultural Censuses* ("Guidelines") which have been recently published jointly by FAO and UNFPA (FAO, 2012).

This paper draws on the Guidelines and on a database of country reports to examine how agricultural data collected through Population and Housing Censuses (PHC) can contribute to build a MSF for agriculture censuses and surveys (AC), with illustrations from the recent experience of Mozambique.

2. Building a MSF for agricultural censuses and surveys on the basis of a PHC

2.1. Definition of master sampling frame

A sampling frame must cover the entire survey population exhaustively and without overlaps. It should provide a list of statistical units from which the sample is selected. Since multistage sampling is the most commonly used design for household and agricultural surveys, a sampling frame providing a list of all statistical units with their characteristics is needed for each stage of the sample selection. Therefore, the development of a sampling frame should be considered together with sample design and survey methodology.

The sampling units used at the first stage (primary sampling units or PSUs) are area units that can be administrative subdivisions, like districts/villages, or areas demarcated for conducting a PHC, like census enumeration areas (EAs). The second stage consists of a sample of secondary sampling units (SSUs) selected within the selected PSUs.

A MSF is a list of area units that covers the whole country and that contains information on a broad range of key characteristics of the unit, including demarcation of the boundaries as well as identification of higher-level units. The specific feature of a MSF is that makes it possible to draw samples for several different surveys or different rounds of the same survey, as opposed to building an ad-hoc frame for each survey. In the context of the Global Strategy, the MSF is a tool that combines information on the socio-economic characteristics of the household and on the agricultural characteristics of the holding, including information on land area. The MSF should therefore allow the selection of samples for both household based surveys and holding based surveys.

2.2. Collection of agricutural data in the PHC

As mentioned in the introduction, one of the approaches proposed in the Guidelines to build a MSF for agricultural censuses and surveys (AC) is the establishment of a list frame on the basis of a recent PHC. In particular, the Guidelines recommend gathering a limited number of agricultural data during the Census operations, either in the listing phase or in the proper data collection phase. When relevant agricultural data is collected at household level during the PHC, *farm or agricultural households*[1] can be identified with their key agricultural characteristics.

The Guidelines are rather flexible regarding the agricultural data to be collected during the PHC as they recommend to introduce in the census questionnaire a variable number of questions according to the country

[1] Defined as a household where one or more members are engaged in own account agricultural production.

situation: from a minimum set of questions to identify farm households to a more detailed agricultural module covering most of the items recommended by the WCA 2010[2] core module, where this is relevant and feasible.

The recommended core set of questions includes information used to identify the farm households and to measure the size of their holding. More precisely, the following core data items are collected:

Item A1. Whether the household is engaged in any form of own-account agricultural production (including livestock, fishery, aquaculture or forestry). This question allows the identification of the farm household.

Item A2. The area of land (or number of plots) used for agricultural purposes. This questions can be used to improve the sample design and the efficiency of the estimates.

2.3. Use of agricultural data collected in the PHC to build the MSF

As indicated above, a MSF is basically a list of units that covers the whole population with no omission or overlaps and that can be used to draw samples for different types of surveys. The agricultural data collected in the PHC can be used to derive two types of MSF: a list of all farm households, or as list of enumeration areas with information on the number of farm households.

During the PHC, the country is divided into Enumeration Areas (EA) in which complete enumeration of all households is conducted. A considerable amount of preparatory activities is dedicated to preparing EA maps with precise boundaries using handheld GPS devices and these maps are now digitized in many countries. GPS can also be used to geo-reference the Households.

From data collected during the PHC, a complete list of all households and their characteristics can be made available. When relevant agricultural data is collected and processed, as explained above, a complete list *of all farm households* (with their geo-referencing coordinates if included in PHC questionnaire) will be available. Given the complete coverage of all households during PHC, farm households located both in rural and urban areas will be identified by inquiring whether they are engaged or not in own-account agricultural activities.

This list of all farm households can be used directly as MSF for AC covering the holdings in the household sector. Depending on the agricultural questions included in the PHC, list of specific types of farm households can be established and used efficiently for targeted surveys. For example, farm households with livestock (for a livestock surveys) or farm households that grow rice (for a rice production survey). This type of MSF is most effective when the agricultural census or surveys are conducted jointly with the PHC (as in the case of Canada), or soon after, as it becomes quickly outdated.

In most countries, where the AC is not jointly conducted with the PHC and where *a multistage survey design* is used, a MSF as list of enumeration areas should be considered. In this case, the list of all EAs with their associated agricultural data (number of farm households and their size) can be used to build a MSF for selecting the PSUs. Sometime, the EAs from PHC may need to be combined in one unit when the number of farm households is too small and some EAs may be deleted if there is no farm household. This list can then be used for selecting samples for all upcoming AC.

When two stage sample designs are used, random sample of EAs are selected and screened to obtain an updated list of agricultural holdings (second stage sampling frame) to be used for selecting ultimately the sample of farm households for AC.

2.4. Benefits of collecting agricultural data during a PHC

When relevant agricultural data is collected during a PHC, the survey design can be enhanced (in terms of stratification, sample size and selection, sample allocation) and consequently data quality can also be improved significantly. Agricultural data, in particular, can be used to better define the target population (by selecting an appropriate threshold), to improve the stratification at first and second stage and to better delineate the spatial distribution of the population of various units in a cluster design. Examples of data items used to define strata includes scale of operation of holdings, based on size of land being operated, or type of agricultural activities, based on items such as number of each species of livestock, or area under specific crop of national importance. More relevant and accurate measure of size can be derived in probability proportional to size sampling and calculation of variability and other parameters.

[2]FAO World Programme for Census of Agriculture 2010.

Collecting basic agricultural data during the PHC means that the same infrastructure, logistics, personnel and equipment can be used to construct a frame that can be used by subsequent AC. Some countries have experienced up to 50 percent reduction in the cost of the agricultural census by including basic questions in the PHC.

Moreover, collecting agricultural data during a PHC allows to cover farm households located in urban areas (normally not covered by the AC) or concentrated in specific districts or engaged in very specific agricultural activities. By collecting data on both socio-economic variables and agricultural variables, cross-tabulations and more in-depth analysis can be performed, contributing to more integrated data in line with pillar 2 of the Global Strategy.

2.5. Where this approach can be used

As specified in the Guidelines, the MSF derived from PHC will only cover household based agricultural holdings and production. Therefore, this approach will be most relevant in countries where agricultural production is mainly based on the household sector which is the case in many developing countries with a large subsistence farming sector. The MSF obtained through this approach will need to be complemented by a separate frame made of the list of corporations and other non-household based holdings in order to have a complete coverage of the agriculture sector.

3. Main methodological issues

3.1. Linking holdings to households

One of the main methodological issues is the conceptual difference between the statistical unit of the PHC (household) and the AC (agricultural holding). The WCA 2010 defines two types of agricultural holdings: (i) holdings in the household sector and (ii) holdings in the non-household sector, such as corporations and government institutions.

In developing countries, most agricultural holdings belong to the household sector and, in the majority of cases, a one to one correspondence can be established between agricultural holdings, farm households and households. In these cases, it is therefore possible to unambiguously identify farm households in the PHC. When this correspondence is not one to one (i.e. one household manages more than one farm or a farm is managed by more than one household) there is a risk of a coverage error in the MSF. To address this issue the Guidelines provide detailed recommendations on the identification of farm holdings.

3.2. Coverage

The exhaustive coverage of all farm households in the PHC will crucially depend on the quality of the field operations (questionnaire compilation and supervisor control). Since agriculture is not the main focus of the PHC, there is a real risk of under-coverage of farms, as for various reasons, not all households declare their agricultural activity. When multi-stage surveys are conducted immediately after the PHC, a new complete listing of all farm households in selected PSUs could give an indication of the level of potential under-coverage.

Another reason for the incomplete coverage of agriculture farms in the MSF is due to the fact that the PHC only includes holdings of the household sector. Therefore an additional list of holdings in the non-household sector must be established using information from government regulatory agencies, producers' associations, telephone directories, or other administrative sources.

3.3. Updating the MSF

The MSF derived from the PHC can become rapidly out-dated and a growing coverage problem will emerge as the time separating the PHC and agricultural surveys increases. Unless there are effective mechanisms in place for updating and maintaining the register, it can quickly become irrelevant.

Updating the sampling frame could be considered at different sampling stages. For the first stage, what is required is an updated list of all EAs in the country. In past PHCs, a heavy and costly cartography exercise had to be undertaken prior to each round of the PHC to update maps of the EA. In recent years, many countries have shifted to preparing geo-referenced and digitised EA maps (using extensively GPS) as part of the PHC process. A database of all EAs in the country will therefore be available with agricultural related data collected during the PHC.

The availability of geo-referenced and digitised EA maps will facilitate the maintenance and updating of the EAs maps. In fact, this information can be combined with satellite images (with land cover and use information) to build an area frame that is much easier to update.

Updating of sampling frames for the second stage can be done by using rotating sample selection of PSUs and performing a complete enumeration of selected PSUs.

4. Country experience

An analysis of the questionnaires used in the PHC conducted as part of the 2010 round shows that almost 50 countries collected some agricultural information. Only a few countries, however, have used this information to build a MSF for AC. Mozambique is one of these few countries. In this section the experience of Mozambique in building a MSF using data from integrated population and housing and agricultural censuses is discussed.

4.1. Mozambique[3]

The 2009/11 Mozambique Census of Agriculture and Livestock (CAP II) was timed to follow the 2007 Mozambique PHC, which provided an effective area sampling frame for CAP II. The census enumeration areas (EAs) were defined as the primary sampling units (PSUs) for the area frame, and a stratified two-stage sample design was used for CAP II. The area frame developed from the PHC was also supplemented by a list frame of large farms. All the large farms in this list frame were matched to those enumerated in CAP II, in order to avoid any duplication.

The 2007 PHC questionnaire included a limited number of questions to identify households that operate farms, and to classify the farm holdings by size. In particular, the following questions were included:

- G1: Does any member of the household practice agriculture? Yes/No
- G2: Does the household have any tanks for aquaculture? Yes/No. If yes, how many?
- G3: Does any member of the household practice traditional fishing? Yes/No
- G4: Does this household have cashew nut trees? Yes/No. If yes how many?
- G5: Does this household have coconut trees? Yes/No. If yes, how many?
- G6: Types and numbers of animals on the holding.(G 6.1 Cattle; G 6.2 Goats; G 6.3 Sheep; G 6.4 Pigs; G 6.5 Chickens; G 6.6 Ducks)

[3]This text is mainly extracted from [INE Mozambique 2012] and adapted for the paper.

The PHC data enabled the farm holdings to be classified by size based on total farmland, cultivated area and number of livestock (see Table 1) and provided the number of agricultural households in each EA. This information was used to improve the efficiency of the sample design by providing the measure of size for sampling the EAs using probability proportional to size (PPS) within each district.

It was also necessary to establish the minimum cut-off criteria for identifying agricultural households. Only households above the cut off were considered in the agriculture census. The necessary data on size was provided by the PHC. The total number of EAs in the final sampling frame for CAP II was 44,859: 35,333 rural EAs and 9,526 urban EAs.

The data from the PHC was primarily used in the first stage sample selection of PSUs. The second stage sampling was carried out during the agriculture census, with a new listing of all of the households and farm holdings in each sample EA selected at the first sampling stage.

A census map was produced for each sample EA that identified the EA boundaries, and Global Positioning System (GPS) devices were also used during the listing stage to ensure complete coverage of the households in each EA. The listing sheet included information on the name of the head of household and the address, as well as information on the total cultivated and irrigated area, the number of cashew, coconut and fruit trees, and the number of animals by type. This information was used for a second stage stratification of the farm holdings by size.

All of the farms identified in the listing as medium or large were included in the sample with certainty at the second stage, and a sample of small farms was selected with equal probability. It was necessary to match the large farm holdings identified in the listing to the list frame of large farms in order to avoid any duplication.

Another important consideration is the number of households to be selected in each PSU (EAs in this case). Based on the experience from the first agricultural and livestock census (CAP I) and the annual agricultural surveys, 10 small farm holdings were selected per sample EA for CAP II. This is consistent with the optimum number of sample households per cluster used for similar agricultural censuses and surveys conducted in other countries.

The total number of sample EAs for CAP II was 3,502 EAs. Different alternatives were considered for allocating the sample districts. For example, proportional allocation could have improved the efficiency

Table 1
Classification of farm holdings by size for CAP II

Class of holding	Land area	Cultivated area	Livestock
Large holdings	100 hectares	OR 50 ha. of cultivated land including both annual and permanent crops	OR 100 cattle OR 500 small ruminants and pigs OR 2000 poultry
Medium holdings	25 hectares	OR 10 ha. of cultivated land including both annual crops and organized plantations OR 5 ha. irrigated land with annual crops	OR 10 cattle OR 50 small ruminants and pigs OR 200 poultry
Small holdings	Less than 25 hectares	AND less than 10 ha. annual cultivation AND less than 5 ha. irrigated	AND less than 10 cattle AND less than 50 small ruminants and pigs AND less than 200 poultry

Source: INE, Mozambique 2012, figure 3.1.

of the sample design for national-level estimates, although the sample allocated to the smaller districts may not have been sufficient for obtaining reliable estimates. Alternatively, equal sample allocation could have improved the level of precision for the smaller districts, but would have been less efficient for estimates at the provincial and national levels. As a compromise it was decided to use power allocation (where the sample is allocated in proportion to x^{λ}, and λ is a value between 0 and 1) and it was decided to set λ equal to 0.4. This resulted in a minimum allocation of 10 sample EAs to the smallest districts, and a maximum sample of 46 EAs to the largest district.

The total number of small farm holdings in the CAP II sample was 38,217. The total number of medium size farm holdings in the sample was 2,449. Following the selection of sample EAs, it was found that some of the EAs in the 2007 Census frame had an incorrect urban/rural code, so these codes were updated in the CAP II data. The 833 large farms, identified in the list frame, were excluded from the area frame. However, if additional large farms were found in the sample EAs that were not in the list frame, they were included with certainty at the second sampling stage, and received the same weights as the medium size farm holdings.

The approach described was successfully used in Mozambique together with the introduction of CAPI (Computer Assisted Personal Interview) and GPS technology, which was the first time this had been done in an agricultural census in Africa. This resulted in the more timely release of census data (about 6 months after data collection) and a good level of accuracy for the data on major crops at national and provincial level, as indicated by the CVs for corn area presented in the annex.

5. Conclusions

Experience shows that, in many countries, linking the PHC with the AC can result in many advantages for the national statistical system. It can significantly contribute to the implementation of pillar 2 of the *Global Strategy to Improve Agricultural and Rural Statistics*, which is to integrate agricultural statistics into the national statistical system through the development of an effective MSF. Collecting limited and well-defined agricultural data during the PHC can substantially contribute to building an efficient MSF for agricultural censuses and surveys in many developing countries. This can result in substantial cost saving and in improving data quality through enhanced survey design, as illustrated by the experience in Mozambique.

However, the approach is more relevant to countries where agricultural production is mainly based on the household sector since non-household based holdings will not be covered by frames derived from the PHC. Furthermore, issues of misclassification/under reporting during the PHC may result in the under estimation of farm households. Moreover, updating and maintaining the MSF also remains a challenge, even if the use of rotating the sample selection of PSUs can contribute to keeping the MSF of EAs closer to the real population.

References

[1] Food and Agriculture Organization of the United Nations: *A System of Integrated Agricultural Censuses and Surveys: World Programme for the Census of Agriculture 2010*, FAO, Rome, 2005.

[2] Food and Agriculture Organization of the United Nations and United Nations Population Fund: *Guidelines for Linking Population and Housing Censuses with Agricultural Censuses: with selected country practices*, FAO, Rome, 2012.

[3] Instituto Nacional de Estatistica/Ministerio da Agricultura e Desenvolvimento Rural: *Mozambique Census of Agriculture and Livestock 2009–2011: Census Methodology*, Maputo, 2012.

[4] United Nations Economic Commission for Africa: *The 2010 Africa round of population and housing censuses. Draft implementation handbook: A guide for African countries*. UNECA, Addis Ababa, 2008.

[5] United Nations Economic Commission for Europe: *Conference of European Statisticians Recommendations for the 2010 censuses of population and housing.* Report no. ECE/CES/STAT/NONE/2006/4. United Nations, New York and Geneva, 2006.
[6] United Nations Statistics Division: *Household Sample Surveys in Developing and Transition Countries* ST/ESA/STAT/SER.F/96, United Nations, New York, 2005.
[7] United Nations Statistics Division: *Principles and recommendations for population and housing censuses, Revision 2*. Statistical Papers Series M, No. 67/Rev.2. UNSD, New York, 2008.
[8] World Bank and Food and Agriculture Organization of the United Nations: *Global Strategy to Improve Agricultural and Rural Statistics*. Report no 56719-GLB. World Bank, Washington DC, 2011.

Annex: Total area planted in corn by Urban/Rural areas and by province

Crop	Estimate	SE	CV(%)	95% confidence interval		DEFF	No. farm holdings
				Lower	Upper	in the sample	
Mozambique	1,411,453	23,566	1.7%	1,365,248	1,457,659	9.6	37,286
Residence							
Urban	156,892	6,248	4.0%	144,641	169,143	3.0	4,608
Rural	1,254,561	22,722	1.8%	1,210,010	1,299,113	9.3	32,678
Province							
Niassa	148,555	7,672	5.2%	133,513	163,596	4.0	3,279
Cabo delgado	97,596	4,563	4.7%	88,649	106,542	4.3	3,230
Nampula	129,501	5,750	4.4%	118,227	140,775	5.4	3,525
Zambézia	193,905	9,813	5.1%	174,664	213,146	10.0	3,867
Tete	236,431	9,770	4.1%	217,275	255,588	4.6	4,620
Manica	208,273	11,790	5.7%	185,155	231,390	7.4	3,366
Sofala	141,921	5,405	3.8%	131,323	152,518	2.8	3,676
Inhambane	85,153	4,587	5.4%	76,160	94,147	4.0	4,268
Gaza	115,172	6,836	5.9%	101,768	128,576	5.9	4,414
Maputo provincia	47,660	3,774	7.9%	40,260	55,059	2.8	2,477
Maputo cidade	7,287	782	10.7%	5,754	8,820	1.1	564