# GHANA STATISTIC SERVICE – FAO – USDA/NASS

## Regional Workshop on
## Sampling for Census of Agriculture 2010 and Agricultural Surveys

***Training Materials***

*Accra, Ghana*
19 – 23 April 2010

# TABLE OF CONTENTS

# SESSION 1

## AGRICULTURAL CENSUS IN THE NATIONAL SURVEY PROGRAMME

# Global Strategy for Improving Agricultural Statistics: Integration of the WCA2010 in the National Survey Programs and the National Strategy of Statistical Development

*N. Keita, FAO*

This text is mainly extracted from the Global Strategy to Improve Agricultural Statistics adopted by the 41st Session of the United Nations Statistical Commission in February 2010 and from the FAO World Programme of Census of Agriculture 2010. The text was adapted to the purpose of this workshop.

The main idea is that, in order to address a wide variety of data users' needs (existing basic needs and emerging needs) in a cost-effective manner, agricultural censuses and surveys should not be designed and implemented as ad-hoc operations. They should be part of an integrated agricultural statistics programme which should be also well integrated into the National Statistical System through the National Strategy for Development of Statistics (NSDS).

The main emphasis is on technical tools and methodologies which can facilitate this integration such as linking population and agricultural censuses, building a centralised master sampling frame, designing an integrated survey framework and an integrated database.

The new integrated and modular approach to agricultural census and surveys in this context is an important element of the statistics programme and sampling techniques become essential for proper implementation of this approach.

## 1. Importance of reliable data on agriculture for policy makers

Policy makers and development practitioners who are responsible for developing investment strategies to promote economic growth find many challenges in the changing face of agriculture in the 21st century. In addition to its productive role providing the food, clothing, fuel, and housing for a growing world population, agriculture assumes other roles, the importance of which has more recently been recognized. In addition to its essential role in food security, agricultural development is now seen as a vital and high-impact source of poverty reduction. It is also seen as a source of serious environmental problems and a major contributor to global warming, water scarcity and pollution, and land degradation. At the same time its potential as a source of environmental services needs to be defined, monitored, and evaluated.

In order to support effective policies aimed at addressing this wide range of basic and emerging issues of the 21st century, there is a growing need of relevant, reliable and timely data produced on a sustainable basis.

However, various assessment of the current status of agricultural statistics in many developing countries, and particularly in African countries reveals that they do not have the capacity to respond to the basic needs and emerging needs. The situation is characterised by:

- Countries' capacity in agricultural statistics have significantly declined since early '80s, mainly because of reduction of resources allocated to agricultural statistics by countries and development partners.
- Basic data requirements are not met and a declining number of countries are reporting basic agricultural statistics
- Emerging data needs linking the economic, environmental and social dimensions (impact of agriculture on environment; livelihood of rural household; biofuels; water and land use, etc.) are not met
- Lack of coordination between National Statistical Offices and Ministries of Agriculture which often results in duplication of efforts and conflicting numbers
- Agriculture often left out of National Strategy for Development of Statistics
- Forestry, fisheries outside of national and agricultural systems

As a consequence, there is a lack of sound basis for agricultural development and food security policy formulation, implementation, monitoring and evaluation in many African countries at a time when this data is most needed.

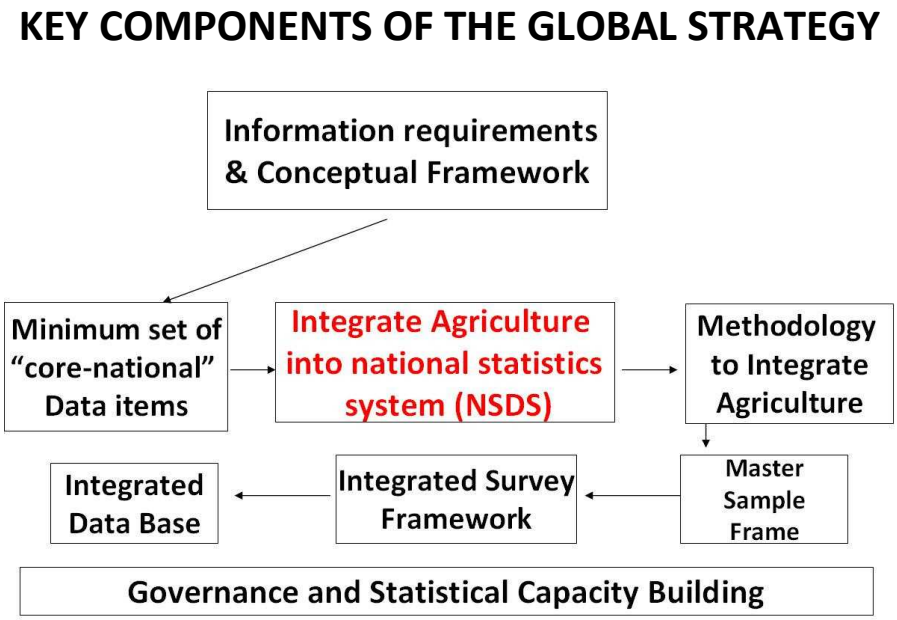
## 2. Global Strategy to Improve Agricultural Statistics

The Global Strategy is an initiative undertaken at international level under the auspices of the United Nations Statistical Commission (composed of the Directors of National Statistics Offices-NSOs of all countries in the World) to address the issues related to declining capacity in agricultural statistics' mainly in developing countries.

The purpose of the Global Strategy is to provide a framework for national and international statistical systems that enables them to produce and to apply the basic data and information needed to guide decision making in the 21st century. This Strategy is based on three pillars.

- The first is the establishment of a minimum set of core data that countries will provide to meet current and emerging demands.
- The second is the integration of agriculture into national statistical systems in order to satisfy the demands of policy makers and other users who rely on comparable data across locations and over time. The integration will be achieved by implementing a set of methodologies that includes the development of a Master Sample Frame for agriculture, the implementation of an integrated survey framework, and with results available in a data management system.
- The third pillar is the foundation that will provide the sustainability of the agricultural statistics system through governance and statistical capacity building.

The Strategy can be summarised in the following figure:




**3. Determining priorities for core national data item**

The Global Strategy identifies a minimum core data items that each country is expected to produce (see table in annex A).

However, data for some core items will not be required every year either because they do not change much from year to year or because they are difficult and expensive to obtain annually. Countries will also have additional items to add to the list of core items to meet national data needs. *Teff* for example is a major crop and food source in Eritrea and Ethiopia, but not in other countries. Items such as rice on the other hand are major global food sources, but are not produced in every country.

Each country therefore needs to select which core items to include in its national system. It must add other items relevant to its economy, determine how frequently data will be provided, and the scope of the national coverage required. For example, the core data do not include fruits and vegetables, or other livestock items that contribute to a country's food supplies and household income. Each country should consider how these should be included in its national system.

Annual data are generally required for those items which, combined, account for more than 75% of a country's value of production. Items with production that can vary significantly from year to year should be included, particularly if the production fluctuations are a major source of risk for vulnerable households and food supplies. Items that account for a significant proportion of land used, and that have short term effects on land use and the environment should be represented as well. Including items that are produced by only a small number of households or holdings or that account for only a small share of the

country's land has sample design and resource implications. For example, sampling theory shows that the relative variance of the estimated mean is approximated by the relative variance of the positive sample units plus the relative variance of the estimated proportion of positive population units.

$CV^2 (Y) = CV^2 (Y_p) + CV^2 (P)$ where $Y_p$ is the mean of the positive responses and P is the proportion of the population that has the item

Assuming that only a third of the households or holdings have a particular item, the sample size will have to be 4 times larger than if 3/4 have the item in order to achieve the same level of precision. If only 10 percent of the households or holdings have the item, then sample sizes triple over what is needed if a third have the item and would be 12 times greater than if (P) > .75 for the same level of precision.

| Proportion of population having the item | Sample size required for same level of precision |
|---|---|
| 75% | n |
| 30% | 4*n |
| 10% | 12*n |

**The general conclusion of this exercise is that minor and relatively rare commodities should be confined to the 5 to 10 year agricultural census, and omitted from more frequent surveys**. The exception would be if the sample frame contains sufficient data that can be used in the survey design to target the rare items.

The next step is to review the rural development indicators for monitoring and development and include those relevant to the national situation (see table in annex B). Then each country should determine the level of geographic coverage and detail to be provided for the core plus additional items added. The same issue raised above about the proportion of households or holdings that have the item will also determine the level of geographic detail or other breakdown that can be provided from the sample surveys. These have implications about the methodology to be used and resources required. The annual collections of data will rely upon sample surveys which will limit the geographic detail that can be provided. **Therefore, it may be only through an agricultural census that detailed geographic or size distribution data can be provided.**

The question of what level of detail is required and how often data are required may be difficult to answer. The table below presents a decision matrix that is useful in many contexts. For example, it should be determined for each item whether the data will be provided for the entire country or only major producing areas.

It is generally true that policy makers will want data for within country administrative areas such as provinces; if so, this should be included in the national framework.

Frequency of coverage by geographic and structural detail:

| | Level of geographic and structural detail | | | |
|---|---|---|---|---|
| **Data Item** | **Major production areas only-production by holdings** | **National coverage of production by holdings** | **Within country administrative areas—production by holdings** | **Inclusive of households and HH plots** |
| **Crop A** | **Annual** | **Annual** | **Decennial census** | **Decennial census** |
| **Crop B** | **Bi annual** | **Bi annual** | **Decennial census** | **Decennial census** |
| **Crop C** | **Decennial** | | | |
| **----** | | | | |
| **Crop Z** | | | | |
| **Livestock A** | | | | |
| **Livestock B** | | | | |
| **-----** | | | | |
| **Livestock Y** | | | | |
| **Aquaculture and Fishery** | | | | |
| **Forestry** | | | | |
| **Inputs** | | | | |
| **Household income** | | | | |
| **Change in Land cover** | | | | |

Time and available resources result in a necessary compromise between frequency, level of geographic detail, and other breakdowns.
These categories need to be considered for each data item.

At this stage, each country should have an overall picture of the content of its national statistical system for agriculture including the rural, forestry, and fishery components and the coverage and frequency of the data provided. Input from policy makers and other data users should shape this final picture.

**4. Importance of an integrated statistical system for addressing basic and emerging data needs**

Integration is a core element of the Global Strategy. The Strategy considers that the process of improving agricultural statistics will begin with the integration of agriculture into the national statistical system. This integration will be accomplished by the development of a master sample frame for agriculture to ensure relevance and completeness, its use in implementing a coordinated data collection program to produce timely and accurate data that are coherent and comparable, and a strategy for data dissemination to ensure accessibility. This integration of agriculture into the national statistical system is needed for several reasons.

One of the shortcomings of current statistical systems in most countries is that data are collected by sector, using different sampling frames and surveys. The division of data by sector leaves no opportunities to measure the impact of an action in one sector on another. Surveys are often conducted on an ad-hoc basis with no linkages to a master sampling frame or the use of geo-referenced units for data collection. It is therefore difficult to integrate data coming from various surveys for in-depth analysis with cross tabulation of variables. Data on crop and livestock production are drawn from separate surveys, which are based on separate samples. The separate data provide no basis for analyzing the characteristics of farms that produce both crops and livestock, or for comparing them to farms that specialize in one or the other. Household surveys are often conducted in isolation from production surveys with no coordination or with sample sizes too small to disaggregate the data into the rural/farm sectors. The results generated from these surveys are also not integrated into a common database for access by data users.

More than one governmental organization is often involved in the collection and analysis of agricultural, fishery and forestry data without coordination. While the National Statistical Office may produce the agricultural census, the annual production data could come from the ministry of agriculture, and the contribution of the fishery and aquaculture sectors may come from another authority and may be ignored or neglected by the National Statistical Office. In some cases, different organizations produce statistics for the same items, with different results which confuse the data users and make it difficult to aggregate results across countries. This means that results then differ also at the international level if those organizations use different sources to populate their data bases.

Integrated statistical systems can resolve many of these problems by avoiding duplications of effort, preventing the release of conflicting statistics, and ensuring the best use of resources. Concepts, definitions, and classifications become standardized, allowing more systematic collection of data across sources. These practical advantages of integrated data systems together with the increasing need for reliable and comparable data in a context of globalization and international concern about environmental issues point to the need for integrated national statistical systems. The FAO World Program for the Census of Agriculture argues forcefully for the development of such integrated systems.

In some countries, centralized organizational structures are already in place, and national statistical offices maintain the principal responsibility for agricultural statistics. However, this centralized role may not always meet the needs of the line ministries such as the ministry of agriculture. For that reason, the statistical responsibilities in many countries are decentralized with ministries of agriculture producing the agricultural statistics. Both systems have advantages and disadvantages. National statistical offices have experience with statistical methodology and sample frames that other ministries do not have. However, the other ministries have more knowledge about agriculture, forestry, fisheries and land use. The purpose of the Global Strategy is to propose a framework for integration that builds off the strengths of both systems.

The integration of agriculture into the national statistical system will be based on statistical methodology using tools that establish a closer link between results from different statistical processes and different statistical units. This can be achieved by the development of a

master sampling frame, the adoption of sample designs such as overlapping samples, and the synchronization of questionnaire designs and surveys.

The master sample, sample designs, and the survey framework need to be considered together because there are choices such as whether to monitor the same farms and households or whether to use different samples, and collect some of the same variables across surveys. It is also necessary that countries have some flexibility in how the master sample frame and resulting survey designs are implemented to consider their national requirements as well as statistical capabilities.

The statistical methodology to be used also needs to consider some basic data quality dimensions—timeliness, completeness, comparability, and accuracy. Measures for each quality dimension will be considered in the development of the strategy. The following sections provide the strategy to create a master sample frame followed by the sample and survey frameworks to achieve the integration.

The Strategy also builds on recent developments in agriculture statistics including the use of satellite imagery for monitoring land use, estimation of crop areas, and providing early warnings of changing growing conditions to name a few examples. In addition, the development of global position systems (GPS) makes it possible to geo- reference observations and data collection to the land cover provided by the satellite imagery. The emergence of the internet and other technology such as the use of personal digital assistants (PDA) equipped with GPS systems for data collection and their connection to databases has tremendous potential for shortening the period between data collection and dissemination with improved data quality.

## 5. Tools and methodology for integration

### *Master sampling frame*

As indicated above, at technical level, the Strategy identifies the development of a master sample frame for agriculture as a starting point for integration. This will be the foundation for all data collection based on sample surveys or censuses. The master sample frame is to be constructed based on the requirements to include both households and farms as statistical units. It provides a linkage between the census framework and land use.

Examples of Sample Frames used for Agricultural Statistics include:

*Population census enumeration areas.* The population census is usually conducted using an administrative structure where cartographic or other mapping materials are used to divide the country into enumeration areas which is the first level of data aggregation. Depending on the country's capabilities, the only results from the population census in some countries are the enumeration area totals for numbers of people, households, etc. Therefore, the sampling frame is basically the listing of enumeration areas and associated aggregated data from the census. Random samples of enumeration areas are selected and screened for households from which subsamples are selected for household surveys--a two stage sampling process. Some countries use their administrative structure of counties, townships, and villages as their framework for the census with the village becoming the enumeration

area. Villages are also used as a first stage sampling unit in countries where the village is where the farm households are generally located.

*Household registers from the population census.* Countries with the statistical capacity are able to develop a register of all households included in the population census. The list of population households is the sample frame used for household surveys. One problem is that the list of households becomes out of date with households changing or dissolving and new households formed. Unless administrative data or other means are used to keep the population register up-to-date, survey results contain an increasing coverage bias over time.

*Agricultural census enumeration areas.* In many countries, the cartographic materials and data from the population census are used for the agricultural census. The sampling frame consists of enumeration areas and aggregated data from the census data collection. As in (a) above, random samples of enumeration areas are selected and screened for farms or agricultural holdings for agricultural production surveys.

*Registers of farms from the Agricultural census.* As in (b) above, countries with the capacity use the agricultural census to develop registers of farms. This provides a powerful sampling tool because it allows a choice of many alternative sampling designs. A major weakness is that the registers rapidly become out of date. Out of date population and farm registers erode all of the data quality dimensions because the completeness of coverage changes over time, thus affecting the comparability and accuracy of the resulting estimates.

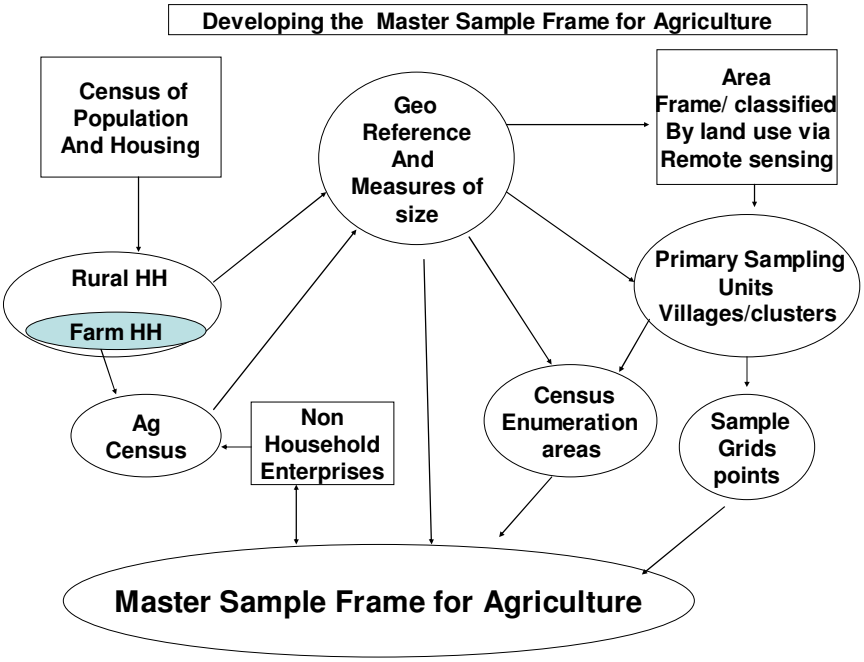
*Registers of farms based on administrative sources such as business registrations or tax collections.* This process is used in some developed countries. It offers the advantages of the registers from the agricultural census, but again, needs to be updated regularly. A disadvantage of the administrative sources is that they may not include the total population, especially units below a threshold required to be registered or pay taxes. In other words, while they will be inclusive of commercial farms, they are not likely to include small scale farms, and subsistence farming units.

*Area Sample Frames.* An area sample frame is the land mass of the country or the space within a country containing the populations of interest. Both maps and satellite images are used to divide the country into administrative areas such as provinces districts, etc. Satellite imagery can be used to subdivide the administrative areas into land use categories such as cropland, rangeland, woodlands, urban areas, etc. Sampling units of segments of land with identifiable boundaries can be formed, or each land use stratum can be divided into square grids and a sample of points becoming the sampling units. During the data collection process, rules of association are used to connect farm holdings or households to the segments or points. An area frame is suitable for obtaining information about variables associated with land such as crops, livestock, forests, and water. Depending on the process used, area frames can be costly and time consuming to construct. However, recent innovations using satellite imagery and two-stage sampling of points have reduced both the cost and time. An advantage of an area frame is that the frame does not go out of date, it is complete in its coverage, and provides a basis to geo reference survey data with the underlying land use. It also provides ground truth useful for classifying satellite imagery by land cover. The primary disadvantage of area frames is that the sampling is based on land

use and not the size and type of agricultural holding. Sampling variability becomes a problem if there is a large range in size of the agricultural holdings. A summary of the methodology of area frame sampling is provided by Gallego. Another disadvantage is that data collection costs exceed those based on registers where telephone or mail can be used instead of personal interviews.

*Multiple Frames.* A combination of the above frames is used, often involving the use of an area frame in conjunction with one of the list frames to take advantage of the strengths and weaknesses of each. The FAO provides an overview of multiple frame sampling. This is an appropriate where there is a large variation in the sizes and types of agricultural holdings with a subset of large commercial farms. The list of commercial farms can be stratified by size and type, and the area frame ensures the population is completely covered by providing coverage of the small and subsistence farms.

The concept of building a master frame can be summarised as follow:




***Integrated survey framework***

The Global Strategy advocates for an integrated survey framework that will be established to provide data measured consistently across time and comparable across countries using an annual survey of selected core items and periodic data from a set of rotating panels covering economic and environmental issues.

All data collection is to be based on sample units selected from the master sample frame, and integrated into the survey framework.

The timing and frequency of data collection are major issues for agricultural statistics. Crops have different production cycles that are seasonal while livestock production is determined not only by the respective reproductive cycles, but also the continuous production of

commodities such as milk and eggs. Aquaculture has characteristics similar to livestock production. The rural labor force is also affected by the seasonal nature of agriculture, which affect opportunities for work and earnings. The timing of data collection affects the quality of the data, especially if a lengthy recall is required. As a result data collection should coincide with harvest periods. For example, if crop yields are determined by crop cutting surveys, then these have to be measured shortly before harvest. Fish capture requires frequent sampling and surveys—for instance twice a week or once every five days—in order for the data to reflect developments such as frequent and unpredictable changes in species composition.

The classical methodology is to select independent samples and conduct separate surveys for each of the categories of data. While the optimum sample design often leads to the selection of samples specific to crops, livestock, and the respective economic, environmental, and social surveys, this limits data analysis across the respective categories.

Single purpose surveys generally make it easier to target the selected sample such as crops or livestock, especially where both are not present on most farms, or when present, differ considerably in size. It is difficult to use stratified designs using many different measures of size. There are recent developments in sampling theory that can provide an alternative using selection probabilities based on the measures of size for a number of different variables. This design is termed "Multiple Probability Proportional to Size" (MPPS) because the relative size of each farm (or enumeration area) is determined for more than one item of interest. The use of this method in China is described by Steiner (2007). It takes advantage of efficiencies of Probability Proportionate to Size sampling while adding the use of multiple measures of size. The use of MPPS is appropriate for multiple purpose surveys where the population sample units each only have a subset of the items of interest.

> China's Integrated Statistical System
> MPPS sampling using multiple variables from the Census of Agriculture is used to support an expanded survey program and to integrate the statistical needs for different levels of government.

For the purposes of data analysis, it is desirable to select one large sample to provide all of the data for production, the economic situation of the holding, its environmental impact, and the social-well being of the household. It would also be desirable for the same sample to be used over time for longitudinal data analysis. While the MPPS sample design provides the basis to use a single sample, at the same time it requires lengthy and complex questionnaires to include all items of interest. For this reason, a strategy to collect data for some core items annually coupled with periodic data collection for other items is required to allow analysis across subjects.

The integrated survey framework should be based on the minimum set of core and national data and the determination of how frequently they are required.

a. Determine the set of core items for which at least annual data are required. For those core items not needed annually, group them by category including economic variables such as farm structure expenditures and income; environmental measures such as the use of fertilizers and chemicals and land and water use; Social variables

such as household income and well being; and other items of national interest (minor crop/livestock items, for example). Data for these items will come from rotating panel surveys based on a subsample of the core survey.

b. Select for example a replicated sample for the annual core items using MPPS. In other words, instead of selecting one large sample, select several replicates. This allows a process to include some of the sample units in the survey across time for longitudinal analysis. Diagram B shows 12 replicates; 1 through 5 for year 1, 2 through 6 for year 2, etc. This provides longitudinal data, but also limits the number of times for respondent burden considerations.

c. Design a survey questionnaire to obtain the annual core data items. Each year the core questionnaire should contain supplemental questions regarding one of the subject matters described above. For example, in year one replicates 1 through 5 will be surveyed using the core questionnaire which will also contain key questions about economic variables. The core questionnaire can either obtain all information required, or a subsample could be selected for the collection of the detailed data. In year 2, replicates 2 through 6 will be surveyed using the core questionnaire which will contain questions about environmental issues. By year 4 all of the subject matters will have been included.

d. Each year, one of the sets of panel data will be linked to the annual core items. Also note that starting with year 4, at least one of the replicates will have been surveyed by all of the rotating panel questionnaires in addition to the core questions.

The table below provides an overview of a survey framework based on replicated samples which are surveyed each year for the annual core data items. In addition, each year the core questionnaire contain a set of supplemental questions for one of the subject matters that round out the minimum set of core data.

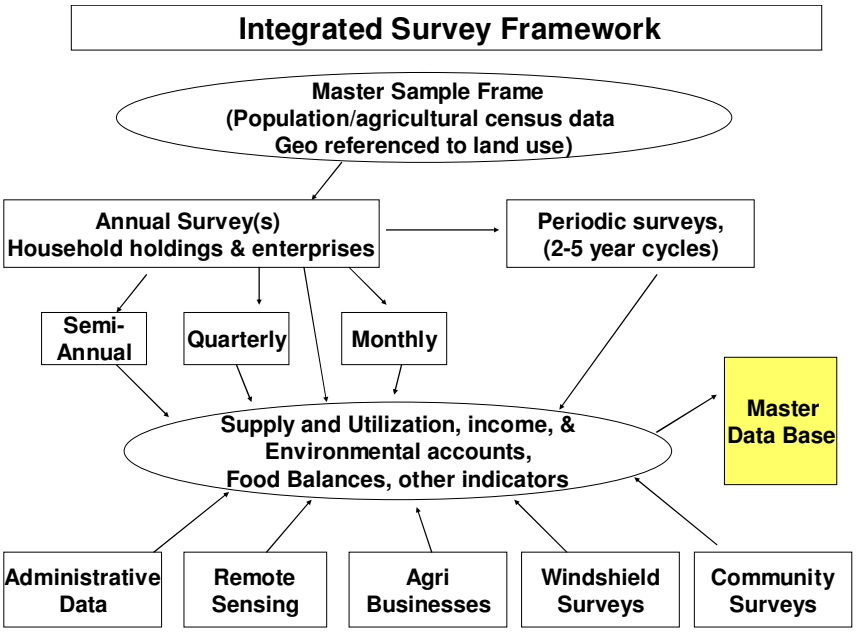
**Example of a replicated survey design with the use of an annual core questionnaire and rotating sets of supplemental questionnaires.**

| Replicate / Year | Rep 1 | Rep 2 | Rep 3 | Rep 4 | Rep 5 | Rep 6 | Rep 7 | Rep 8 | Rep 9 | Rep 10 | Rep 11 | Rep 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | **Every replicate receives same core questionnaire every year for annual core data items. Each core questionnaire contains questions about the rotating panel surveys for sub-sampling purposes.** | | | | | | | | |
| **1** | **A** | **A** | **A** | **A** | **A** | | | | | | | |
| **2** | | **B** | **B** | **B** | **B** | **B** | | | | | | |
| **3** | | | | | | **C** | **C** | | | | | |
| **4** | | | | | | **D** | **D** | **D** | | | | |
| **5** | | | | | | **A** | **A** | **A** | **A** | | | |
| **6** | | | | | | **B** | **B** | **B** | **B** | **B** | | |
| **7** | | | | | | | **C** | **C** | **C** | **C** | **C** | |
| **8** | | | | | | | | **D** | **D** | **D** | **D** | **D** |
| **9** | | | | | | | | | **A** | **A** | **A** | **A** |
| **10** | | | | | | | | | | **B** | **B** | **B** |
| **11** | | | | | | | | | | | **C** | **C** |
| **12** | | | | | | | | | | | | **D** |

**Detailed Questionnaires for Rotating panel surveys**
**A. Economic items including Farm structure, expenditures, income**
**B. Environmental items including inputs, chemicals, tillage, water use, land use**
**C. HH income, consumption, employment**
**D. Items of national interest**

The above survey design provides a strategy to collect data for core items—some annually, others on a 4 year rotating cycle. Each country will need to make its own decisions about the content of each of the components. Once the above design is in place, the next consideration is whether some of the data collections for the annual core items should take place more frequently during the year. One example would be to conduct a midyear survey to determine crop yields before harvest, another to obtain the final production and stocks.

The integrated survey framework also offers the opportunity to compare sample unit data across time providing a major validation tool to improve data quality. The integrated framework also provides the opportunity to use alternative estimators. While the direct unbiased estimators based on the sample design form the foundation, they can be supplemented using ratio and regression estimators, and/or model based estimators using census results. The use of multiple estimators can improve data accuracy and reliability.
The integrated survey framework shown in the figure below provides an overview of how the annual and periodic surveys are connected in the data system. Note that within year surveys can also be conducted using sub-samples from the annual survey.




The survey framework also takes into account the additional data sources that need to be included in the overall framework. These include:

a. *Administrative data.* Governmental interventions such as subsidies, regulation, and legislation often require agricultural holders to report production information. Land ownership and cadastral surveys provide useful information for constructing registers. Food inspections, animal health inspections, and trade data provide input to the utilization accounts.

b. *Remotely sensed data.* These include vegetative indices that show overall crop conditions and information about changes in land cover and use. The

survey framework should include the need to provide ground truth data if remote sensing information is to be used to estimate cropland areas.

c. *Agri-businesses* are the source of utilization data and prices.

d. *Expert Judgment and Windshield surveys* can be used to collect data from experts whose judgments inform evaluations of agricultural conditions. For instance, the *Sourcebook* refers to a procedure in which experts travel a specified route on a periodic basis and record the condition of crops, which provide an input into crop yield forecasts.

e. *Community surveys.* The World Programme for the Census of Agriculture provides an overview of data that can be collected at the village level. These data include information about the infrastructure and services available to households and agricultural holdings, occurrences of food shortages, frequency of natural disasters, etc.

The integrated survey framework will provide annual data for a core set of items on agricultural production and other variables determined by the national statistical system. The survey framework enables longitudinal analysis of the core data and it provides linkages to the data that are collected regarding economic, environmental, and social issues. The use of the master sample frame ensures that the data collection is connected to land use as well. The remaining pillar of integration is the management of the data to maximize their use for analysis.

***The Data Management System: Master database***

The official statistics that are gathered are to reside in a data management system. These are the basic principles of the Strategy. Their implementation will require improved governance across the national statistical system. The concept of a master sample frame will be extended to include a data management system for all official statistics related to agriculture.

The data management system fulfils three main functions—access to official statistics for dissemination purposes, storage and retrieval of survey results, and access to farm, household, and geo-referenced data for research. The data management system should:

i. Support the dissemination of data to ensure the official statistics are readily available, clearly identified by source and time, and are comparable for aggregation purposes, both within and across countries. If more than one institution is involved in the national statistical system, there should either be a single database, or the databases should be coordinated to avoid duplication of official statistics. Such duplication can lead to different numbers, causing confusion among those using the data. These data could be part of CountrySTAT and FAOSTAT, and becomes a public good for data access.

ii. Provide the framework for the storage of the aggregated survey results and geo referenced land use data along with the supporting administrative and other data sources. Not all survey results are published, however, they should be available for research and analysis purposes. As described above, the sample and survey design enables the use of ratio and regression estimators requiring linkage to previous data.

iii. Build on the capabilities provided by the master sample frame's linkage to land use. The data management system should provide for the storage and maintenance of the farm and household survey data and for the linkage between the different sets of data that are geo-referenced to a common land use. For example, there will be five consecutive years of core production data for the same sample units plus data from the rotating panel surveys. The strength of the integrated survey system will come from the data analysis capabilities provided by this data set.

The value of the integrated database will increase over time as the database itself grows. It will enable more analyses across time, and it can be used to improve data quality by comparing survey information with census data or between surveys over time. The output of the aggregated values will be the input to Country Stat following its methods and principles.

The integration of agriculture into the national statistical system through the implementation of a master sample frame, an integrated survey framework, and an integrated database, will require countries to review their current governance structures. Some countries will have to make changes in order to meet the challenges of coordination and to ensure that the statistical system is sustainable.

## 6. FAO Integrated Agricultural Census/survey programme as a component of Agricultural sector statistical master plan within NSDS

***Importance of the Census of Agriculture in integrated statistical system to meet data needs***

As indicate above, agricultural census is an essential element of the statistical system for meeting priority data requirements of a wide variety of data users. The FAO World Programme for census of agriculture 2010 identifies the main domains for use of census data:

*Monitoring the Millennium Development Goals*

The new modular approach used for the current round of agricultural censuses, based on the census core and supplementary modules together with the programme of agricultural surveys, enhances the usefulness of the agricultural census/survey programme as a source of data for MDG monitoring.

*Poverty monitoring and analysis*

Achieving sustainable economic growth with the focus on combating poverty has become the key development goal for governments around the world, as reflected in the MDGs and, in particular, Goal 1. Most of the poor live in rural areas, often in isolated conditions, where they face problems of poor natural resources, underdeveloped infrastructure, lack of access to markets, fluctuating commodity prices, lack of employment opportunities, and vulnerability to natural disasters. The agricultural census helps to better understand the causes of poverty and provide baseline data for monitoring poverty alleviation programmes.

The community-level data collection, introduced for the first time in the 2010 programme, can provide a useful source of data on infrastructure issues affecting farmers' incomes, especially relating to the access farmers have to agricultural produce markets.

*Food security monitoring and analysis*

The importance of combating hunger while achieving economic growth is one of the cornerstones of the MDGs, as reflected in Goal 1. A wide range of data is needed to monitor progress towards this goal, and the agricultural census can play a role in this regard.

On the food availability side, data from the agricultural census helps in understanding the structure of the food production industry and the constraints faced by farmers in increasing agricultural production, as well as suggesting strategies for increasing agricultural productivity. Cropping patterns can be studied along with information on the use of irrigation, farm machinery and improved varieties of seed to help develop programmes for increasing food production.

*Measuring the role of women in agriculture*

The contribution of women to agricultural development is often not well-understood because of the lack of data and the problems in accurately measuring women's involvement in agricultural production activities. The agricultural census can be an important vehicle for studying the social and cultural patterns of agricultural and rural development as they relate to women, the distribution of agricultural work within households, and the interactions between different household members in the management and operation of agricultural holdings.

*Agricultural planning and policy-making*

*Study of a specific crop.* Census tables specific to agricultural holdings with the particular crop – for example, coffee – can be used to measure the number and location of coffee growers, the distribution of coffee growers by plantation area, cropping systems used by coffee growers, labour requirements for coffee growing, etc.

*Study of a specific livestock production system.* Census tables specific to agricultural holdings with the particular livestock type – for example, sheep – can be used to measure the number and location of sheep producers, the distribution of sheep producers by flock size, the integration of sheep raising with cropping activities, etc.

*Structure of agriculture in a particular area.* Census tables relating to the particular geographic area, such as a district, can highlight the main crops grown and livestock raised in the district, the agricultural practices used in the district in comparison with other districts, employment characteristics in the district, etc.

*Inter-relationship between crop and livestock production.* Census tables can be prepared showing the number of holdings with specific combinations of crop and livestock types.

*Sources of farm labour*. Census tables can be prepared to show the types of farm labour inputs for specific farming systems and the role of household and outside labour.

*Farm typology studies.* The agricultural census can be useful for classifying holdings by type, as an aid to developing agricultural development policies. For example, holdings can be sub-divided into whether they are subsistence or market oriented, and different policies and programmes can be developed for each group.

*Studies of small holdings.*

*Improving current agricultural statistics*
The agricultural census can provide reliable current data relating to crop and livestock production for the census year, and this can be useful as a benchmark for improving current crop and livestock statistics.

*Providing baseline data for monitoring agricultural development projects*
Typically, an agricultural development project aims to achieve certain outcomes in a defined project area. Baseline data are needed to help assess whether the project has been successful. An agricultural census provides detailed structural data for small geographic areas, making it an ideal source of baseline data.

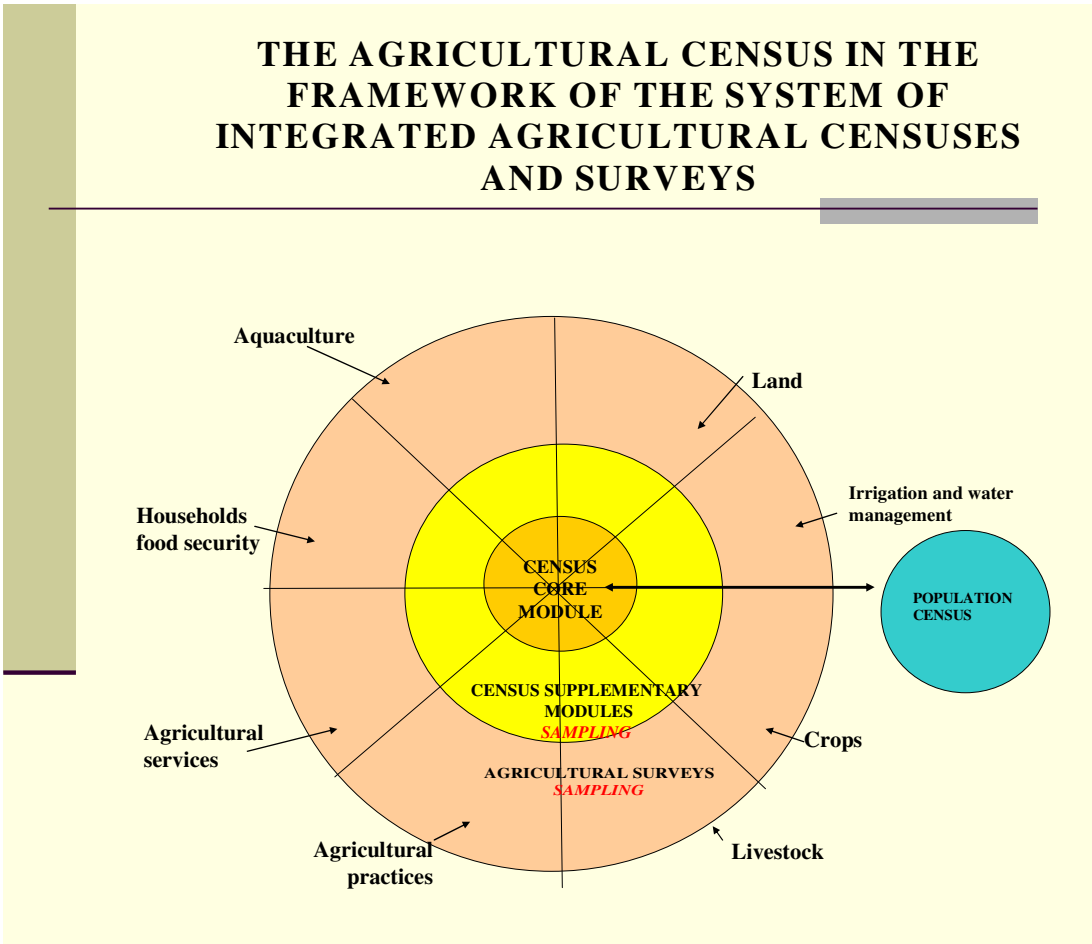
*Providing data for the private sector*
As well as providing data for government planning and policy-making, an agricultural census is also a valuable source of data for the private sector. The main interest for the private sector is usually in data to help make commercial decisions. A food processing company could use agricultural census data on the number of growers and area for specific crops in each district to help identify suitable sites for its processing plants. An input supplier could use census data on input use for each crop by district to better understand market opportunities. Farm machinery suppliers could make use of data on the area of each type of crop grown and the number of growers to assess the potential demand for their products. A company planning to establish a business in a particular location could use census data to assess the availability of labour and the pool of skills available in that location.

***The Census of Agriculture in an Integrated Agricultural Statistics System***
As indicated in the paragraphs above, integration, in a statistical sense, means that each statistical collection is carried out, not in isolation, but as a component of the national statistics system. In an integrated agricultural statistics system, the census of agriculture provides certain types of data as part of an integrated set of data on food and agriculture, needed for decision-making in food, agriculture and rural development.

The main advantages of an integrated statistics system can be summarised as follow:

a. It is possible to plan and develop a comprehensive statistical programme, without duplication of statistical activities or the release of conflicting statistics, while ensuring the efficient and balanced use of available statistical resources.
b. Concepts, definitions and classifications used in the different statistical activities can be made compatible, making it easier to interpret and analyse related data from different sources.
c. Any one statistical collection, such as the census of agriculture, can be restricted to a coherent and manageable set of items, in the knowledge that other related data are available in a comparable form from other sources.

Planning and implementing an integrated agricultural statistics system requires an efficient organization, trained personnel at various levels, and secured budgetary allocations over a period of years. Efficient organization implies strong cooperation between users and producers of agricultural statistics. The approach to integrated and modular approach to agricultural census and survey programme promoted by FAO can be summarised as follow:




## 7. Importance of sampling in integrated census survey programme

In the above approach, a limited number of core data items to be included in the census of agriculture are to be collected through completed enumeration. Some of these data items can be included in an agricultural module of a population census (Annex C provides African countries' plans for 2010 population census). Data items to be covered by Census supplementary modules and follow-up agricultural surveys are to be collected on a sample basis. This will be further developed through the other sessions of this workshop.

The effective implementation of this approach therefore requires intensive use of sampling and staff qualified in this technique and this workshop aims at contributing to this goal.

## ANNEX A: Minimum set of core data

| Group of Variables | Key Variables | Core data items | Frequency |
|---|---|---|---|
| *Economic* | | | |
| - Output | Production | Core crops (e.g wheat, rice, etc.)<br>Core livestock (e.g. cattle, sheep, pigs, etc.)<br>Core forestry products<br>Core fishery and aquaculture products | Annual |
| | Area harvested and planted | Core crops (e.g wheat, rice, etc.) | Annual |
| | Yield / Productivity | Core crops, core livestock, core forestry, core fishery | Annual |
| - Trade | Exports in quantity and value | Core crops, core livestock, core forestry, core fishery | Annual |
| | imports in quantity and value | Core crops, core livestock, core forestry, core fishery | Annual |
| Stocks | Quantities in storage at beginning of harvest | Core crops | Annual |
| - Stock of Resources | Land cover and use | Land area | [1] |
| | Economically active population | Number of people in working age by sex | |
| | Livestock | Number of live animals | |
| | Machinery | e.g. Number of Tractors, harvesters, seeders etc. | |
| - Inputs | Water | Quantity of water withdrawn for agricultural irrigation | |
| | Fertilizers in quantity and value | Core Fertilizers by core crops | |
| | Pesticides in quantity and value | Core Pesticides (e.g. fungicides herbicides, insecticides, disinfectants) by core crops | |
| | Seeds in quantity and value | by core crops | |
| | Feed in quantity and value | by core crops | |
| Agro processing | Volume of core crops/livestock/fishery used in processing food | By industry | |
| | Value of output of processed food | By industry | |
| | Other uses (e.g. biofuels) | | |
| Prices | Producer prices | Core crops, core livestock, core forestry, core fishery | |
| | Consumer prices | Core crops, core livestock, core forestry, core fishery | |
| Final expenditure | Government expenditure on agriculture and rural development | Public investments, Subsidies, etc. | |
| | Private Investments | Investment in machinery, in research and development, in infrastructure | |

[1] The frequency for the following items will be established by the framework provided in the Global Strategy to determine the national priorities for content, scope, and frequency. The frequency requirement will also be considered in the establishment of the integrated survey framework where the data sources will be defined.

| Group of Variables | Key Variables | Core data items | Frequency |
|---|---|---|---|
| | Household consumption | Consumption of core crops/livestock/etc. in quantity and value | |
| Rural Infrastructure (Capital stock) | Irrigation/roads/railways/ communications | Area equipped for Irrigation / Roads in Km / Railways in Km / communications | |
| International transfer | ODA[2] for agriculture and rural development | | |
| *Social* | | | |
| Demographics of urban and rural population | Sex | | |
| | Age in completed years | By sex | |
| | Country of birth | By sex | |
| | Highest level of education completed | 1 digit ISCED by sex | |
| | Labor status | Employed, unemployed, inactive by sex | |
| | Status in employment | Self Employment and employee by sex | |
| | Economic sector in employment | International Standard Industrial Classification by sex | |
| | Occupation in employment | International Standard Classification of Occupations by sex | |
| | Total income of the household | | |
| | Household composition | By sex | |
| | Number of family/hired workers on the holding | By sex | |
| | Housing conditions | Type of building, building character, main material, etc. | |
| *Environmental* | | | |
| Land | Soil degradation | Variables will be based on above core items on land cover and use, water use, and other inputs to production. | |
| Water | Pollution due to agriculture | | |
| Air | Emissions due to agriculture | | |
| *Geographic location* | | | |
| GIS coordinates | location of the statistical unit | Parcel, Province, Region, Country | |
| Degree of urbanization | Urban/Rural area | | |

[2] Official Development Assistance

## Annex B: Menu of Indicators for Agricultural Statistics

| | Indicator | Data Requirements | Data Sources | Technical Notes |
|---|---|---|---|---|
| | **Sector Wide indicators for agriculture and rural development** | | | |
| 1 | Gross Domestic Product (GDP)— | | Censuses and surveys of firms, farms, and households for small holders. | Value added should include unreported activities as well as the value of informal or small scale operations. Annual estimates between census or surveys based on extrapolations based on other indicators. |
| 2 | GDP growth from Agriculture value added. | Estimates of total production and value for all commodities produced in the country; including that from small holders/household plots minus estimates of the cost of inputs such as seed, feed, energy, fertilizer, labor, etc. Agriculture includes forestry and fisheries | Censuses and surveys agricultural enterprises, farm and rural households, administrative and processor. data | SNA concepts followed. Problems include estimation of output consumed by the household and the annual coverage of all commodities for which only periodic census data are available. Annual estimates made using previous census and other administrative data if available. |
| 3 | Amount of public spending on agriculture, subsidies, and infrastructure | Government budget allocations, and spending related to agriculture. Agriculture includes forestry and fisheries | Ministry of Finance, National Accounts, Planning commissions, Donor reports | The definition for public spending on agriculture should follow the UN Classification of Functions of Government (COFOG) for agriculture |
| 4 | Amount of public spending on rural infrastructure including health and education | Government budget allocations, and spending related rural areas | Ministry of Finance, National Accounts, Planning commissions, Donor reports | Rural defined using national description |
| 5 | Change in Investment in capital stock | Inventories of machinery and equipment owned by agricultural holdings, buildings such as milking purposes, animal breeding stock, area of semi-permanent crops such as trees and vineyards, number of trees and vines | Agricultural resource surveys of holdings and agricultural enterprises | Machinery and equipment inventories should be by purpose (tillage, harvesting, etc.) and size |
| 6 | Demographics of agricultural and rural population | Rural population and number of rural households, , number of agricultural households and population living in them, age and education levels. Agriculture includes forestry and fisheries | Census of Population, Census of Agriculture, Household surveys, administrative records | Rural defined using national description |
| 7 | Rural poor as a percent of total poor population | Household income and consumption estimates for national and rural | Household Surveys. International Comparison Program for comparisons | Countries should use poverty estimates based on PPPs and extrapolate between ICP |

| | Indicator | Data Requirements | Data Sources | Technical Notes |
|---|---|---|---|---|
| | | poverty lines. Purchasing Power Parities for comparisons across countries | across countries | benchmarks |
| 8 | Rural hungry as a percent of total poor population | Household income and food consumption estimates for national minimum energy requirements. | Household Surveys. International Comparison Program for comparisons across countries | Countries should use hunger estimates for monitoring food deprivation levels |
| 9 | Food production index | Area, production and yield for food crops, livestock numbers and production of meat, milk, eggs, fish captured and cultured, and other food products, non-food use of food products, food imports and exports | Agricultural Census, surveys of agricultural enterprises, processors, fish landings, administrative data such as imports, exports. Food Balances and Household consumption surveys | Follow FAO guidelines for inclusions and exclusions |
| 10 | Change in value of Trade—imports and exports | Imports and exports—quantities and values of agricultural products including fishery and forest products | Customs inspections—in some countries the customs offices collect the data which then are turned over to the national statistical office for compilation | National statistical offices should collaborate with customs officials to ensure coding and classifications follow international guidelines |
| | **Indicators for subsectors of agricultural and rural** | | | |
| 11 | Productivity of Crop production as measured by crop yields | Quantity harvested per unit of area such as hectare and area harvested. Area harvested, distinguished between irrigated harvested crops and rainfed harvested crops | Census of Agriculture, crop cutting surveys. Production sample surveys, processor surveys, such as oil seed crushers, cotton ginners | Difficult to measure with mult-cropping or with crops that can be harvested > once a year. Crop cutting can over estimate yields |
| 12 | Change in components of crop balances | Area Harvested, Quantity harvested, quantities imported/exported, change in stocks, quantities by utilization such as food, bio fuels, own consumption, for every crop including those produced for fiber and oil | Surveys of agricultural enterprises, administrative data on trade, processors by utilization, household surveys for own consumption | Crop balances should reflect the growing cycle and marketing year which could be different from the calendar year. |
| 13 | Livestock value added | Estimates of quantity and value of production of meat, and poultry, milk, eggs, by products such as hides and skins, wool mohair minus costs of inputs such as feed and replacement stock | Surveys of agricultural holdings, enterprises such as slaughter plants, dairies, processors. Household surveys for own consumption | Own consumption should be included, difficult to measure. |
| 14 | Change in components of Livestock and | Number of animals born, acquired, slaughtered, deaths from disease. | Surveys of agricultural holdings at least annually but more often for | Data collection intervals should reflect the reproductive cycles. This suggests annual for cattle, |

| | Indicator | Data Requirements | Data Sources | Technical Notes |
|---|---|---|---|---|
| | poultry Balances by species | Number of animals by purpose such as breeding, meat, milk, wool, and by age breakdowns relevant to specie. (see FAO 2010 Census) | species with more frequent births during a reference period. This ranges from annually for cattle to monthly for egg production. | semi- annual for pork, quarterly or shorter for poultry, milk, |
| 15 | Change in productivity of Capture Fish production | Quantity of fish taken by unit of fishing effort; Scientific estimates of fish stock and exploitation rates; | National fishery surveys, surveys at landing sites, on-board observers, national, regional and global assessment results; | |
| 16 | Change in productivity of aquaculture | Estimates of quantity and value of production of fish by species minus costs and quantity of inputs such as seed, feed and fertilizers | Surveys of aquaculture enterprise, and holdings, aquaculture census, market certifications, | |
| 17 | Change in components of fish balances | Quantities and value of captures form coastal and offshore waters, rivers and lakes including non-landed catch; Quantities and value of products from aquaculture; utilizations including own consumption and discards, imports and exports, inputs such as seed and feed; outputs such as stocking; for each aquatic species | National fishery surveys, fishery census, aquaculture census, surveys of fishery and aquaculture enterprises, , processors, market information, administrative and inspection sources | See CWP Handbook, FAO coding and classification |
| 18 | Change in components of forestry balances | Quantity and value of removals of products from forested areas and respective utilizations | Appropriate ministries, satellite imagery, price surveys or processor data | |
| 19 | Commodity Price indexes | Market reports of prices being offered by commodity and location. Prices received by the enterprise at the first point of sale, | Market observers, Surveys of enterprises, agro enterprises purchasing commodities from agricultural enterprises | Care needed to ensure units of measure for pricing are comparable |
| 20 | Consumer Price indexes | Monthly/seasonal prices paid by the consumer | Consumer Price Index, | Care is needed to ensure highly seasonal products do not distort the price series. |
| 21 | Early warning of change in food security | Monthly/seasonal prices paid by the consumer | Windshield surveys of crop conditions, amount of precipitation, satellite imagery of vegetative indexes, changes in trade data, animal disease outbreak | These do not have to be statistically rigorous, mainly to provide an early warning that other interventions are needed |
| | **Climate Change, land, and the environment** | | | |
| 22 | Change in Land Cover and use | Land Cover Classification System (LCCS), Area and | Land use surveys, satellite imagery. Geo referenced | Ground truth data required to provide more detailed |

| | Indicator | Data Requirements | Data Sources | Technical Notes |
|---|---|---|---|---|
| | | geo-referenced for Cultivated land, Grass/pasture, inland water, marine water, wetlands, shrubland, woodland, fallow/idle cultivated land, barren land, urban/developed areas, areas equipped for irrigation. | data on economic situation of agricultural holdings needed to understand effect of policy decisions on land use. | breakdowns of cultivated land, especially for crops in small plots. Difficult to apply in detail where multi- cropping is used. |
| 23 | Change in proportion of land area covered by forests, rate of deforestation | Area geo referenced to map materials | Ministry responsible for forestry, satellite imagery | Follow LCCS classification |
| 24 | Percent of land and water area formally established as protected areas | Land and water area and geo referenced to mapping material | Responsible ministry—satellite imagery | Follow LCCS coding with expansion covering inland and marine water bodies |
| 25 | Irrigated land as percent of total cropland<br><br>Productivity of irrigation | Total cropland and area irrigated by source of water for irrigation—(surface water, groundwater, treated wastewater, etc.)<br>- by method (surface, sprinkler, localized irrigation)<br><br>Crop yields from irrigated land compared to yields from non irrigated areas. | Agricultural Census, other crop related surveys or water user survey | irrigation refers to the artificial application of water to assist in the growing of crops (and pastures). Can be done by letting water flow over the land ("surface irrigation"), by spraying water under pressure over the land concerned ("sprinkler irrigation"), or by bringing it directly to the plant ("localized irrigation") |
| 26 | Withdrawal of water for agriculture as a percent of total water withdrawal | Area under irrigation, number of irrigations, irrigation intensity and requirements by crop, water withdrawal and turn over rate for aquaculture consumption, per capita consumption by people and animals | Appropriate ministries, special studies or surveys to estimate water use in agriculture and aquaculture, surveys of aquaculture enterprises and holdings. | Should include both surface and ground water. Coding and classifications should be defined |
| 27 | Change in soil loss from watersheds | Reduction in crop yields, reduction in area of cultivated land | Appropriate ministries, geo referenced data with satellite imagery | |
| 28 | Change in affect of inputs on the environment | Fertilizer, pesticide, and other chemicals applied to the soil, water bodies, and plants by type of crop and watershed area, stocking | Agricultural census and or follow-up surveys to measure fertilize and chemical use, tillage methods | Data should be geo referenced to land cover and use |
| | **The agricultural and rural economy** | | | |
| 29 | Number of | Include Unpaid labour of | Labour force surveys of | Need to establish standards for |

| | Indicator | Data Requirements | Data Sources | Technical Notes |
|---|---|---|---|---|
| | family and hired workers on the holding | the operator of the holding and family members plus number of hired workers | holdings | minimum ages of workers and the number of hours worked per week to be considered a worker. Need to define reference period Need to ensure female workers are counted |
| 30 | Number of household members employed by farm and non farm | The employment status for work off the agricultural holding for each household member | Labour force surveys—household surveys | Need to distinguish defined employment from unpaid household service work such as domestic chores. |
| 31 | Change in Farm and Rural non farm household income from all sources | Income to the household by sector, crop, livestock, etc. Income from investments or employment outside the agricultural holding | Rural Household Survey. | Rural to be classified using range in population density using national definitions |
| 32 | Percent of rural population using services of formal banking institutions | Total number of rural households, number using credit or savings services | Central Bank or commercial banks, special surveys, agricultural census | |
| 33 | Change in sales of agro enterprises | Sales, net profits of enterprises providing services to agriculture | Special surveys | Use standard accounting principles |

**Annex C: African countries Plans for Population censuses in 2010 round**

| Year | Number of countries | Countries |
|---|---|---|
| 2005 | 1 | Cameroon |
| 2006 | 5 | Burkina Faso, Egypt, Lesotho, Libya, Nigeria |
| 2007 | 3 | Ethiopia, Mozambique, Swaziland |
| 2008 | 6 | Algeria, Burundi, Congo, Liberia, Malawi, Sudan |
| 2009 | 5 | Chad, Djibouti, Guinea Bissau, Kenya, Mali |
| 2010 | 10 | Angola, Cape Verde, Congo (DRC), Côte d'Ivoire, Ghana, Guinea, Mauritania, Senegal, Togo, Zambia |
| 2011 | 7 | Botswana, Eritrea, Mauritius, Namibia, Niger, Sao Tomé and Principe, South Africa |
| 2012 | 7 | Benin, Equatorial Guinea, Rwanda, Seychelles, United Republic of Tanzania, Uganda, Zimbabwe |
| 2013 | 4 | Central African Republic, Comoros, Gabon, Gambia |
| 2014 | 3 | Morocco, Sierra Leone, Tunisia |
| n.s. | 2 | Madagascar, Somalia |
| TOTAL | 53 | |

## New features of the World Programme of Census of Agriculture 2010 (WCA2010): Modular Approach to Data Collection, *M. Srivastava, FAO*

*The new World Programme for the Census of Agriculture 2010 has been issued as guidance to countries in the preparation for their agricultural censuses to be undertaken during the decade 2006-2015. It has many new features as compared to the previous Programmes, in particular the data items to be covered, the relation between the agricultural census and related surveys and censuses, and the role of agricultural census in monitoring emerging concerns at world summits. The paper presents these new aspects focusing on the recommendations to: (i) collect data at community level to help meet the increasing demand for such data by policy makers, planners and development partners; (ii) include new data items on agricultural services, agricultural practices and aquaculture; (iii) adopt a modular approach for the census taking; (iv) establish linkage between agricultural census and other major statistical undertakings such as population census; and (v) the role of agricultural census in obtaining indicators useful for monitoring the Millennium Development Goals.*

### 1. Introduction

The new World Programme for the Census of Agriculture (WCA 2010) has been issued by FAO in 2005 for the decade 2006 – 2015. This is the ninth programme in the series of World Censuses of Agriculture, and the seventh decennial programme issued by FAO. The programme would be useful to the countries, which plan to conduct their national censuses during this period. FAO World Programme for the Census of Agriculture (WCA) has helped countries to carry out their national agricultural census at least once every decade using standard international concepts, definitions and methodology.

The new programme, besides maintaining the conventional role of agricultural census as provider of key structural data at lower geographic level, sees the agricultural census as the heart of the national food and agricultural statistics system. Its role towards monitoring of Millennium Development Goals (MDGs), and monitoring and measurement of poverty, contribution of women to agriculture and food security is being increasingly realized. The new programme gives special attention to this aspect. For preparing targeted programme for development of rural areas, increasing need is being realized for Community-level data on rural infrastructure, such as effectiveness of agricultural produce markets. The programme shows how agriculture related community-level data could be collected easily with agricultural census.

The additional new features of this programme are the following:

- A modular approach to agriculture data collection has been introduced. The core census covers all agricultural holdings but collects only a limited range of items. Supplementary modules use samples chosen from the core census;
- Emphasis is given to integrating national agricultural and population censuses for greater synergy and cost-effectiveness;
- The option of combining the agricultural and the aquaculture censuses is included;
- The option to include households that are not agricultural producers is considered;

- Data content, concepts, definitions and classifications have been updated and improved.
- A concept of sub-holding has been introduced to better capture the data on contribution of women.

In order to increase the utility of data collected through the agricultural census the concepts and classifications recommended for WCA 2010 have been aligned to standard international practices, particularly to the national accounting framework. In particular, the classification of crops has been harmonized with the latest version of Common Product Classification (CPC). The classification of land use has been fine tuned to collect data for seven elementary land categories which could be further broken down by countries as per their needs. A classification on livestock and one on agricultural machinery have been included in the new document. The scope of agricultural census within the framework of International Standard Industrial Classification of all Economic Activities (ISIC) has been examined in detail to establish clear linkage with the national accounts framework.

**2. Expanded list of data items and Modular approach**

The previous programme recommended to countries a set of about 10 "essential data items" for inclusion in their agricultural census. Also guidelines were provided to countries on a few "optional data items" for inclusion in their census programme as per their need. The new programme (WCA 2010) has recognized the need for international comparison as well as the challenge faced by countries to collect data on a variety of new subjects. The new programme give guidance to countries on a large set of data items, in addition to 16 key items recommended for inclusion of census programmes of all countries. The concepts relating to many new data items have been explained, particularly in the areas of Agricultural practices, Agricultural services, Aquaculture, Irrigation and Water management, Livestock rearing cycles, and Management of the holding. Additional data-items have been grouped in 12 themes.

The new programme offers to the countries a great amount of flexibility to design their survey programme as per their needs, priorities and resource availability through adoption of a **modular approach**. Under this approach, a core census module, to be conducted on a complete enumeration or large sample basis, will provide a limited range of key data at lower administrative level. The programme lists 16 items for inclusion in the **core census module** of all countries for key international comparison purposes. The core module would also provide frames for further surveys to be carried out on the basis of agricultural surveys. One or more **census supplementary modules** could be conducted on a sample basis as part of the census or soon after it to provide more in-depth data on a selected field. These optional items would be important for national policy-making, and would at the same time be useful for making inter-country comparisons. As these modules which will be integrated to the core census items, even though carried out at different points of time, they will offer a cost-effective solution to data production and enhanced scope for analysis.

**3. Community-level data**

Responding to a great demand of data at village or commune level, the new programme has specially addressed this issue. The programme identifies a list of data items which may not be suitable for collection with the agricultural holding as the statistical unit. The appropriate statistical unit for such data would be the community. These 33 data-items identified in the programme cover the data domains of: Geography of the Community, Socio-economic conditions, Community infrastructure services and Development programmes. The programme encourages the countries to include this component according to national circumstances and data requirements. While selecting the data items for inclusion in the agricultural census, it is to be noted that ample opportunities to interface this data with other similar data from other sources also exist.

**4. Agricultural Census in the National Statistical System**

National censuses are extremely large and costly undertakings. However, in the past, many countries have seen their agricultural census as an independent statistical exercise separated from their programme for collection of agricultural statistics. Such an approach tends to be costly and often leads to divergent data from various surveys. The new programme provides for agricultural censuses to be conducted as the central component of the system of agricultural sample surveys. Under this scenario, the agricultural census/survey programme can be viewed as having two related elements: (i) the agricultural census; and (ii) the series of agricultural sample surveys based on the agricultural census, called the thematic agricultural surveys.

There is also considerable commonality between the population and housing census and the census of agriculture. For this reason, FAO has worked closely with United Nations Statistics Division to explore ways in which the relationship between the two censuses activities can be strengthened that can save costs and enhance the usefulness of the data. The possibilities of integration of the two censuses exist in the following ways:

- use of common concepts, definitions and classifications;
- sharing field materials;
- using the population census as a household frame for the census of agriculture;
- making use of agriculture-related data from the population census;
- collecting additional agriculture-related data in the population census;
- linking data from the two censuses;
- conducting the two censuses as a joint field operation.

The WCA 2010 stresses that the opportunity for coordinating the various national census activities should be actively explored at an early stage in the census planning process, and be taken into consideration in developing national statistical plans. Countries should make the necessary administrative arrangements to ensure close collaboration between the census development teams and to ensure that all avenues for coordination are explored.

Yet another opportunity for integration of agricultural census with other census and surveys exists with aquaculture census, wherever aquaculture production industry is important and

such a census is to be conducted. There is a strong interest in many countries to link aquaculture with agriculture by carrying out the agricultural and aquaculture censuses together. Agriculture and aquaculture are seen to be closely related because:

- Aquaculture involves the raising of fish in captivity or cultivation of aquatic plants, which is analogous to rearing of livestock or cultivation of crops under agricultural production.
- Aquaculture is often integrated with agricultural production, such as in rice-cum-fish culture.
- Aquaculture and agriculture commonly share the same inputs, such as machinery and labour.

The WCA 2010 provides the option to conduct an aquaculture census at the same time as the census of agriculture to cover all aquaculture production, and yet keeping it conceptually separate from agricultural census. Countries are strongly urged to examine this option if aquaculture is an important economic activity in their country.

The integrated approach to survey planning could particularly prove cost-effective in countries where carrying out a statistical operation such as a census is a rare opportunity. Integrating the surveys and censuses also increases the scope of analysis of data gathered through various statistical operations, and hence enhances the utility of existing data to policy makers.

**5. Agricultural Census and World Summits**

FAO has been entrusted the task of estimating and monitoring targets set by World Food Summit by 2015. As country level policies are the main instruments for achieving the Millennium Development Goals (MDGs), country-level monitoring of the progress towards the MDGs has thus become an important element in formulating economic development strategies, and countries have begun to focus on the need for MDG-related indicators as a key component of the national statistical programme. A variety of data sources are being sought for this purpose. Considering that a census of agriculture is one of the largest national statistical operations undertaken by a country, WCA 2010 programme advocate its use as a potential source of data for monitoring the MDGs. The programme shows that at least 8 indicator to monitor 5 different MDGs could be obtained to a great extent through agricultural census programme. Agriculture census data could be particularly useful in monitoring:

- Eradication of extreme poverty,
- Achievement of Universal education,
- Promotion of gender equality and empowerment,
- Ensuring environmental sustainability,
- Development of global partnership fro development.

The new modular approach used for the current round of agricultural censuses, based on the census core and supplementary modules, together with the programme of agricultural surveys, enhances the usefulness of the agricultural census/survey programme as a source of data for MDG monitoring. Countries should look to carrying out regular agricultural surveys, based on the census of agriculture frame, to provide additional MDG-related data to complement the data collected in the census of agriculture.

Newer strategies for sustainable development and agrarian reform are being sought. Focus of these approaches often tends to be the challenges confronting the under-privileged communities in rural areas around the globe. FAO has been given the mandate at the International Conference on Agrarian Reforms and Rural Development (ICARRD) in March 2006 to identify the indicators to monitor the implementation of the declaration. The declaration reiterated the importance of traditional and family agriculture, and other small-holder production as well as the roles of rural communities in contributing to food security and the eradication of poverty. Community-level data, at the village or the commune level, can be useful for this purpose. Information on the infrastructure and services available to holdings would help in formulating, executing and evaluating community projects. Data on whether the community is prone to natural disasters or subject to seasonal food shortages can also be of interest for food security analysis.

**REFERENCES**

FAO (2005): A system of integrated census and surveys, Volume 1, World Programme for the Census of Agriculture 2010. Rome.

FAO (1995): Programme for the World Census of Agriculture 2000. Rome.

FAO (1983): Community Level Statistics. Rome.

# SESSION 2

## STRATEGIES FOR MODULAR APPROACHES TO COLLECTION OF STRUCTURAL DATA

# Operationalisation of the Modular Approach and Integration of Census and Surveys

*M. Srivastava, FAO*

## 1. Background

The 2010 round of agricultural censuses has been developed after a review of country experiences with the previous programme. Many countries are concerned about the costs of agricultural censuses, as well as the demands made on technical and other resources. Countries are finding it difficult to meet the increasing demand for often complex data on current issues, such as food security, agricultural practices, and the environment. Greater use is being made of sampling methods in agricultural censuses. Countries would also like to link agricultural and population census data.

The need to better integrate the agricultural census into the ongoing system of agricultural statistics, in particular, and in the entire national system of collection of statistics in general, is also of a concern to many countries. Needless to say that the integration with the statistical system provides a cost effective solution to data providers in meeting diverse data demands but also facilitates multi-dimensional analysis of data which would not be possible in a fragmented system of surveys. The integration while providing the flexibility to plan and implement the agricultural surveys by different agencies according to their specific needs, also avoids duplication in data collection efforts at country level. Modular Approach to planning of census and surveys is a method to prepare an integrated plan of data collection.

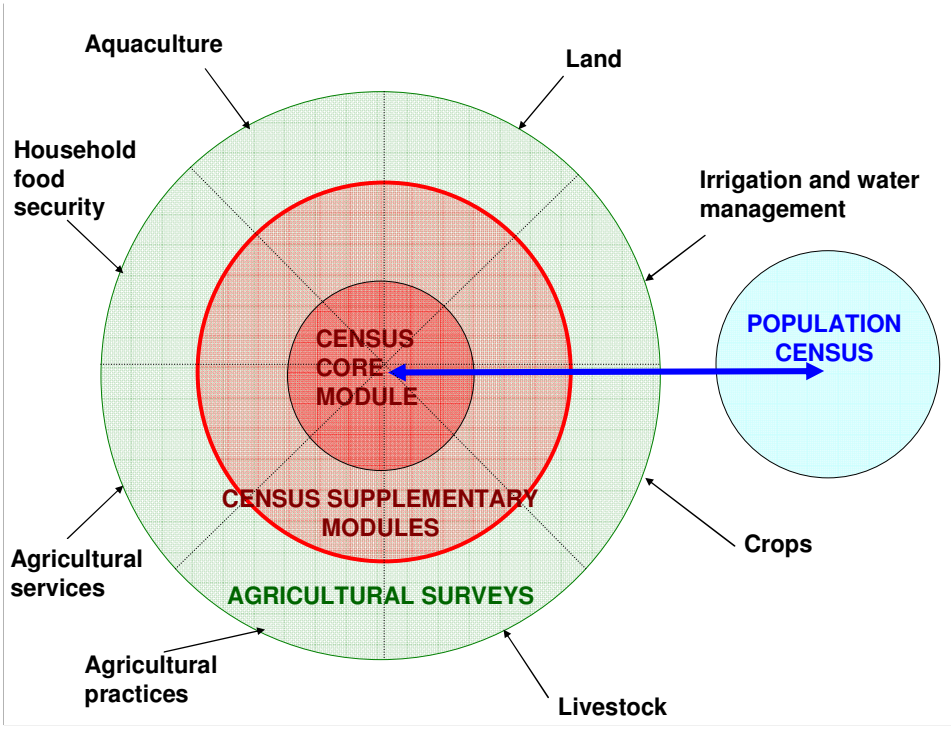
## 2. The framework of an integrated agricultural census - survey programme

The new programme provides for agricultural censuses to be conducted as the central component of the system of agricultural sample surveys. Under this scenario, the agricultural census-survey programme can be viewed as having two related components: (i) the *agricultural census*; and (ii) the series of agricultural sample surveys based on the agricultural census, called the *thematic agricultural surveys*. Each country is encouraged to develop its agricultural census within the overall framework of the system of agricultural surveys, and to plan the thematic agricultural surveys as a part of the agricultural census planning.

The thematic agricultural surveys will be based on sampling frames obtained from the agricultural census. These surveys will provide more detailed data for agricultural census topics, such as crops and agricultural practices. They will also include new topics, such as time use and cost of production, as well as ongoing agricultural surveys such as crop and livestock production.

A schematic representation of agricultural censuses in the framework of an integrated agricultural census/survey programme is shown in Figure 1. It shows items under selected headings, such as land, irrigation, etc., according to their suitability for inclusion in the census core module, in a census supplementary module, or in a thematic agricultural survey.

**Figure 1: Agricultural censuses in the framework of an integrated agricultural census/survey programme**




## 3. Wider scope of agricultural census

As in the past, the 2010 round of agricultural censuses focuses on the collection of structural data on agriculture from agricultural holdings. However, two options to widen the scope of the agricultural census are provided:

(i) In traditional agricultural censuses, units engaged in aquaculture production are included only if they also have some crop and livestock production activities. In view of the growing importance of aquaculture in many countries, the option is provided to cover all aquaculture production units.

(ii) Some countries may wish to expand their agricultural censuses to collect socio-economic data about other rural households that are not agricultural holdings, such as those containing farm labourers.

## 4. Modular approach

A modular approach is proposed to be used for agricultural censuses in the 2010 programme. A *core census module*, to be conducted on a complete enumeration or large sample basis, will provide a limited range of key data. One or more *census supplementary modules* will be conducted on a sample basis, using the core census module as a frame, to provide more in-depth data. This is essentially the short/long questionnaire approach already used by some countries.

A list of 16 items has been identified as being suitable for inclusion in the core census module. These items have been selected on the basis of: (i) their importance for national policy-making

and for making inter-country comparisons; (ii) the need to establish sampling frames; and (iii) the importance of providing data at low-level administrative units. The 16 items are:

- identification and location of holding
- legal status of agricultural holder
- sex of holder
- age of holder
- household size
- main purpose of production of the holding
- area of holding according to land use types
- total area of holding
- land tenure types on the holding
- presence of irrigation on the holding
- presence of temporary crops by type on the holding
- presence of permanent crops by type and whether in compact plantation
- number of animals on the holding by each livestock type
- presence of aquaculture on the holding
- presence of forest trees and other wooded land on the holding
- other economic production activities on the holding's enterprise.

Further 87 items have been identified as suitable for enumeration by sample methods in the census supplementary modules. These items may be as important as the 16 recommended core items, but do not need to be collected on a complete enumeration basis because they are not needed for sampling frame purposes or because data are not required for low-level administrative units. These 87 items are classified under 12 headings, representing possible topics for census supplementary modules:

- land
- irrigation and water management
- crops
- livestock
- agricultural practices
- agricultural services
- demographic and social characteristics
- economic activity of household members
- household food security
- aquaculture
- forestry
- management of the holding

Under each of these themes it is possible to imagine a <u>hierarchy of data items (questions)</u>, depending on details. For example, consider the data items on Livestock.

*<u>For the holding</u>*

| | |
|---|---|
| 0013 | Number of animal (of each species) on the holding |
| 0401 | Type of livestock production system |
| 0402+ | Use of veterinary services |

*For each livestock type*

0411 Number of animals: age, sex and breed
0412 Number of animals according to purpose
0413+ Number of milking animals according to milk status
0414+ Number of animals born
0415+ Number of animals acquired
0416+ Number of animals slaughtered
0417+ Number of animals disposed of
0418+ Number of animals died from natural causes
0419+ Types of feed
0420+ Value of livestock production
042A Production of meat by each livestock type
042B Value of sales of livestock products

One would note that these items are arranged in increasing order of details on livestock theme. The choice of segregating these data items in the census and surveys will depend upon the country situation. However, in most countries, it would be appropriate ask questions relating to births, deaths, acquisition, disposal and meat production of livestock in a specialized sample surveys rather than an agricultural census. Such a survey will be needed to provide production in the livestock sector as well as for preparing inter-censual estimates of livestock census population. The table below exemplifies the treatment of livestock related information in census and/or surveys.

| Type of information | Questions/ data items | Type of survey | Periodicity/ sampling fraction |
|---|---|---|---|
| List of holdings by livestock types, livestock population by small administrative unit | Is the holding engaged in livestock production? If so, number of animals | Population census or agriculture census | decadal, complete enumeration |
| Livestock system | .Rearing system, feed system, use of vet services | Agricultural census/ Livestock survey | decadal / 5 yearly, sample |
| Livestock production | Livestock death, birth, acquisition, slaughter, milk status | Special livestock Survey | Annual/ sufficiently large sample |
| Livestock production per animal | yields of milk, meat, honey | Yield survey | 5 yearly, Thin sample |

The modular approach can also be illustrated using the example of aquaculture. One aquaculture item i.e. presence of aquaculture, is recommended for inclusion in the core census module to provide data on the number of holdings with aquaculture and to help establish a sampling frame. The following five items are recommended for enumeration by sampling methods.

- area of aquaculture by type of site
- area of aquaculture by type of production facility

- area of aquaculture by type of water
- sources of water for aquaculture
- type of aquaculture organism cultivated

In countries where there is a high incidence of aquaculture activity on agricultural holdings, it may be good to attach a special module to agricultural census and use a long-short questionnaire approach. In some other countries, it may be suitable to carryout a separate survey on aquaculture based upon the sampling frame provided by agricultural census.

The number of items in the 2010 round of agricultural censuses is larger than in previous programmes but a majority of these are optional for the agricultural census and are more suitable for thematic surveys based on the agricultural census. In developing their core census modules, countries will be encouraged to include all 16 recommended core items, plus other items required to meet national data needs or to help establish sampling frames. For example, an item on fertilizer use might be included in the core census module for sampling frame purposes if an agricultural inputs supplementary module will be required. Normally, the core census module should cover fewer items than in agricultural censuses in the past.

A country is not expected to collect all the recommended supplementary items; Instead it should select topics for supplementary modules according to national policy priorities and data needs. It is expected that countries will carry out one or two census supplementary modules in association with the core census module. Usually, a supplementary module will include data from several headings. For example, a household food security module could include data from the land, crops, and agricultural practices categories, as well as household food security items. Such surveys could be country-wide or be confined to a specific area (e.g. province) where the characteristic under study is predominant. These surveys could be implemented by the Census agency or could be carried by any other agency with a technical contribution from the Census agency.

The dividing line between the census supplementary modules and the follow-up thematic surveys is in general thin. Decision regarding inclusion of item in the census supplementary module or follow-up survey will be taken considering:

(1) the level of precision needed in the estimates
(2) the lowest administrative level at which the data is needed, and
(3) the periodicity at which the data needs to be updated.

There may also exist possibilities of implementing a specific census supplementary module only in any specific region where the questions contained in the module are more relevant. Such options need to be explored at the census-survey planning stage keeping in view resource availability, users demands and the work load for the field work.

**5. Examples of Modular Approach**

The modular approach could be operationalised as:

(1) Long and short-questionnaire, as in Thailand, or

(2) A phased Census-Survey programme based on complete enumeration and sample surveys, as in India and Niger

***Example 1:*** **Thailand** implemented short-long questionnaire approach, whereby every fourth agriculture household was selected to answer the long form. Broadly, the data item coverage of these forms was as follow:

| **Short Form** | **Long Form** (Detailed information) |
|---|---|
| Land Use: Parcel number and Total Area | Land Use: Parcel number and Total Area |
| | Livestock |
| Crop cultivated: Rice, Para rubber, Permanent Crop, Field crop, Vegetables, Forest | Crop cultivated: Rice, Pararubber, Permanent Crop, Field crop, Vegetables, Forest, Flowers and ornamental plants, herbs |
| Land tenure: Owner Land, Non owner land, Type of document of owner land | Land tenure: Owner Land, Non owner land, Type of document of owner land |
| | Fresh water culture |
| | Fertilizer and pesticide |
| | Employment on holding |
| | Machinery |
| | Household income and debt for agriculture |

It is also possible to decide on use of long or short question based upon size of the holding.

***Example 2:*** **Indian programme for agricultural census** comprises 3 phases:

Phase 1 (Listing): A listing of crop holdings is carried out using the land cadastre or door-to-door enquiry.

Phase 2 (Main Census): Information on land use, land tenure, cropping pattern and irrigation infrastructure is collected through enumeration work in 20% of the villages randomly selected. In the States and Union Territories without Land Records, a stratified two-stage sampling design is applied.

Phase 3 (Input Survey): Data on application of inputs along with details of multiple cropping, agricultural machinery, credit and livestock are collected through a stratified sub-sample selected from 7% of villages from the 20 villages where census was conducted. The stratification is based on size of holdings collected during the listing. This is conducted one year after the main census.

***Example 3:*** **Niger** collected on the 16 data items recommended for the core agricultural census as part of agriculture module of the population census. This module included questions on land use, land tenure and crop-cultivation and sedentary livestock. This provided sampling frames of most of the agricultural surveys which were implemented as a phased programme. There was a special module on Nomadic and semi-nomadic livestock.

Other thematic survey module included: Food security module, Vegetable survey module, Livestock productivity module and Annual agricultural crop production survey modules.

**6. Planning a Modular Approach**

Typically, the planning for any large scale data collection exercise should begin by taking stock of all the on-going statistical activities in a country, including efforts made by the National Statistical Office, Line Ministries and Departments, Non-governmental Organizations and even the private sector. In most cases, such an exercise reveals some duplication in efforts. Given that the funds for statistical activities are generally scarce, every effort should be made to avoid duplication in data collection efforts; also to avoid confusing the user with multiple estimates of the same characteristics. A list of surveys conducted in the country with coverage of data items in them should invariably be prepared. An integrated census-survey programme could then contribute to improvement in these surveys: by way of providing better sampling frames or a more representative sampling design.

Often, it is a good idea to make an assessment of user demand with inputs from stakeholders in agricultural census-survey programme. This assessment could be carried out through:

1. A formal User Survey to assess to expectation for different categories of data users. An example of this is attached in Annex -1.
2. A Working group of data producers and users from different categories of stake holders
3. Structured interviews of prominent users
4. Regular comments received through the web-sites and E-mail.

No matter what method is adopted for carrying out an assessment of user requirement, it is always a good idea to present the synthesised assessment in a User-Producer workshop to validate the conclusions as also to gain the support of the user community in the national plan for collection of agricultural statistics. This workshop may also serve as building partnerships in the data collection efforts by way of sharing of manpower and financial resources needed for implementation of an integrated programme for census and survey.

Usually, in situations where data sources are not developed and conduct of an agricultural census is a rare occasion, the list of data demands tends to be long and unmanageable in any single census and/or survey. At this stage it will be good to specify for each data item:

- the degree of details with which the data is required,
- the administrative level at which the data is needed,
- the acceptable level of precision in the estimates,
- the frequency/ periodicity at which the data is needed,
- the number of users a data items, and
- the use of the data.

Having obtained these details, a survey planner is in a better position to decide whether a particular data item is suitable of complete enumeration or sample enumeration. The complete enumeration items will indeed go Census supplementary Module with the Core Census module. The sampling modules could also be implemented with the census using long-shot questionnaire approach. Thematic Sample Survey which will collect very detailed data will indeed be done after the census. Examples of these surveys are: a livestock survey to estimate meat production per animal slaughtered, crop yield survey, survey of commercial vegetable producers etc.

In most country situation the statistician will be expected to provide scenarios of data collection programme along with associated costs for the whole programme for census and surveys as well as spread of expenditure. The choice of a programme will, in most cases, will be made on the basis of availability of resources. Thus it will be good to build a bit of flexibility in the programme so that it could easily expand when the resources become available. This issue is particularly relevant for availability of frames for conducting specialized surveys. For example, if a meat production survey was not foreseen, even if the budget is not immediately available, at the census planning stage, it will be costly to build a special frame for such a survey when the resources become available.

In the ultimate analysis, given the innumerable considerations and trade-offs involved in process of development of a statistical system, the task of a survey planner involves a bit of creativity. The task of developing a feasible integrated plan for census and surveys is an art rather than just the statistical principles and management of field operations.

**ANNEX D**

**.....MODEL EXAMPLE for LEARNING.....**

**Ministry of Economic Development of AFRILAND**
**Department of Statistics**

# Survey on Assessment of Needs of Users of Agricultural Statistics

## - - - Q U E S T I O N N A I R E - - -

***Dear respondent,***

***Department of Statistics of AFRILAND, supported by an International DONOR and FAO is committed to plan and implement a new system of agricultural statistics in conformity with international norms and standards and in accordance to the needs of all categories of users. To this goal we kindly ask you to fill in the questionnaire below. Besides obtaining your opinion on the present status of Agricultural Statistics and your preferences for obtaining the data, the survey aims to precisely assess the demand for various types of statistics.***

***Your feed back will provide vital input into the process of building of a new system for collection of agricultural statistics and would help the Department set priorities.***

***Thank you for your cooperation and understanding!***

## Part A: User Profile

*(Please mark by putting **X** in the box corresponding to your profile)*

**1. Nature of Organization in which you are working?** *(please mark only one box)*

- Government Department/ Ministry ☐ (1)
- Parliament ☐ (2)
- District office ☐ (3)
- Bank ☐ (4)
- Development Research Institution ☐ (5)
- Academic Institution ☐ (6)
- NGO (non-governmental organization) ☐ (7)
- International Organization/Agency ☐ (8)
- mass-media ☐ (9)
- Private company ☐ (10)

Other (please specify) ☐ (11)

2. **Nature of your duties in the organization** (mark one or several boxes, if the case)

- Policy analysis and policy formulation for development ☐ (1)
- Management and execution of projects ☐ (2)
- Identification of investment opportunities and formulation of development projects ☐ (3)
- Technology dissemination ☐ (4)
- Market management ☐ (5)
- Consulting activities ☐ (6)
- Teaching and research ☐ (7)

Other (please specify) ☐ (8)

**3. Which specific aspect of Food and Agriculture sector you are particularly concerned?** *(mark one or several boxes, if the case)*

- Crop Production ☐ (1)
- Livestock rearing ☐ (2)
- Fishing and Aquaculture ☐ (3)
- Food supply ☐ (4)
- Food safety ☐ (5)
- Crop health monitoring ☐ (6)
- Livestock health ☐ (7)
- Rural development ☐ (8)

Other (please specify) ☐ (9)

**4. What activities are your primary concerns?** *(mark one or several boxes, if the case)*

- Production process ☐ (1)
- Intermediate consumption ☐ (2)
- Post harvest storage and processing ☐ (3)
- Marketing ☐ (4)

Other (please specify) ☐ (5)

**5. How frequently do you need statistics for your work?** *(please mark only one box)*

- Almost daily ☐ (1)
- Once a week ☐ (2)
- Occasionally ☐ (3)
- Rarely ☐ (4)

**6. Which data sources you mostly rely on?** *(mark one or several boxes, if the case)*

- Department of Statistics ☐ (1)
- Ministry of Agriculture ☐ (2)
- Other ministries/government agencies ☐ (3)
- National Bank of Afriland ☐ (4)
- Local authorities ☐ (5)
- Various registers ☐ (6)
- Other administrative sources ☐ (7)
- Research inquiries/studies ☐ (8)
- International organizations ☐ (9)

Other (please specify) ☐ (10)

## Part B: General assessment of existing sources of agriculture statistics

*Please give your general opinion about current state of data on agriculture sector, particularly on the aspects linked to your activity.*

*Mark the option by inserting corresponding number in the box for each quality component*

| | | **Availability vs. needs** | **Source and evaluation** *(To be completed only if the answer in the previous column is (1)* | | | |
|---|---|---|---|---|---|---|
| | | **Do you need the data of this group and are they available to you?** | **Which is the data source you usually use?** | **How completely your needs are met?** | **How timely the data are available?** | **How reliable are the data** |
| **Major data groups** | | codes:<br>(1)- Available<br>(2)- Needed but not available<br>(3)- Not needed | codes:<br>(1)- Department of Statistics;<br>(2)- Ministry of Agriculture;<br>(3)- Other administrative sources<br>(4)- Scientific studies<br>(5)- International organizations;<br>(6)- Other sources | codes:<br>(1)- completely<br>(2)- adequately<br>(3)- only partially<br>(4)- with efforts<br>(5)- not at all | codes:<br>(1)- available on time<br>(2)- available with delay but useful<br>(3)- available when the operational need is over | codes:<br>(1)- reliable and accurate<br>(2)- workable<br>(3)- not dependable |
| | | **1** | **2** | **3** | **4** | **5** |
| **1. Holdings and holders** | 1 | | | | | |
| **2. Land operated by holdings** | 2 | | | | | |
| **3. Irrigation and water management** | 3 | | | | | |
| **4. Temporary crops** | 4 | | | | | |
| **5. Permanent crops** | 5 | | | | | |
| **6. Livestock** | 6 | | | | | |
| **7. Animal production** | 7 | | | | | |
| **8. Fertilizers** | 8 | | | | | |
| **9. Agricultural chemicals** | 9 | | | | | |
| **10. Machinery and equipment** | 10 | | | | | |
| **11. Agricultural buildings** | 11 | | | | | |
| **12. Agricultural practices** | 12 | | | | | |
| **13. Producer prices** | 13 | | | | | |
| **14. Labour force** | 14 | | | | | |
| **15. Agricultural services** | 13 | | | | | |
| **16. Use of agricultural production** | 16 | | | | | |
| **17. Cost of production** | 17 | | | | | |

## Part C: Expectations from the new integrated system on agricultural statistics

*Here we would like to find out your preferences on how to receive agricultural statistics data.*
*Please indicate your answers by inserting the appropriate code number in the cells of the table below*

| Data items | | Degree of need to your work? | Desired periodicity | Lowest administrative level at which data is needed? | Acceptable error | Acceptable time-lag |
|---|---|---|---|---|---|---|
| | | | (to be completed only if the answer in the previous column is not (5)) | | | |
| | | ***Codes:***<br>(1) Can' to without it<br>(2) Hard to do without it;<br>(3) Need it<br>(4) Good to have it<br>(5) Do not need it | ***Codes:***<br>(1) - daily<br>(2) - weekly<br>(3) - monthly<br>(4) - quarterly<br>(5) - annually<br>(6)- periodicity of more than one year<br>(7)- occasionally | ***Codes:***<br>(1) -Country<br>(2) -Region<br>(3) -District<br>(4) -Sakrebulo<br>(5) - Village | ***Codes:***<br>(1) ± 2% (2) ±5 % (3) ±10 % | ***Codes:*** (1) 1 day<br>(2) 5 days<br>(3) 10 days<br>(4) 1 month<br>(5) 2 months<br>(6) a quarter<br>(6) half a year<br>(7) a year<br>(8) more than one year |
| | | **1** | **2** | **3** | **4** | **5** |
| **1. Holdings and holders** | | | | | | |
| - Distributions by various characteristics | 11 | | | | | |
| - Holder characteristics | 12 | | | | | |
| - Other (please specify)..... | 13 | | | | | |
| **2. Area operated by holdings** | | | | | | |
| - By land tenure | 21 | | | | | |
| - By land use | 22 | | | | | |
| - Other (please specify)..... | 14 | | | | | |
| **3. Irrigation and water management** | | | | | | |
| - Area irrigated by crops | 31 | | | | | |
| - Area irrigated by method of Irrigation | 32 | | | | | |
| - Source of irrigation water | 33 | | | | | |
| - Payment terms for irrigation Water | 34 | | | | | |
| - Drainage | 35 | | | | | |
| - Other (please specify)..... | 36 | | | | | |
| **4. Temporary crops (by each individual crop)** | | | | | | |
| - Area sown/ harvested | 41 | | | | | |
| - Harvest | 42 | | | | | |
| - Yield | 43 | | | | | |
| - Ongoing information on Sowing | 45 | | | | | |
| - Ongoing information on Harvesting | 46 | | | | | |
| - Production forecasts | 48 | | | | | |
| - Other (please specify)..... | 49 | | | | | |
| **5. Permanent crops (by each individual crop)** | | | | | | |

| Data items | | Degree of need to your work? | Desired periodicity | Lowest administrative level at which data is needed? | Acceptable error | Acceptable time-lag |
|---|---|---|---|---|---|---|
| | | | (to be completed only if the answer in the previous column is not (5)) | | | |
| | | ***Codes:***<br>(1) Can' to without it<br>(2) Hard to do without it;<br>(3) Need it<br>(4) Good to have it<br>(5) Do not need it | ***Codes:***<br>(1) - daily<br>(2) - weekly<br>(3) - monthly<br>(4) - quarterly<br>(5) - annually<br>(6)- periodicity of more than one year<br>(7)- occasionally | ***Codes:***<br>(1) -Country<br>(2) -Region<br>(3) -District<br>(4) -Sakrebulo<br>(5) - Village | ***Codes:***<br>(1) ± 2% (2) ±5 % (3) ±10 % | ***Codes:*** (1) 1 day<br>(2) 5 days<br>(3) 10 days<br>(4) 1 month<br>(5) 2 months<br>(6) a quarter<br>(6) half a year<br>(7) a year<br>(8) more than one year |
| | | **1** | **2** | **3** | **4** | **5** |
| - Area of compact plantations | 51 | | | | | |
| - Number of trees | 52 | | | | | |
| - Harvest | 53 | | | | | |
| - Yield | 54 | | | | | |
| - Nurseries | 55 | | | | | |
| - Planting/felling of trees | 56 | | | | | |
| - Ongoing information on harvesting | 57 | | | | | |
| - Production forecasts | 58 | | | | | |
| - Other (please specify)..... | 59 | | | | | |
| **6. Livestock ( by each individual specie)** | | | | | | |
| - Total numbers | 61 | | | | | |
| - Numbers by age and sex (big animals) | 62 | | | | | |
| - Numbers by breeds | 63 | | | | | |
| - Numbers by purpose | 64 | | | | | |
| - Births, felling, acquisitions, disposals | 65 | | | | | |
| - Feed type used | 66 | | | | | |
| - Weight gaining | 67 | | | | | |
| - Veterinary services | 68 | | | | | |
| - Other (please specify)..... | 69 | | | | | |
| **7. Animal Production (by each individual specie)** | | | | | | |
| - Production | 71 | | | | | |
| - Productivity | 72 | | | | | |
| - Other (please specify)..... | 73 | | | | | |
| **8. Fertilizers (by crops)** | | | | | | |
| - Area fertilized | 81 | | | | | |

| Data items | | Degree of need to your work? | Desired periodicity | Lowest administrative level at which data is needed? | Acceptable error | Acceptable time-lag |
|---|---|---|---|---|---|---|
| | | | (to be completed only if the answer in the previous column is not (5)) | | | |
| | | ***Codes:***<br>(1) Can' to without it<br>(2) Hard to do without it;<br>(3) Need it<br>(4) Good to have it<br>(5) Do not need it | ***Codes:***<br>(1) - daily<br>(2) - weekly<br>(3) - monthly<br>(4) - quarterly<br>(5) - annually<br>(6)- periodicity of more than one year<br>(7)- occasionally | ***Codes:***<br>(1) -Country<br>(2) -Region<br>(3) -District<br>(4) -Sakrebulo<br>(5) - Village | ***Codes:***<br>(1) ± 2% (2) ±5 % (3) ±10 % | ***Codes:*** (1) 1 day<br>(2) 5 days<br>(3) 10 days<br>(4) 1 month<br>(5) 2 months<br>(6) a quarter<br>(6) half a year<br>(7) a year<br>(8) more than one year |
| | | **1** | **2** | **3** | **4** | **5** |
| - Area fertilized by crops | 82 | | | | | |
| - Quantities used | 83 | | | | | |
| - Quantities used by crops | 84 | | | | | |
| - Other (please specify)..... | 85 | | | | | |
| **9. Agricultural chemicals ( by crops)** | | | | | | |
| - Area treated | 91 | | | | | |
| - Area treated by crops | 92 | | | | | |
| - Quantities used | 93 | | | | | |
| - Quantities used by crops | 94 | | | | | |
| - Other (please specify)..... | 95 | | | | | |
| **10. Agricultural machinery (by type of activities, except tractors )** | | | | | | |
| - Quantities by capacity | 101 | | | | | |
| - Quantities by age and working condition | 102 | | | | | |
| - Use of machinery and equipment | 103 | | | | | |
| - Use of machinery and equipment by source | 104 | | | | | |
| - Other (please specify)..... | 105 | | | | | |
| **11. Agricultural buildings (by type, for vegetal production and animals)** | | | | | | |
| - Quantities | 111 | | | | | |
| - Capacity | 112 | | | | | |
| - Types of tenure | 113 | | | | | |
| - Other (please specify)..... | 114 | | | | | |
| **12Agricultural practices** | | | | | | |
| - Incidence of a practice | 121 | | | | | |
| - Other (please specify)..... | 122 | | | | | |
| **13. Producer prices in agriculture (by type of agricultural products)** | | | | | | |

| Data items | | Degree of need to your work? | Desired periodicity | Lowest administrative level at which data is needed? | Acceptable error | Acceptable time-lag |
|---|---|---|---|---|---|---|
| | | | (to be completed only if the answer in the previous column is not (5)) | | | |
| | | ***Codes:***<br>(1) Can' to without it<br>(2) Hard to do without it;<br>(3) Need it<br>(4) Good to have it<br>(5) Do not need it | ***Codes:***<br>(1) - daily<br>(2) - weekly<br>(3) - monthly<br>(4) - quarterly<br>(5) - annually<br>(6)- periodicity of more than one year<br>(7)- occasionally | ***Codes:***<br>(1) -Country<br>(2) -Region<br>(3) -District<br>(4) -Sakrebulo<br>(5) - Village | ***Codes:***<br>(1) ± 2% (2) ±5 % (3) ±10 % | ***Codes:*** (1) 1 day<br>(2) 5 days<br>(3) 10 days<br>(4) 1 month<br>(5) 2 months<br>(6) a quarter<br>(6) half a year<br>(7) a year<br>(8) more than one year |
| | | **1** | **2** | **3** | **4** | **5** |
| - Prices | 131 | | | | | |
| - Indices | 132 | | | | | |
| - Other (please specify)..... | 133 | | | | | |
| **14. Labour force** | | | | | | |
| - Number of workers | 141 | | | | | |
| - Hours worked | 142 | | | | | |
| - Other (please specify)..... | 143 | | | | | |
| **15 Agricultural services (by type of activities)** | | | | | | |
| - Access to the service | 151 | | | | | |
| - Use of the service | 152 | | | | | |
| - Sources of the service | 153 | | | | | |
| - Other (please specify)..... | 154 | | | | | |
| **16. Use of agricultural production (by agricultural products)** | | | | | | |
| - In physical terms | 161 | | | | | |
| - In value terms | 162 | | | | | |
| - Other (please specify)..... | 163 | | | | | |
| **17. Cost of production** | | | | | | |
| - Intermediate consumption | 171 | | | | | |
| - Capital expenditure | 172 | | | | | |
| - Taxes | 173 | | | | | |
| - Other (please specify)..... | 174 | | | | | |
| **18. Please insert in the left column below any other general or specific items not covered above which you need in your work, and fill in the table for them** | | | | | | |
| | | | | | | |
| | | | | | | |
| | | | | | | |
| | | | | | | |

| Data items | | Degree of need to your work? | Desired periodicity | Lowest administrative level at which data is needed? | Acceptable error | Acceptable time-lag |
|---|---|---|---|---|---|---|
| | | | (to be completed only if the answer in the previous column is not (5)) | | | |
| | | ***Codes:***<br>(1) Can' to without it<br>(2) Hard to do without it;<br>(3) Need it<br>(4) Good to have it<br>(5) Do not need it | ***Codes:***<br>(1) - daily<br>(2) - weekly<br>(3) - monthly<br>(4) - quarterly<br>(5) - annually<br>(6)- periodicity of more than one year<br>(7)- occasionally | ***Codes:***<br>(1) -Country<br>(2) -Region<br>(3) -District<br>(4) -Sakrebulo<br>(5) - Village | ***Codes:***<br>(1) ± 2% (2) ±5 % (3) ±10 % | ***Codes:*** (1) 1 day<br>(2) 5 days<br>(3) 10 days<br>(4) 1 month<br>(5) 2 months<br>(6) a quarter<br>(6) half a year<br>(7) a year<br>(8) more than one year |
| | | **1** | **2** | **3** | **4** | **5** |
| | | ☐ | ☐ | ☐ | ☐ | ☐ |

## Part D. Preferences for receiving data

*(Please mark by putting* **X** *in the appropriate box/boxes)*

**D1 How would you like to receive the data?** *(please mark only one box)*

Printed format ☐ (1) Electronic formats like CDs ☐ (2)

Both in electronic and printed format ☐ (3)

(if answer is (1) go to d3)

**D2. In which form would you prefer to receive the information about the electronic format?** *(mark one or several boxes, if the case)*

Electronic formats (floppy, CDs) ☐ (1) Through periodic E-mails ☐ (2)

From internet ☐ (3) Other ( please specify) ☐ (4)

**D3. In which form would you like to receive the data?** (mark one or several boxes, if the case)

Numbers, Tables ☐ (1) Graphs, charts ☐ (2)

Maps ☐ (3) Other ( please specify) ☐ (4)

**D4. Would you prefer to receive the above with or without analysis of results that facilitate decision making?** *(please mark only one box)*

With ☐ (1) Without ☐ (2)

**Part E. Suggestions on mechanism for involvement of users**

*(Please mark one or more alternatives you prefer by putting* **X** *in the appropriate box/boxes)*

| | |
|---|---|
| Discuss the programme on agriculture statistics at meetings where representatives of users take part | ☐ (1) |
| Create special working groups for agricultural statistics | ☐ (2) |
| Use frequently opinion survey for the users of statistics on agriculture | ☐ (3) |
| Insert short questionnaire in statistical publications on agriculture to be filled in by users | ☐ (4) |
| Other ( please specify) | ☐ (5) |

**Part F. Please expose your general views about the future system of agricultural statistics in Afriland, specifically on the following topics:**

a) Coordination and integration of the system among various institutions

_____

_____

b) Sharing of human and financial resources for data collection and processing of the main agricultural statistics works

_____

_____

c) Which part of the system you consider possible to take over on your charge?

_____

_____

d) Which kind of administrative sources that can be used for statistical purposes you consider appropriate to develop/maintain/improve?

_____

_____

**Thank you for the collaboration!**

## Group Exercises on Modular Approaches to Census and Survey Program

**Learning Objective**

*To be able to assess the user requirements, prioritize them and suggest a cost-effective feasible plan for collection of agricultural data through census and surveys.*

**Activities**

**First 60 minutes**

Divide the group in 2 parts: Data Providers (5 persons) and Data Users (10 persons). Each person in the group assumes the role of representatives of one of the stake holder listed in Annex-1.

**Imagine** that you all belong to a specific country (AFRILAND) which is planning to carry out its long over due agricultural census. The National Statistics Office (NSO) acknowledges the urgent need for an agricultural census as the ongoing sample surveys are not able to provide reliable estimates. However, it neither has the financial nor the technical capacity to plan and execute the census. This census will form the basis of future development of current agriculture statistics. Many stakeholders have high expectations from this exercise. Some development agencies are looking forward to the census data for planning their development projects (Please feel free to imagine one or two such projects).

The NSO has prepared a budget estimate of US$ 4.00 million for a comprehensive agricultural census by complete enumeration of all households and enterprises. The total funds available for the census are not even half of this amount. Some donors are, however, willing to fund up to US$ 2.00 million provided the census meets their key data requirements for international comparison and national project planning.

It is almost certain that a Population Census will be conducted shortly before the planned date of agricultural census. The cartographic work of listing of enumeration areas has already been started.

**Assignment**

To come-up with feasible plan of agricultural census this satisfies most of the stakeholders in the light of above scenario.

**Steps:**

**Step I:** Prepare a short list of data items which are relevant in the context of AFRILAND using an exhaustive list of themes and data items (Annex -2) preloaded on your computer. Use the preloaded file **Group Activity Technical Session 2**. First select themes and then a manageable list of data items for the activity.

**Step II:** Data Provider Group conducts a mock Workshop to discuss the short listed data items with Data Users Group to understand their expectations in terms of:

- Administrative Level (country, province, district, village) at which a specific data is needed
- Level of accuracy expected
- Periodicity
- Urgency in data requirement/other sources of data
- How data will be used?

Synthesize the inputs from Data Users in the excel sheet using codes (options) given on top of the Columns C, D, E and F.

**The group of 15 persons is split in two sub-groups having representation of both Data Providers and Data Users. Each of the two Groups uses the same short-list of data items and follows on to the exercise below.**

**Next 60 minutes**

**STEP-III:** Each of the two sub-groups uses the same out put from STEP-II. The whole sub-group now models as

Based upon the inputs of Data Users and the technical judgement of the group fills the last two columns of the sheet, basically deciding the suitability of data items for sample survey or census.

**STEP-IV:** Data providers prepare a cost-effective plan of agriculture Census and Surveys. The plan should *inter alia* indicate:

- How the plan takes advantage of activities of the Population Census
- The strategy for agricultural census: sampling and/or complete enumeration
- Different modules for census and surveys: data item in each module
- Time schedule for implementing each module.
- Linkages between the modules.
- Length of the module, in terms of response time

The group is free to imagine the following numbers for preliminary evaluation of cost implication of their census-survey plan.

- Number of Households in country
- Number of agricultural holdings of different types
- Unit cost for visiting a household etc.

**Final Presentation (60 minutes)** (4 group presentations)

Data providers present their plans to the audience comprising Heads of Departments in the user Ministries and Donor organizations. (Roles may be pre-assigned)

## Annex E

### List of Institutions for Selection of Role Models in Group Activity

**Data Provides** (National agency responsible for Agricultural Census and Surveys)

1. Methodological Unit
2. Field Operation Unit
3. Data processing Unit
4. Data Dissemination
5. User Coordination Unit

**Data User** (Stake holders)

1. Department of crops
2. Department of Livestock Development
3. Department of Fisheries and aquaculture
4. Ministry of Irrigation and water resources
5. Ministry of Fertilizers
6. Ministry of Food Processing Industries
7. Ministry of Rural Development
8. Ministry of Local Self Government
9. Association of Meat Exporters
10. Farmers' Association
11. Association of Agricultural Enterprises/ Cooperatives
12. Agriculture Exporter's Association
13. International Agencies and donor's representatives

## Annex F

**Items recommended for international comparison**[3]

0001 Identification and location of agricultural holding
0002 Legal status of agricultural holder
0003 Sex of agricultural holder
0004 Age of agricultural holder
0005 Household size
0006 Main purpose of production of the holding
0007 Area of holding according to land use types
0008 Total area of holding
0009 Land tenure types on the holding
0010 Presence of irrigation on the holding
0011 Types of temporary crops on the holding
0012 Types of permanent crops on the holding and whether in compact plantations.
0013 Number of animals on the holding for each livestock type
0014 Presence of aquaculture on the holding
0015 Presence of forest and other wooded land on the holding
0016 Other economic production activities of the holding's enterprise

**Items for consideration**
**(The group is free to add important new items, as required)**

Theme 01 – Land
*(Reference group: holdings with land in Item 0008)*

*For each parcel*
0101 Location
0102 Area
0103 Land tenure
0104 *(For rented parcels)* Terms of rental
0105 Presence of shifting cultivation
0106 Number of years since cleared

*For the holding*
0111 Presence of soil degradation: type and degree

Theme 02 – Irrigation and water management
*(Reference groups: Items 0201–0205 – holdings with irrigation in Item 0010; Item 0206 – holdings with temporary crops or permanent crops in Items 0011 and 0012; Item 0207 – holdings with land in Item 0008).*

*For the holding*
0201 Area of land irrigated according to land use type
0202 Area irrigated according to method of irrigation
0203 Area irrigated for each crop type
0204 Sources of irrigation water
0205 Payment terms for irrigation water
0206 Other types of water management practices
0207 Presence of drainage equipment

---

[3] If the agricultural census is conducted in conjunction with the aquacultural census, an additional core item on area of aquaculture by type of site is included.

Theme 03 – Crops
*(Reference groups: Items 0301–0303 – holdings with temporary crops in Item 0011; Items 0311–0314 – holdings with permanent crops in Item 0012; Items 0321–0327 – holdings with temporary crops or permanent crops in Items 0011 and 0012).*

*For each temporary crop type*
0301 Area of temporary crops harvested
0302 *(For selected crop types)* Area of temporary crops harvested according to end-use
0303 *(For selected crop types)* Production of temporary crops harvested

*For each permanent crop type*
0311 Area of productive and non-productive permanent crops in compact plantations
0312 *(For tree crops)* Number of permanent crop trees in compact plantations and scattered plantings
0313 *(For selected crop types)* Area of productive permanent crops in compact plantations according to end-use
0314 *(For selected crop types)* Production of permanent crops

*For the holding*
0321 Area of land used to grow temporary crops as a secondary land use
0322 Use of each type of fertilizer
0323 Area fertilized for each type of fertilizer and major crop type
0324 Source of seed inputs for each major crop type
0325 Type of seed for each major crop type
0326 Area of nurseries
0327 Area of cropped land under protective cover

Theme 04 – Livestock
*(Reference group: holdings with livestock in Item 0013)*

*For the holding*
0401 Type of livestock production system
0402 Use of veterinary services

*For each livestock type*
0411 Number of animals: age and sex
0412 Number of animals according to purpose
0413 Number of milking animals according to milk status
0414 Number of animals born
0415 Number of animals acquired
0416 Number of animals slaughtered
0417 Number of animals disposed of
0418 Number of animals died from natural causes
0419 Types of feed

Theme 05 – Agricultural practices
*(Reference group: all holdings)*

*For the holding*
0501 Use of agricultural chemicals
0502 Use of good agricultural practices
0503 Use of organic agricultural practices
0504 Use of genetically modified crops according to crop type
0505 Selected machinery and equipment used on the holding according to source
0506 Non-residential buildings according to use

0507 Percentage of each major agricultural product sold

Theme 06– Agricultural services
*(Reference group: holdings in sector "single-holding household" in Item 0002)*

*For the holding*
0601 Receipt of credit for agricultural purposes
0602 Source of credit
0603 Type of collateral for credit
0604 Period of loan or credit
0605 Sources of agricultural information
0606 Sources of agricultural extension services
0607 Travelling time to nearest periodic or permanent agricultural produce market

Theme 07 – Demographic and social characteristics
*(Reference group: holdings in sector "single-holding household" in Item 0002)*

*For the holding*
0701 Whether holding is part of an agricultural household
0702 National/ethnic group of household head or agricultural holder

*For each household member*
0711 Sex.
0712 Age
0713 Relationship to household head or other reference person
0714 Marital status
0715 Educational attainment

Theme 08 – Farm labour
*(Reference group: Items 0801–0814 – holdings in sector "single-holding household" in Item 0002; Items 0821–0823 – all holdings)*

*For each household member of working age*
0801 Activity status

*For each economically active household member*
0811 Status in employment of main job
0812 Occupation of main job
0813 Time worked in main job
0814 Time worked on the holding

*For the holding*
0821 Number of employees on the holding: time worked and sex
0822 Form of payment for employees
0823 Use of contractors for work on the holding according to type of contract

Theme 09 – Household food security
*(Reference group: holdings in sector "single-holding household" in Item 0002)*

*For the household*
0901 (a) Whether household members could not afford to eat what they normally eat at any time during a twelve-month reference period.
(b) Months in which food shortage occurred.

(c) Reasons for food shortage.
(d) How the household's eating patterns were affected by food shortage.
(e) Steps taken to alleviate food shortage.

0902 Whether the household fears a food shortage during a future twelve-month reference period
0903 Frequency of normally eating selected food products
0904 Effects of natural disasters
0905 Extent of loss of agricultural output due to natural disasters

*For children aged under 5 years*
0911 Height and weight

Theme 10 – Aquaculture
*(Reference group: holdings with aquaculture in Item 0014)*

*For the holding*
1001 Area of aquaculture according to type of site
1002 Area of aquaculture according to type of production facility
1003 Type of water
1004 Sources of water for aquaculture
1005 Type of aquacultural organism cultivated

Theme 11 – Forestry
*(Reference group: holdings with forest and other wooded land in Item 0015)*

*For the holding*
1101 Area of forest and other wooded land as primary land use
1102 Area of forest and other wooded land as a secondary land use on agricultural land
1103 Main purpose of forest and other wooded land
1104 Whether agro-forestry is practised

Theme 12 – Management of the holding
*(Reference group: holdings in sector "single-holding household" in Item 0002)*

*For each holding*
1201 Identification of sub-holdings
1202 Identification of sub-holders

*For each sub-holding*
1211 Sex of sub-holder
1212 Age of sub-holder
1213 Area of crops managed for each crop group
1214 Number of livestock managed for each livestock group

# SESSION 3

## SAMPLING FRAME
## AREA FRAME FOR AGRICULTURAL SURVEY

# General presentation on Building, Using and Maintaining Sampling Frames: Area Frame, Multiple Frame, Master Sampling Frames

*Terry Holland, USDA/NASS*

## Building & Using Area Sampling Frames for Agricultural Censuses & Surveys

presented by
Theresa "Terry" Holland
National Agricultural Statistics Service
U.S. Department of Agriculture

Regional Workshop on Sampling
for Census of Agriculture 2010
& Agricultural Surveys

20 April 2010
Accra, Ghana

**Definitions**

### sampling frame:

- a means by which a **target population** may be sampled
- a list of all **sampling units** & a set of rules for identifying **population units**

## target population

all the items (people, farms, animals, businesses, etc.) about which information is needed

## sampling units

well-defined units that allow access to the target population

## population units

individual elements of the target population



## List frames:

farmers

agri-businesses

fields or orchards

## Area frame:

segments of land

## NASS List Frame

***What is it?***

data to identify, locate & contact farmers & agri-businesses

name
address
telephone number
state, district & county
Social Security Number
Employer Identification Number

data about the farm/business

total acres
individual crop acres
grain storage capacity
peak livestock inventories
peak number of hired workers
⋮

## NASS List Frame

***How is it constructed?***

sources for new names & data

growers organizations
farm program lists
state & local tax records
state & local license records
lists from other federal, state & local agencies
newspaper & magazine articles

sources for updating names & data

on-going NASS surveys
Census of Agriculture

## NASS List Frame

***How is it used?***

| | |
|---|---|
| ❶ classify | identify farmers &/or agri-businesses likely to have item(s) of interest |
| ❷ stratify | group similar units together based on size or amount of item(s) to be measured |
| ❸ sample | select units from each group |
| ❹ survey | collect data for selected units |
| ❺ summarize | expand data using probabilities of selection |

## NASS List Frame

**Illinois – Quarterly Crops/Stocks Surveys:**

| stratum | boundaries | population | sample size | sampling interval | percent in sample |
|---|---|---|---|---|---|
| 62 | capacity 1 - 9,999 | 6387 | 100 | 63.9 | 2 |
| 65 | cropland 200 - 599 | 7221 | 210 | 34.4 | 3 |
| 66 | capacity 10,000 - 49,999 | 11231 | 400 | 28.1 | 4 |
| 72 | cropland 600 - 2,499 | 7627 | 500 | 15.3 | 7 |
| 73 | sorghum 1+ | 2495 | 200 | 12.5 | 8 |
| 78 | capacity 50,000 - 499,999 | 5912 | 550 | 10.7 | 9 |
| 79 | cropland 2,500 - 5,499 | 474 | 100 | 4.7 | 21 |
| 95 | cropland 5,500+ | 29 | 29 | 1.0 | 100 |
| 97 | capacity 500,000+ | 23 | 23 | 1.0 | 100 |
| | total | 41399 | 2112 | | |

*1997*

**Illinois – Quarterly Hogs Surveys:**

| stratum | boundaries | population | sample size | sampling interval | percent in sample |
|---|---|---|---|---|---|
| 80 | hogs 1 - 99 | 1711 | 70 | 24.4 | 4 |
| 82 | hogs 100 - 499 | 1138 | 220 | 5.2 | 19 |
| 84 | hogs 500 - 999 | 366 | 225 | 1.6 | 61 |
| 86 | hogs 1,000 - 1,999 | 289 | 255 | 1.1 | 88 |
| 88 | hogs 2,000 - 2,999 | 132 | 125 | 1.1 | 95 |
| 90 | hogs 3,000 - 4,999 | 116 | 116 | 1.0 | 100 |
| 92 | hogs 5,000 - 14,999 | 116 | 116 | 1.0 | 100 |
| 98 | hogs 15,000+ | 28 | 28 | 1.0 | 100 |
| | total | 3896 | 1155 | | |

2005



## Strengths:

- ✓ can use inexpensive data collection methods (mail, telephone)
- ✓ can target specific or rare commodities
- ✓ can reduce variability due to sampling
- ✓ cost efficient

## Weaknesses:

- ✗ does not cover entire population
- ✗ goes out-of-date quickly
- ✗ increased non-sampling errors due to data collection methods
- ✗ requires on-going maintenance
  - build
  - update
  - remove duplication
  - remove out-of-scope records



***Coverage:***

| | |
|---|---|
| **Number of Farms** | **70%** |
| *by Value of Sales (June 2008)* | |
| **$100,000+** | **93%** |
| **$10,000-99,999** | **82%** |
| **$1,000-9,999** | **58%** |
| *by Type of Farm (June 2008)* | |
| **Crops** | **89%** |
| **Livestock** | **71%** |
| **Specialty** | **64%** |
| *by Commodity* | |
| **Corn** *(June 2009)* | **93%** |
| **Soybeans** *(June 2009)* | **92%** |
| **Winter Wheat** *(June 2009)* | **92%** |
| **Hogs** *(December 1, 2008)* | **98%** |
| **Cattle** *(January 1, 2009)* | **89%** |
| **Land in Farms** | **91%** |

## Sampling Techniques:

- Simple Random Sampling (SRS)
- Systematic Sampling
- Stratified Sampling
- Probability Proportional to Size (PPS)
- Multivariate Probability Proportional to Size (MPPS)
- Permanent Random Number (PRN)



***What is it?***

land area of the U.S
divided into segments
using physical boundaries

associate farms, crops, animals, etc.
with land inside the segments

***How is it constructed?***

using… satellite imagery
digital maps
GIS software
aerial photography

❶ divide land area into strata based on land use & likelihood of finding agriculture

❷ subdivide land use strata into strata blocks

❸ select a sample of strata blocks

❹ subdivide selected strata blocks into segments



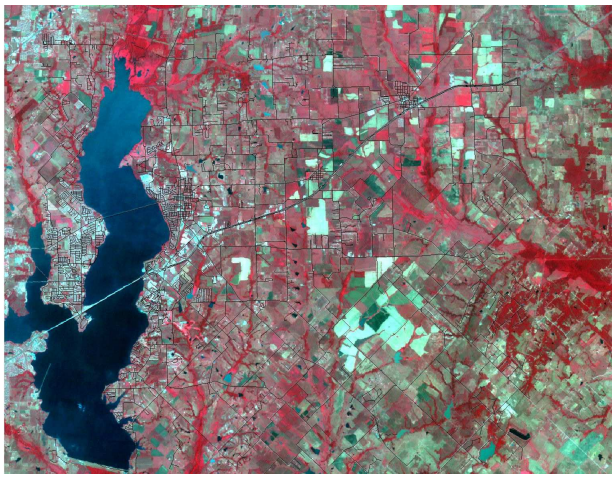
## General Land Use Categories

| | |
|---|---|
| general cropland | 75% or more cultivated |
| general cropland | 50-74% cultivated |
| general cropland | 15-49% cultivated |
| agri-urban | less than 15% cultivated, residential mixed with agriculture |
| range & pasture | less than 15% cultivated |
| residential & commercial | no cultivation |
| non-agricultural | |
| water | |

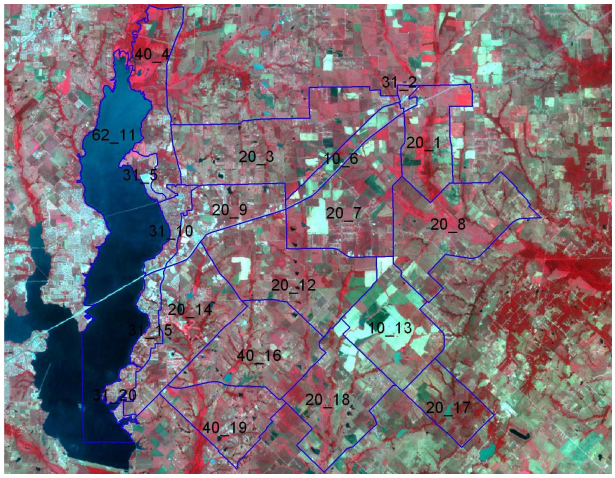
***Satellite imagery:***

NASS Area Frame

***Digital map features:***

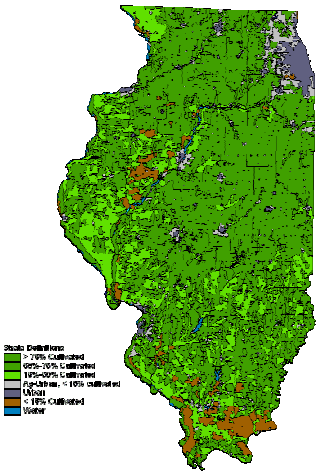


***Strata blocks - primary sampling units (PSUs):***

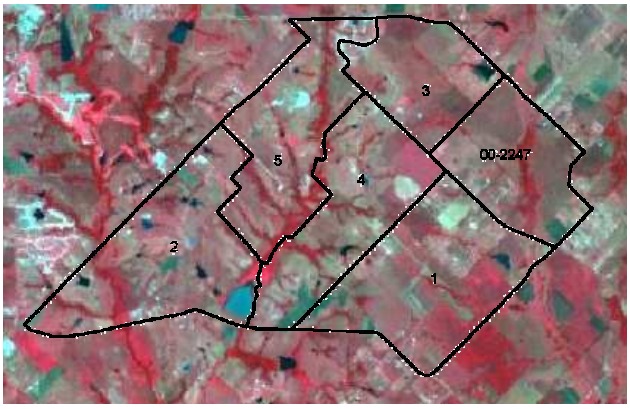



NASS Area Frame

***Land use stratification for Illinois:***

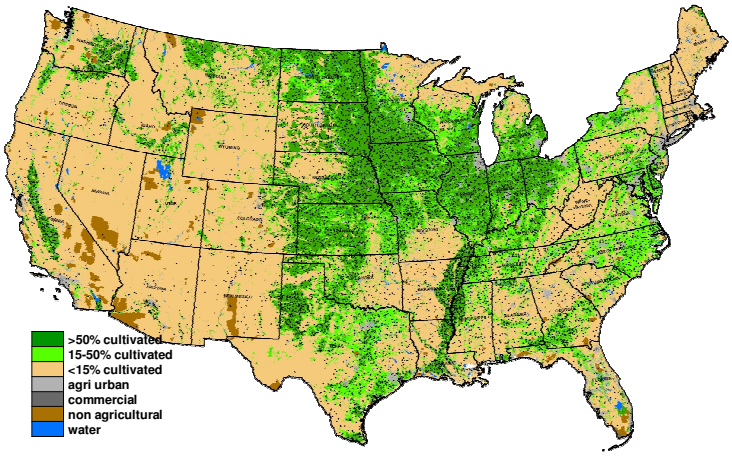


***Selected PSU & segment:***






***Land Use Strata & Sampled Segments:***



### *How is it used?*

❶ sample — select a sample of segments – generally keep segments in sample for 5 years, rotate 20% of sample each year

❷ survey — account for all land & animals inside segment boundaries, obtain information about all farms with land inside segments

❸ summarize — expand data using probabilities of selection (based on land area)



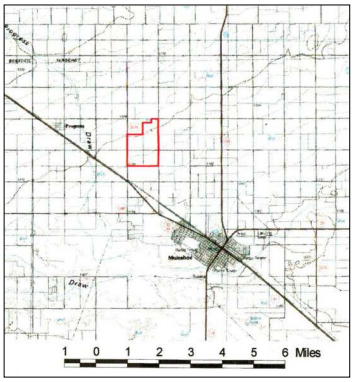
**Illinois – Area Sample Design (2006)**

| stratum | boundaries | total land $mi^2$ | segment size | total number of segments | number of sampled segments | expansion factor |
|---|---|---|---|---|---|---|
| 11 | >75% cultivated | 30923 | 1.00 | 30936 | 250 | 124 |
| 12 | 51-75% cultivated | 8513 | 1.00 | 8512 | 70 | 122 |
| 20 | 25-50% cultivated | 10834 | 1.00 | 10836 | 50 | 217 |
| 31 | agri-urban: >100 homes/$mi^2$ | 2681 | 0.25 | 10718 | 10 | 1072 |
| 32 | commercial: >100 homes/$mi^2$ | 676 | 0.10 | 6768 | 4 | 1692 |
| 40 | <25% cultivated | 1984 | 1.00 | 1981 | 15 | 132 |
| 50 | non-agricultural | 216 | pps | 53 | 2 | 27 |
| | total | 55827 | | 69804 | 401 | |

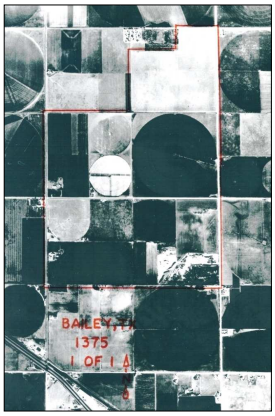
2007

## *Data Collection Tools:*

**Road map**

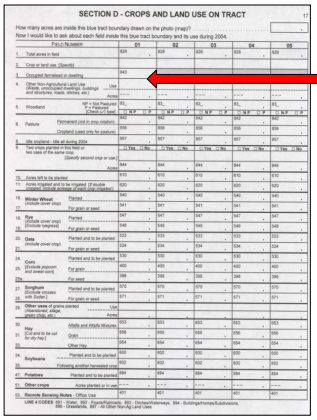



**Aerial photo**






## *Data collection:*



SECTION D - CROPS AND LAND USE ON TRACT

How many acres are inside this blue tract boundary drawn on the photo (map)? ____

Now I would like to ask about each field inside this blue tract boundary and its use during 2004.

| Field Number | 01 | 02 | 03 | 04 | 05 |
|---|---|---|---|---|---|
| 1. Total acres in field | 629 | 629 | 629 | 629 | 629 |
| 2. Crop or land use (Specify) | | | | | |
| 3. Occupied farmstead or dwelling | 343 | | | | |
| 4. Other Non-Agricultural Land Use (Waste, unoccupied dwellings, barnyards and structures, roads, ponds, etc.) — Use | | | | | |
| Acres | --- | --- | --- | --- | --- |
| 5. Woodland — NP = Not Pastured, P = Pastured, C = Cut for Feed | NL | NL | NL | NL | NL |
| | ☐ N.P ☐ P | ☐ N.P ☐ P | ☐ N.P ☐ P | ☐ N.P ☐ P | ☐ N.P ☐ P |
| 6. Pasture — Permanent (not in crop rotation) | 342 | 342 | 342 | 342 | 342 |
| Cropland used only for pasture | 356 | 356 | 356 | 356 | 356 |
| 8. Idle cropland – idle all during 2004 | 357 | 357 | 357 | 357 | 357 |
| 9. Two crops planted in this field or two uses of the same crop. (Specify second crop or use.) | ☐ Yes ☐ No | ☐ Yes ☐ No | ☐ Yes ☐ No | ☐ Yes ☐ No | ☐ Yes ☐ No |
| Acres | 344 | 344 | 344 | 344 | 344 |
| 10. Acres left to be planted | 610 | 610 | 610 | 610 | 610 |
| 11. Acres irrigated (Acres to be irrigated. If double cropped, include acreage of each crop irrigated.) | 320 | 320 | 320 | 320 | 320 |
| 16. **Winter Wheat** (Include cover crop) — Planted | 540 | 540 | 540 | 540 | 540 |
| For grain or seed | 541 | 541 | 541 | 541 | 541 |
| 18. **Rye** (Include cover crop) — Planted | 547 | 547 | 547 | 547 | 547 |
| 19. **Barley** (Include targeted) — For grain or seed | 548 | 548 | 548 | 548 | 548 |
| 20. **Oats** (Include cover crop) — Planted and to be planted | 533 | 533 | 533 | 533 | 533 |
| For grain or seed | 534 | 534 | 534 | 534 | 534 |
| 24. **Corn** — Planted and to be planted | 550 | 550 | 550 | 550 | 550 |
| 25. Corn (Include popcorn and sweet corn) — For grain | 400 | 400 | 400 | 400 | 400 |
| For seed | 398 | 398 | 398 | 398 | 398 |
| 27. **Sorghum** (Include crosses with Sudan) — Planted and to be planted | 570 | 570 | 570 | 570 | 570 |
| For grain or seed | 571 | 571 | 571 | 571 | 571 |
| 29. **Other uses** of grains planted (Intercrop, silage, popcorn, etc.) — Use | | | | | |
| Acres | | | | | |
| 30. **Hay** — Alfalfa and Alfalfa Mixtures | 553 | 553 | 553 | 553 | 553 |
| 31. (Cut and to be cut for dry hay) — Grass | 558 | 558 | 558 | 558 | 558 |
| 32. Other Hay | 554 | 554 | 554 | 554 | 554 |
| 34. **Soybeans** — Planted and to be planted | 555 | 555 | 555 | 555 | 555 |
| Following another harvested crop | 552 | 552 | 552 | 552 | 552 |
| 47. **Potatoes** — Planted and to be planted | 584 | 584 | 584 | 584 | 584 |
| 51. **Other crops** — Acres planted or in use | --- | --- | --- | --- | --- |
| 53. **Remote Sensing Notes** - Office Use | 401 | 401 | 401 | 401 | 401 |

LINE 4 CODES: 551 - Water, 552 - Roads/Railroads, 553 - Ditches/Waterways, 554 - Buildings/Homes/Subdivisions, 555 - Orchards/Vineyards, 557 - All Other Non-Ag Land Uses

***"Segment" sample estimators:***

$$y_{hjkm} = w_{hjkm} x_{hjkm}$$

$$\hat{y} = \sum_h \sum_j \sum_k e_{hjk} \sum_m y_{hjkm}$$

$$e_{hjk} \approx \frac{N_h}{n_h}$$

$h$ = land use stratum
$j$ = stratum block within stratum $h$
$k$ = segment within stratum block $j$
$m$ = farming operation within segment $k$
$e_{hjk}$ = expansion factor for selected segment $k$
$w_{hjkm}$ = weight for farming operation $m$
$x_{hjkm}$ = survey value for farming operation $m$
$N_h$ = number of possible segments in stratum $h$
$n_h$ = number of segments sampled in stratum $h$

NASS Area Frame

***"Segment" sample estimators:***

## Closed

$w_{hjkm}$ = 1
$x_{hjkm}$ = value of item within segment only

## Open

$w_{hjkm}$ = 1 if farmer resides in segment, 0 otherwise
$x_{hjkm}$ = value of item for entire farming operation

## Weighted

$w_{hjkm}$ = percent of total farm area within the segment
$x_{hjkm}$ = value of item for entire farming operation

## Strengths:

- ✓ complete coverage
- ✓ reduced non-sampling errors
- ✓ estimates well for commonly produced commodities
- ✓ versatility
- ✓ longevity

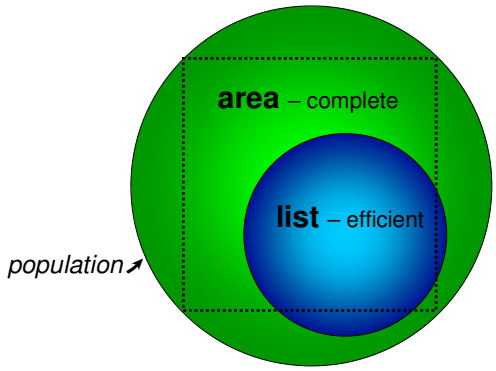


## Weaknesses:

- ✗ expensive (frame construction & data collection)
- ✗ difficult to target specific or rare commodities
- ✗ sensitive to outliers
- ✗ can be inefficient
- ✗ requires definable physical boundaries

***What is it?***

a way to take advantage of strengths
of both **list** & area frames






***How is it used?***

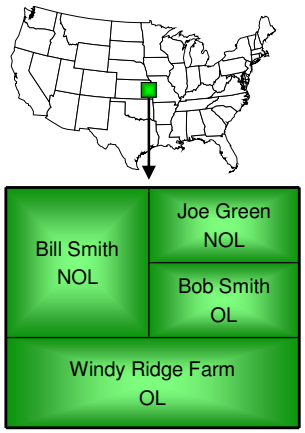
❶ sample — select list & area samples

❷ survey — collect data for selected units from both frames – determine if operations in area sample are on list (OL)

❸ summarize — expand data for list samples & area operations not on list (NOL) using probabilities of selection

*MF expansion = list expansion + NOL expansion*

## Area




## List

Windy Ridge Farm
John Brown
1234 Farm Rd
Anywhere, US 00000

Richard Jones
789 Ranch Rd
Anystate, US 99999

Bob Smith
56 Orchard Rd
Anywhere, US 00000

Dave White
123 Farm Rd
Anywhere, US 00000



## Strengths:

- ✓ together frames cover target population
- ✓ can control variability due to sampling
- ✓ can control costs with large list, small area samples
- ✓ can target specific or rare commodities

## Weaknesses:

- ✗ NOL can be too small
- ✗ overlap determination can be difficult
- ✗ errors in overlap determination can bias estimates
- ✗ list and area frames must be maintained independently



### segments based on latitude & longitude

sampling unit = segment of land using latitude & longitude as boundaries
associate farms, land, animals, etc. with land inside the segment

### segments based on random points

sample unit = segment constructed around random point according to specific rules
associate farms, land, animals, etc. with land inside or touching the segment

### random points

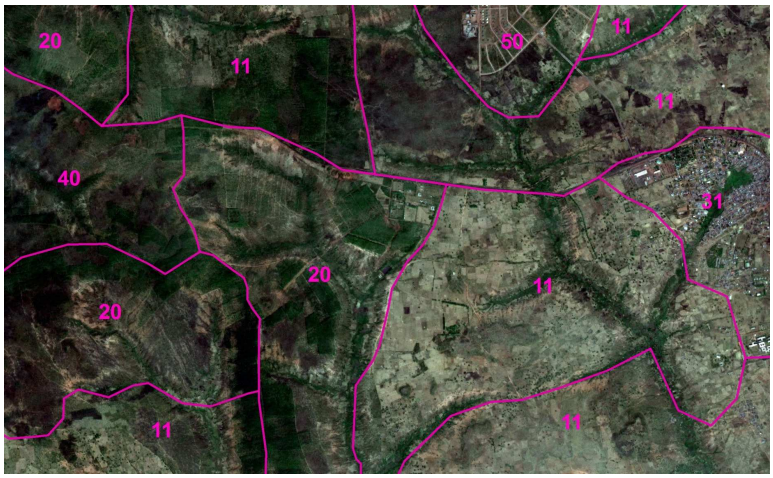
sample unit = random point
associate farms, land, animals, etc. with operator of land at the point

**Nigeria Area Frame – Pilot in Kaduna State**

***Satellite imagery:***

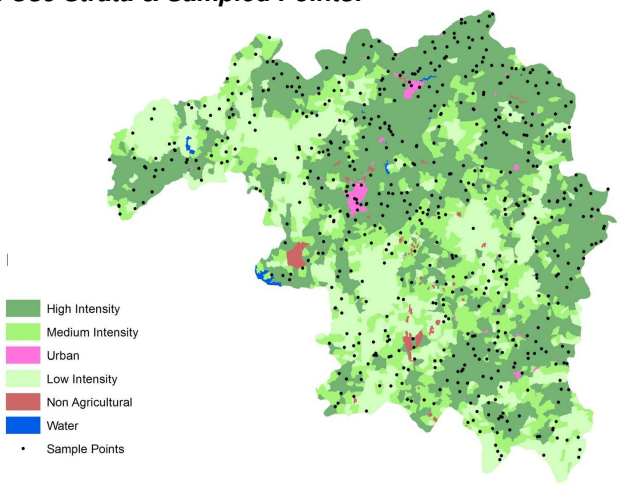
**Nigeria Area Frame – Pilot in Kaduna State**

***Strata blocks:***



## Nigeria Area Frame – Pilot in Kaduna State

***Land Use Strata & Sampled Points:***




## Nigeria Area Frame – Pilot in Kaduna State

***How is it used?***

❶ sample — select a sample of random points

❷ survey —
1. locate point on ground
2. find operator of land under point
3. if operator is a farmer, obtain information about the entire farming operation

❸ summarize — expand data using point-specific probabilities of selection based on total land area in stratum, number of points sampled in stratum, and total land operated by specific farmer

## Nigeria Area Frame – Pilot in Kaduna State

### Kaduna – Area Sample Design

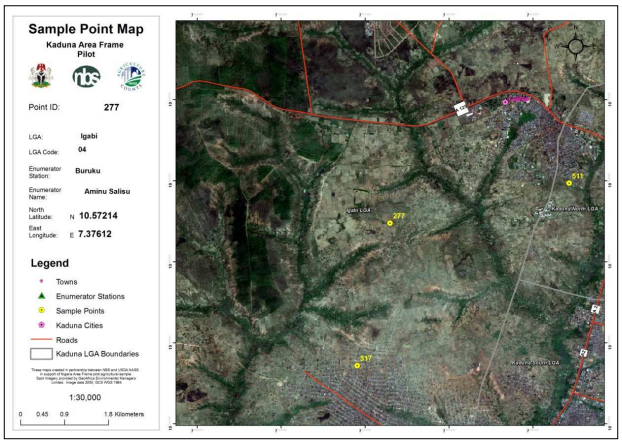
| stratum | boundaries | total land km2 | number of sampled points | expected expansion factors: avg farm = 2 ha | avg farm = 5 ha | avg farm = 100 ha |
|---|---|---|---|---|---|---|
| 11 | agric land, >50% cultivated | 21693.90 | 350 | 3099 | 1240 | 62 |
| 20 | agric land, 15-50% cultivated | 10019.37 | 150 | 3340 | 1336 | 67 |
| 31 | agri-urban | 334.84 | 20 | 837 | 335 | 17 |
| 40 | agric land, <15% cultivated | 11710.74 | 80 | 7319 | 2928 | 146 |
| 50 | non-agric land | 360.52 | 0 | | | |
| 62 | water ≥ 1 $km^2$ | 101.27 | 0 | | | |
| | total | 44220.65 | 600 | | | |

## Nigeria Area Frame – Pilot in Kaduna State

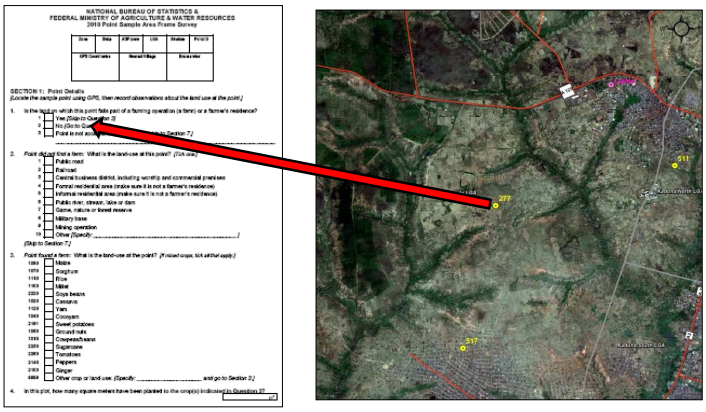
### *Data Collection Tools:*

**GPS**

**Satellite maps**



## *Data collection:*






## *"Point" sample estimator:*

$$y_{ij} = e_{ij} p_{ij} x_{ij}$$

$$\hat{y} = \sum_i \sum_j y_{ij}$$

$$e_{ij} = L_i / n_i l_{ij}$$

$i$ = land use stratum
$j$ = selected point within stratum
$L_i$ = total land in stratum $i$
$n_i$ = total number of points sampled in stratum $i$
$e_{ij}$ = expansion factor for selected point $j$ within stratum $i$
$p_{ij}$ = population indicator for point $j$ in stratum $i$
$x_{ij}$ = survey value for point $j$ in stratum $i$
$l_{ij}$ = total land in farm identified by point $j$ in stratum $i$

# SESSION 4

## SAMPLING FRAMES
## LIST FRAMES FOR AGRICULTURAL CENSUS AND SURVEYS

*A. K. SRIVASTAVA, FAO Consultant*

## 1. Sampling Frames

### *1.1 What is a sampling frame?*

Sampling frame is a collection of sampling units with clear-cut identity for the sampling units. A list of sampling units with serial numbers for the units is an example of a sampling frame. In an agricultural census, normally, agricultural holdings are the sampling units. However, to prepare and maintain an updated list of all agricultural holdings in the entire population for sample selection purposes is neither feasible nor desirable. Instead, the approach of two stage sampling is followed in which enumeration areas (EAs) are selected at the first stage and agricultural holdings are the sampling units at the second stage. Thus, if EAs are to be selected in a stratum (say, district), the list of EAs in the stratum is a sampling frame at the first stage of selection. In an enumeration area, if agricultural holdings are to be selected, a list of such holdings is a sampling frame at the second stage. Thus, we observe that in multi-stage sampling, at each stage of selection, a sampling frame is essential. In fact, availability of a sampling frame is a pre-requisite in any sampling process. An advantage of multi-stage (say, two stage) sampling is that at the second stage, frames need to be prepared only for selected first stage units. In deciding about the frames for agricultural census, care is needed to ensure that all agricultural production units are covered. Large farms and industrial units are normally covered on a complete enumeration basis through update lists, whereas household sector is covered through sampling approach.

It is important to ensure that the sampling frame must provide a complete and up-to-date list of sampling units, without omission and duplications, and without including any units other than the defined sampling units (say, holdings). Omissions and duplications lead to imperfect frames. The sampling frames are used for selection of samples through a random selection procedure. In random sampling, units are selected with pre-assigned selection probabilities. If some units are omitted from the frame, they will have no chance of selection in the sample, while duplicate units will have more chance of selection than they should have. If some unwanted units are included in the frame, it will distort the selection probabilities of the genuine units in the sampling frame. The imperfections in the frames lead to biases and various types of non-sampling errors. Thus, efforts should be made to make the frame free from imperfections.

### *1.1 Sampling frame of enumeration areas*

In agricultural censuses, selection of samples is invariably done in two stages. At the first stage, EAs are sampling units, while at the second stage operational holdings or agricultural households are the sampling units. A list of EAs is normally available from population censuses. Enumeration areas are based on the concept of group of households, for which it is feasible for an enumerator to collect the data needed for population and housing censuses. EAs are also non-overlapping area units, based on clear-cut defined boundaries. In some countries, cartographic maps of EAs, based on satellite data, are now available. In agriculture census and surveys, one has to deal with agricultural holdings and agricultural

households, who are the ultimate respondents. In preparing and maintaining the sampling frame of EAs, role of cartography becomes very important. Since a list of EAs is used as sampling frame, such frames are also referred to as a **list frame** of EAs. These frames have also got an advantage of stability. They are prepared/ updated prior to decennial population censuses. Other kind of frames is **area frames**, which are described separately.

### *1.2 Sampling frames of agricultural households/holdings*

In agricultural censuses, the basic statistical unit has been agricultural holdings. An agricultural holding is defined as "*an economic unit of agricultural production under single management comprising all livestock kept and all land used wholly or partly for agricultural production purposes, without regard to title, legal form, or size. Single management may be exercised by an individual or household, jointly by two or more individuals or households, by a clan or tribe, or by a juridical person such as a corporation, cooperative or government agency. The holding's land may consist of one or more parcels, located in one or more separate areas or in one or more territorial or administrative divisions, providing the parcels share the same production means, such as labour, farm buildings, machinery or draught animals*."

There are two types of agricultural holdings:

i) holdings in the household sector, i.e. those operated by household members and

ii) holdings in non-household sectors, such as corporations and government institutions which are in the form of Large Scale Private and Institutional Farms (PLS&IF).

For the household sector, there is usually a one-to-one correspondence between an agricultural holding and a household with own-account agricultural production activities. There are two special cases where the agricultural holding and household concepts may diverge:

- If there are two or more units making up a household, such as where a married couple lives in the same dwelling as their parents, the two units may operate land independently but, as members of the same household, they make common arrangements for food and pool incomes.
- A household may operate land or keep livestock jointly with another household or group of households. Here, there are two agricultural holding units associated with the household: (i) the agricultural production activities of the individual household itself; and (ii) the joint agricultural operations with the other household.

In the past, some countries have found it difficult to strictly apply the agricultural holding concept in their agricultural censuses and, instead, have defined the agricultural holding to be equivalent to a household with own-account agricultural production. Usually, there is little difference between an agricultural holding and a household with own-account agricultural production. Equating the agricultural holding and household units would have a number of benefits:

- The identification of the holding in the agricultural census would be simplified; it would no longer be necessary to find out about the management of the household's own-account agricultural production activities.

- It would bring the concept of agricultural holding into line with the practice already used in previous agricultural censuses in many countries.
- The use of a common statistical unit – the household – would enable the agricultural census to be more easily linked to the population census.
- It would facilitate the analysis of household characteristics.
- If the scope of the agricultural census was expanded to also include other households not engaged in own-account agricultural production, there would be a common unit between agricultural production units and other households.

In African countries, considering the agricultural households as the sampling units has got distinct advantages taking into consideration the issues mentioned above as well as with regard to operational considerations. In the present write-up, for household sector, we use agricultural holdings and agricultural households synonymously. However, it is important that the definition of a holding used should be clearly stated in the presentation of census results to help in the interpretation of data.

In most of the African countries, the large scale private and institutional farms (PLS&IFs) contribute significantly to Country's economy but they are invariably small in number. Normally, a list of such farms are maintained and updated in every country. For agricultural censuses, no sampling is done for such farms and they are enumerated on a 100 percent basis. However, in a developing economy, the number of such farms keeps on changing and improving continuously, it is important to ensure that before the start of agricultural census, the list of such farms is updated.

For household sector, the target population includes all the agricultural households. Agriculture, in its broader form includes the agricultural activities relating to crops livestock, horticulture etc. The information collected in the agricultural modules of Population and Housing Censuses is of immense help in clearly identifying the agricultural households.

Invariably agricultural holdings are divided into large, medium and small holdings on the basis of holding size, number of livestock etc. Selection procedures for holdings in these categories are not necessarily same. The most common procedure followed for preparation of frames within selected EAs is to prepare list of agricultural holdings for each category separately. These lists provide the list frames within each selected EA.

### *1.3 Exclusion of some holdings*

Ideally, in an agricultural census, all the holdings must be included in the target frame. However, some countries follow specific criteria to exclude some holdings based on holding characteristics. These criteria are invariably based on cut-offs for minimum land holding sizes, and number of different categories of livestock. In a review study for censuses followed in different countries, carried out by FAO, it has been observed that commonly three types of restricting criteria and their possible combination, are followed. These criteria are: a) exclusions based on minimum size of holding; b) exclusions by legal status of holding and c) exclusions based on type and purpose of production on holding. Some examples of

exclusion criteria in African countries (particularly for Uganda and Mozambique) will be demonstrated in the Country discussions.

It must be realized that whatever may be the chosen exclusion process, excluding some types of holdings obviously leads to underestimation of agricultural activity and related production for that country. Differences in the exclusion criteria from one country to the other will lead to comparability problems. However, it is expected that when deciding the exclusion criteria, countries are vigilant to keep these underestimations to a minimum level, maybe in the same order of magnitude as observation errors.

### *1.4 Area Frames*

In area frame methodology, the total land area is divided into small parcels of land (called segments), without overlap or omission and a small number of segments are selected. The selection procedure is usually in two stages where, in the first stage, primary sampling units (PSUs) are selected. These PSUs are then divided into segments (which are approximately equal in area, say, 40 hectares). Within each PSU one segment is selected. Data are collected in respect of agricultural activities of each sample segment. The area frame approach has got an advantage over list frames in the sense that it ensures the completeness of the frame, while list frame is sometimes prone to incompleteness. It also provides an indirect check on the quality of data when areas of the individual fields in the selected segments are reported (Total area of a segment is already known).

***(More detailed discussion on Area Frames is being done separately)***

Area frames are quite often used in conjunction with list frames with the help of **multiple frame methodology.**

### *1.5 Master Sampling Frames*

A master sampling frame is a general purpose sampling frame created from a census, for use in selecting samples for different surveys or different rounds of a periodic survey. The frame is usually maintained by the national statistical office, and is updated on an ongoing basis so that it is available for any survey carried out at any time.

A master sampling frame has several benefits. It is quick and easy to conduct surveys of any kind, because a ready-made frame is already available. The cost of preparing sampling materials and selecting samples is also reduced. Master sampling frames also make it easier to relate data from different surveys and to control the reporting burden on survey respondents.

Master sampling frames suitable for agricultural surveys may be available from either the population census or the agricultural census.

- Population census. The population census master sampling frame is a database of small geographical units, such as villages or EAs, containing key data about each unit, such as the population and the number of households. This can be used to select samples for any type of household survey, including agricultural surveys. A

supplementary frame may be needed for non-household holdings and any other large holdings.

- Agricultural census. Agricultural census master sampling frames are of two types. One type is similar to a population census master sampling frame in that it is a database of small geographical units, such as villages or EAs, containing key agricultural data about each unit, such as the area of major crops and the number of livestock. Another type of agricultural census master sampling frame is a database of agricultural holdings, created and maintained following an agricultural census, containing key agricultural data about each holding, such as the crops grown and the livestock raised. Often, the master sampling frame is a combination of the two frame types, with a database of all important agricultural holdings complementing the EA frame. A master sampling frame from an agricultural census can be used for any type of survey of agricultural holdings.

## 2. Use of Population Census to build a frame for Agricultural Census and Surveys

Population censuses are the main source of all other censuses. In fact, all other censuses, including the agricultural censuses have emanated from it. Besides the use of common concepts and definitions, another application is the use of information collected in population censuses for preparation of frames in agricultural censuses.

In population censuses the basic statistical unit is a household, whereas in agricultural censuses it is agricultural holding. One way could be to use the population census as a household frame for the agricultural census. In this method, the households in the population census could be asked some screening questions to identify the households that are in the scope of agricultural census. On the basis of this screened samples individual agricultural holdings may be identified. The approach clearly distinguishes, in the agricultural census, between data for households and data for agricultural holdings. This is an important element in linking data between the two censuses. Moreover, household data from the population census are also useful for constructing a sampling frame for a sample-based core agricultural census. This could be done by selecting EAs as the PSUs and then within the selected EAs, using population census households for identifying and selecting agricultural holdings.

### *2.1 Existing Agriculture Related Data in the Population Census*

Another option could be to use existing agriculture related data in the population census for identifying agricultural holdings. The population census has got some occupation and industry related information, which may enable to identify agricultural holdings from the population census data. Some of the existing agriculture-related data in population censuses are as follows:

- Main occupation - This item is collected for each economically active person. It can be used to provide tabulations of persons working in an agricultural occupation. This includes household members of agricultural holdings working in an agricultural occupation (on the household's holding or another holding), as well as persons working as employees in an agricultural occupation.

- Main industry - Industry is the activity of the establishment in which the person works in his/her main job. This item is collected for each economically active person, and can be used to provide tabulations of persons working in the agricultural industry. This includes household members of agricultural holdings working in the agricultural industry, as well as persons working as employees in the agricultural industry.
- Status in employment - This item refers to whether the person is working as an own-account worker, family worker, employee, etc., and is collected for each economically active person. This is useful for analysing persons with an agricultural main occupation according to whether they are agricultural holders (an "own-account worker"), working on the household's holding (contributing family worker), or in paid agricultural work (employee).

From the above information, households having any member with both: (i) an agricultural main occupation; and (ii) status in employment *own account worker,* may be identified. Such households are closely related to agricultural holdings.

### *2.2 Collecting Agriculture-Related Data in the Population Census*

Yet another option for using population census for building the frame of agriculture census and subsequent surveys is to collect additional agriculture related data in the population census. This option has been tried in recent population censuses in several countries in which an agricultural module is canvassed as a piggy-back to the main population census questionnaires.

I would like to illustrate this aspect with a recent application in Mozambique. A Population and Housing Census (PHC) or Recenseamento Gerald a Populacao e Habitacao (RGPH) was conducted in Mozambique during 2007-08. In this PHC, information on some of the basic agricultural activities carried out by each and every household was obtained in an agricultural module (Section G). It is important to note that agriculture, in its broad sense, deals not only with crops but also with livestock, aquaculture, horticulture etc. The questions actually asked were as follows:

**Section G - Livestock and Aquaculture activity**

(As obtained in Section G: Population and Housing Census – Agriculture Module)

G 1: Does any member of the household practices agriculture for himself?

Yes No

G2: Does the household have any tank for aquaculture?

Yes No

If yes, how many? ------

G3: Does any member of the household practice fishing through traditional methods?

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: How many animals, this household has got?

G 6.1 Cows/Bullocks ------

G 6.2 Goats ------

G 6.3 Sheep ------

G 6.4 Pigs ------

G 6.5 Chicken ------

G 6.6 Ducks ------

This provided a good base for the preparation of sampling frame of agricultural households. In Mozambique, the livelihood of large number of households depends on agriculture which includes different combinations of growing crops, keeping livestock, practicing aquaculture etc. These types of agricultural practices are expected to be followed in most of the developing countries and more so in African countries.

On the basis of the information collected as above, it was possible to get information for each and every household whether they are doing agriculture (including livestock, aquaculture, horticulture etc) or not. From this information, it was possible to ascertain whether the household is an agricultural household. In the previous agriculture census (*Censo Agro - Pecuario* CAP- I) conducted in 1999-2000, farms were classified as small, medium and large according to cultivated area or number of head of cattle and some holdings were excluded on the basis of combination of land and livestock. In the present agricultural census CAP-II (2009-10), for the choice of a suitable cut off for the exclusion of households, based on the data from PHC (2007/08), following scenarios were considered.

***Scenarios for definition of Agricultural Households***

(AG: Agriculture; P: Livestock; CC: Cashew nuts and Coconuts)

**Scenario 1 – Variable AGP**

All households who respond that

i) At least one member of the household practices agriculture for own use

ii) They are having at least one (i.e.>= 1) livestock (i.e. Cows/bullocks or goats or sheeps or pigs or chicken or ducks)

**Scenario 2 – variable AGP_C**

All households who respond that

i) At least one member of the household practices agriculture for own use,
ii) They are having at least 5 (cashewnut + coconut) trees and
iii) They are having at least one (i.e. >= 1) livestock (i.e. Cows/bullocks or goats or sheep or pigs or chicken or ducks)

**Scenario 3 – variable AGP_L (Cut - offs in Urban areas)**

All households who respond that

i) At least one member of the household practices agriculture for own use
ii) In urban areas: They are having at least 5 (cashewnut + coconut) trees
   In rural areas: They are having at least 1 (cashewnut+ coconut) tree
iii) In urban areas: They are having at least one cow/bullock or at least three (goat + sheep + pig) or at least 5 (chicken + ducks)
   In rural area: they are having at least one livestock (as already defined)

**Scenario 4 – variable AGP_L1 (Cut - offs for rural and urban areas)**

All households who respond that

i) At least one member of the household practices agriculture for own use,
ii) In urban and rural areas : They are having at least 5 (cashewnut + coconut) trees
iii) In urban and rural areas: They are having at least one cow/bullock or at least three (goat + sheep + pig) or at least 5 (chicken + ducks)

**Scenario 5 – variable AGP1_L1 (Cut – offs for rural and urban areas)**

Same cut offs as Scenario 4 with a difference that in this case, cut – offs for small ruminants becomes at least two (goat + sheep + pig)

**Scenario 6 – variable AGP2_L1 (Cut – offs for urban and rural areas)**

All households who respond that

i) At least one member of the household practices agriculture for own use,

ii) In urban areas, they are having at least 5 (cashewnut + coconut) trees,
in rural areas they are having at least one (cashewnut + coconut) tree

iii) In Urban as well as rural areas at least one cow/bullocks or
in urban areas at least three (goat + sheep + pig) and
in rural areas at least one (goat + sheep + pig) or
urban areas at least ten (chicken + ducks) and
in rural areas at least one (chicken + ducks)

The scenarios considered above are presented in following table:

# Matrix- Scenarios

| Variables Section G | Urban/ Rural | Scenario 1 | Scenario 2 | Scenario 3 | Scenario 4 | Scenario 6 |
|---|---|---|---|---|---|---|
| Agriculture (G1=1) | U & R | All Hhs | All Hhs | All Hhs | All Hhs | All Hhs |
| Cashew+ Coconut | U<br>R | -<br>- | >=5<br>>=5 | >= 5<br>>=1 | >= 5<br>>=5 | >= 5<br>>=1 |
| Cows | U<br>R | >=1<br>>=1 | >= 1<br>>=1 | >= 1<br>>=1 | >= 1<br>>=1 | >= 1<br>>=1 |
| Goats+Sheep+ Pork | U<br>R | >=1<br>>=1 | >= 1<br>>=1 | >= 3<br>>=1 | >= 3<br>>=3 | >= 3<br>>=1 |
| Chicken + Ducks | U<br>R | >=1<br>>=1 | >= 1<br>>=1 | >= 5<br>>=1 | >= 5<br>>=5 | >= 10<br>>=1 |
| Agril. Househols | | 3,592,103 | 3,691,651 | 3,659,297 | 3,433,733 | 3,611,629 |
| % of Total Hhs | | 75.0 | 77.0 | 76.4 | 71.7 | 75.4 |

In the above table only five scenarios (excluding scenario 5, which was very much similar to scenario 4 with a minor difference in the cut offs) are presented. Based on the discussions with the experts in the concerned departments, finally scenario 6 was chosen as a criterion for identification of agricultural households. It is interesting to observe that nearly 75 percent of the households were agricultural households based on the criterion developed.

***Analysis of agricultural questions in the population census.***

| Activities | Percent |
|---|---|
| No Agril. activities | 27.8 |

| | |
|---|---|
| Only Agriculture | 24.2 |
| Only Livestock | 0.4 |
| Agri + Livestock | 4.4 |
| Agri + livestock + Chicken | 9.0 |
| Agri + CC | 7.7 |
| Agri + CC + Chicken | 3.8 |
| Agri + live + Chicken + CC | 5.0 |

The main advantage of considering the various scenarios was that it

i) helps in frame preparation,

ii) provides size measure for

allocation of sample and

selection of EAs by probability proportional to size (pps) method,

iii) provides distribution of agriculture, livestock and cashew, coconut crops across provinces and even district-wise,

iv) helps in identifying important pockets for conducting subsequent surveys.

In other countries also, information on agriculture has been obtained in the respective population censuses. However, the type of information and the format in which it is collected varies from one country to the other. For example, in Uganda PHC was conducted in 2002 and agricultural module consisted of information about number of plots and number of livestock in each EA. The information has been used in the selection of EAs in recent agricultural census (2008-09). In Ethiopia, population census was conducted in 2007-08. Information has been collected at the community level for each EA, whether specific principal crops are grown, livestock kept and whether crops are grown in Belg season. The information could be useful in identifying important areas for special crops and livestock which may be useful for conducting special surveys.

In the workshop, it would be useful to explore the types of information relating to agriculture collected in recent population censuses by different African countries and discuss the ways and means for utilizing the information in agricultural censuses and surveys.

**3. Design of Agricultural Sample Census (Core or Thematic Modules)**

As already mentioned above, one of the main features of WCA 2010 is that a modular approach has been suggested with a core module based on complete enumeration to provide key data and one or more census supplementary modules to cover more in depth topics. Following 16 items are recommended for the core module:

1 Identification and location of agricultural holding.
2 Sector of agricultural holding.
3 Sex of agricultural holder.
4 Age of agricultural holder.
5 Household size.
6 Main purpose of production of the holding.
7 Area of holding according to land use types.
8 Total area of holding.
9 Land tenure types on the holding.
10 Whether holding is irrigated.
11 Presence of temporary crops on the holding by crop type.
12 Presence of permanent crops on the holding by crop type and whether in a compact plantation. .
13 Number of animals on the holding by livestock type.
14 Presence of aquaculture on the holding.
15 Presence of forest and other wooded land on the holding.
16 Other economic production activities of the holding's enterprise.

For thematic modules following themes are suggested:

Theme 01 – Land
Theme 02 – Irrigation and water management
Theme 03 – Crops
Theme 04 – Livestock
Theme 05 – Agricultural practices
Theme 06– Agricultural services
Theme 07 – Demographic and social characteristics
Theme 08 – Farm labour
Theme 09 – Household food security
Theme 10 – Aquaculture
Theme 11 – Forestry
Theme 12 – Management of the holding

For more details, Chapter 4 of the document *A system of integrated agricultural censuses and surveys – Volume 1, World Programme for the Census of Agriculture* may please be referred. From sampling point of view, the approach implies intensive use of sampling for thematic modules in addition to the core module which is to be covered through complete enumeration. In many countries, where the time gap between population census and agricultural census is very little, some of the items of the core module already covered in the population census may be of help. Complete enumeration based coverage of core modules may help in getting the frames for the supplementary modules and samples of holdings may straight away be selected for the respective modules. However, a more realistic approach could be to follow a multi stage sampling approach as follows:

- In the population census, the entire country is already divided in non-overlapping and distinct EAs.
- Conduct the core census module by enumerating all holdings in all EAs.
- Select a sample of EAs for the specific module (say, livestock module).
- During the core census enumeration, identify all holdings with livestock in the sample EAs. Special holdings with livestock, such as large units, would also be identified in the non-sample EAs.

- For the livestock module, enumerate all holdings with livestock in the sample EAs and all special holdings with livestock in the non-sample EAs.

The advantage of the multi-stage approach is that the sample selection of EAs can be done by technical staff prior to the fieldwork, rather than requiring each enumerator to do the sample selection. This makes the census field operations easier. A convenient way to organize the census enumeration would be to assign the best enumerators to the sample EAs, to interview each holding for the core module, and, if the holding is within scope of the supplementary module, ask further questions for the supplementary module. All other enumerators would be assigned to the non-sample EAs to collect core data only. Senior field staff could enumerate the special holdings.

## Technical Session 4B

Group Discussion on practical issues relating to sampling frames

***Learning Objective***

**You have by now selected core data items relevant to AFRILAND and decided on what item is suitable for complete enumeration and what should be done by sampling. You are aware that a Population Census is under preparation in AFRILAND and you want to take this opportunity to get an up to date and adequate frame for conducting your agricultural census in a cost effective manner.**

***Activities:*** Same groups, A, B, C, D.

**Question 1:**

Among the core data items that you have identified for complete enumeration, what are the items that you would like to negotiate with the responsible of the population census for inclusion in the population census? Why? How this would support your agricultural census strategy?

**Question 2:**

How will these questions in the population census help you to decide on an operational definition of agricultural holding? What are possible scenarios and what are the implications in terms of agricultural census coverage and the cost of the census?

**GROUP REPORTS ON**

1) **Technical Session 2 B: Group Exercise on Operationalization of Modular Approach**
2) **Technical Session 4 B: Group Discussion on practical issues relating to sampling frames**

The groups are required to prepare a full report on Session 2B and 4B that will be presented **on Friday 23 April 2010**. The report should include:

1) the strategy for development of agricultural statistics with an integrated agricultural census/survey programme;
2) the agricultural data items to be included in the population census and the operational definition of agricultural holding, explaining the rationale and implications;
3) the strategy and plan for a cost effective agricultural census with identification of core module items (also taking into account the population census), census supplementary module items;
4) Items for thematic follow-up surveys.

# SESSION 5

## APPLICATION OF SAMPLING

## 4. Sampling Schemes

Here we briefly describe some of the sampling procedures commonly followed in actual sample surveys. A sample must be representative of the population. The representative character of a sample is normally ensured through random selection procedures.

### *4.1 Random Sampling*

In random sampling procedures, selection of sample is done through a probability mechanism, so that all samples are provided a definite probability of selection. This is why, Random sampling is also known as probability sampling. The simplest of all random sampling procedures is the one in which all possible samples are provided equal probability of selection. It is not always feasible, nor desirable, to generate all possible samples and then select one sample with equal probability. The procedure of selection, normally followed, is *one after another draw* procedure in which units are selected one after the other draw with equal probability of selection at each draw. This procedure is known as *simple random sampling.* In this procedure all population units have equal chance of selection. Thus, simple random sampling may be defined as follows:

***Simple random sampling*** is the method of selecting the units from the population where all possible samples are equally likely to get selected.

It follows that in *simple random sampling* every population unit has the same chance of being selected in the sample. Such sampling procedures are known as Equal Probability Selection Methods (EPSEM). It may be noted that simple random sampling is an EPSEM procedure, but all EPSEMs are not necessarily simple random sampling methods.

### *4.2 Systematic Sampling*

Systematic sampling is yet another method of selecting a sample. In simple random sampling the units were selected randomly at each draw. In systematic sampling the whole sample is selected with just one random number. The procedure is defined as follows:

*Systematic sampling* is a method of sample selection in which only the first unit is selected at random and the rest of the units are automatically selected according to a predetermined pattern.

The most common predetermined pattern is the one in which after the random start, units are selected at equal intervals. This method is also known as *linear systematic sampling*. Suppose we want to select a systematic sample of size n from a population consisting of N units. The method of linear systematic sampling is employed when N is a multiple of n i.e. N=n.k where k is an integer. Let us assume that the nk serial numbers of the population units in the frame are arranged as follows:

| | | | | | | |
|---|---|---|---|---|---|---|
| 1 | 2 | 3 | .. | r | .. | k |
| k+1 | k+2 | k+3 | | k+r | | 2k |
| 2k+1 | 2k+2 | 2k+3 | | 2k+r | | 3k |
| … | … | … | … | … | … | … |
| … | … | … | … | … | … | … |
| … | … | … | … | … | … | … |
| (n-1)k+1 | (n-1)k+2 | (n-1)k+3 | | (n-1)k+r | | nk |

Select a random number r such that $1 \le r \le k$. The number r is called a random start and k is called sampling interval. The selected sample is the population units with serial numbers r, r+k, …, r+(n-1)k.

Systematic sampling is an EPSEM procedure. However, it is not equivalent to simple random sampling. The method has got advantage of simplicity in selection. However, the efficiency of estimation depends on the ordering of the units. With a suitable choice of arrangement, keeping in view the trends in the population, the method has got the potential of performing very well. It has got a limitation that an unbiased estimation of variance is not feasible with this method of sampling.

There are several other variants of systematic sampling, depending upon the systematic pattern used for selection. One such procedure is ***Circular Systematic Sampling***, in which random start is taken between 1 and N and then subsequent units are selected with equal interval after arranging the population units in a circular way. This method takes care of the situation when N is not multiple of n i. e. $N \ne nk$.

### *4.3 Unequal Probability sampling*

Simple random sampling and systematic sampling are EPSEM procedures. However, when units vary considerably in sizes, providing equal chance of selection for every unit may not be a desirable proposition. Under such situations, selection of units with unequal probabilities may provide more efficient estimators. In this scheme, the units are selected with probability proportional to a given measure of size. The size measure is an auxiliary variable closely associated with the study variable. This method is known as varying probability sampling or probability proportional to size (PPS) sampling. For selecting a population unit with PPS, following methods are used:

**<u>Cumulative Total Method</u>:**

Let the size of the $i^{th}$ unit be denoted by $X_i$. Let the total size for N population units be, $X = \sum X_i$, i=1,…., N. Then the selection procedure consists of the following steps:

**Step 1.** Define cumulative totals

$$T_{i-1} = X_1 + X_2 + \ldots\ldots\ldots + X_{i-1}$$
$$T_i = T_{i-1} + X_i \; ; \qquad i=2,\ldots, N$$

**Step 2.** Chose a random number r such that $1 \le r \le k$

**Step 3.** Select i$^{th}$ population unit if $T_{i-1} < r \leq T_i$

The probability of selecting the ith population unit, using this procedure is given by $P_i=X_i/X$. This procedure is described for selecting one unit with probability of selection $P_i$.

It is observed that in this method, it is required to cumulate the sizes and write down these cumulative totals. The procedure becomes a bit tedious when population size N is large. A procedure which does not involve cumulating the sizes was given by D. B. Lahiri (1951) and is described below:

**<u>Lahiri's Method</u>**

**Step 1**. Select a random number (say) I from 1 to N

**Step 2.** Select another random number (say) j, fro 1 to M, where M is either equal to the maximum of the sizes $X_i$; i=1, 2, ….., N or is more than the maximum size in the population.

**Step 3**. If $j \leq X_i$, the i$^{th}$ unit is selected, otherwise, the pair (i,j) of random numbers is rejected and another pair is chosen by repeating the steps 1 and 2.

The procedure is repeated till a unit is selected. This methods ensures that the probability of selection for i$^{th}$ population unit is $P_i=X_i/X$.

**Probability Proportional to Size with and without replacement**

If n units are to be selected with replacement, the procedure is to be applied independently n times. Thus, conceptually, every selected unit is replaced to the population before next unit is selected again with the same probability measures. The estimation procedure for estimating population total and for estimating the sampling variances is simpler in this case. We are not describing the estimation procedure here.

For PPS sampling without replacement, the selected units are to be excluded from the population for subsequent draws and the selection and estimation becomes much more complex. Since estimation become somewhat involved, attempts have been made to make the estimation procedures somewhat simpler. We shall not get into vast area of sampling literature in varying probability without replacement. However, it is worthwhile to mention about a special category of varying probability without replacement procedure, which is commonly used in Agricultural censuses as well as annual surveys. For sampling procedures with varying probabilities without replacement, if inclusion probabilities of sampling units are proportional to size measures, then estimation becomes very simple. Such procedures are called Inclusion Probability Proportional to Size (IPPS) procedures or $\pi ps$ procedure, since inclusion probabilities are sometimes denoted by $\pi_i's$. One such procedure is PPS systematic sampling which is described below.

**Probability Proportional to Size (PPS) Systematic Sampling**

The procedure is described here for selection of EAs within a specific stratum. Define
N = number of EAs in the stratum

n = number of EAs to be selected in the stratum

$z_i$ = the measure of size (MoS - number of agricultural households in this case) for the $i^{th}$ EA in the stratum

$$Z = \sum_{i=1}^{N} z_i$$

$$p_i = z_i / Z \qquad i=1, \ldots\ldots\ldots\ldots, N$$

$$\pi_i = np_i \qquad i=1, \ldots\ldots\ldots\ldots, N$$

The $\pi_i$ values are the selection probabilities for the ith EA. The pps systematic sampling selection procedure is described in following steps:

**Step 1:** In this step, the procedure of implicit stratification is described. We consider a sub-administrative (say, SA) level as implicit strata. Sort the list of EAs in the stratum by SA level. Within a PA, arrange the EAs in ascending order of MoS; then in the next SA arrange the EAs in descending order of MoS. Continue this sorting by alternating between ascending and descending sorting from one SA to the next. This type of sorting helps in improving the efficiency of pps systematic sampling.

**Step 2:** Check that $np_i < 1$, i.e. $z_i$ is less than $Z/n$ for all i in the stratum.

**Step 3:** Compute cumulative totals

$C_1 = \pi_1$

$C_2 = C_1 + \pi_2$

…..

$C_{N-1} = C_{N-2} + \pi_{N-1}$

$C_N = C_{N-1} + \pi_N$ (Note that $C_N = n$)

**Step 4:** Generate a random number "r" between 0 and 1. Compute the numbers $r_i$ = r+i-1 with i = 1, 2, 3, …..,n+2.

**Step 5:** Select the n EAs with the labels $i_1, i_2, i_3, \ldots\ldots, i_n$ such that

$$C_{i_1-1} < r_1 \le C_{i_1}$$

$$C_{i_2-1} < r_2 \le C_{i_2}$$

$$C_{i_3-1} < r_3 \le C_{i_3}$$

……..

$$C_{i_n-1} < r_n \le C_{i_n}$$

The procedure yields a sample of size n with pps systematic sampling and the selection probabilities are given by $\pi_i = np_i$ ; i=1, 2, 3, ……,N.

In this procedure, the estimation of population total becomes very simple and the estimator is given by

$$\hat{Y} = \frac{1}{n}\sum_{i=1}^{n}\frac{y_i}{p_i}$$

### *4.4 Stratification*

The selection procedures described so far i.e. simple random sampling; systematic sampling and unequal probability sampling are all methods of selection. In order to improve the efficiency of estimators, some times the population is divided into certain number of groups, such that the variability within the group is minimum, whereas it is maximum between groups and then smaller numbers of units are selected within each group following any of the random selection methods as described earlier.

The procedure of partitioning the population into groups, called strata and then drawing a random sample independently from each stratum, is known as **stratified random sampling.** In fact, stratification is a control measure applied in the process of selecting a sample for improving the precision of estimates.

Certain considerations need to be addressed while deciding about the stratification plans:

1. How many strata to be formed?
2. How to allocate the total sample size to different strata?
3. How to form the strata?

The basic considerations for all these questions are that the sampling variance should be minimized for a given cost or vice versa i.e. minimize the cost for a specified precision level. Following considerations should be kept in mind for stratified sampling.

1. The strata should be non-overlapping and should together comprise the whole population.
2. The strata should be homogeneous with respect to the study variable.
3. Several rules are available in literature for demarcating the strata boundaries. However, when it is difficult to stratify the population with respect to the study variable or a highly correlated auxiliary variable, the administrative convenience may be considered as a basis for stratification.

**Number of strata**

Regarding the question of number of strata, it may be observed that the normally the efficiency of stratified sampling estimators increases with increase in number of strata. But rate of reduction in the variance decreases as the number of strata increases. Also, the cost of the survey is affected by increase in the number of strata. Based on certain theoretical and empirical considerations, Cochran (1977) observed that if an increase in number of strata (say, L) beyond 6 necessitates any substantial decrease in sample size n in order to keep the cost constant, the increase will seldom be profitable. This indicates that if one is interested in overall estimates for the population mean/total, number of strata around six or seven should be reasonable enough. However, if estimates are wanted also for geographical subdivisions of the population, the argument for larger number of strata is stronger.

**Allocation of sample to different strata**

Although the total sample size is generally obtained based on cost and variance considerations, the decision about allocating the total sample size to different strata has to be made before selecting the sample. Following methods are commonly available:

1. Equal allocation
2. Proportional allocation
3. Compromise allocation
4. Optimum allocation

We explain the allocations with the help of an example in which districts are not only strata, but the interest also lies in getting reliable district level estimates and also national level estimates. Enumeration Areas (EAs) are the sampling units. A given sample size of EAs is to be allocated to different strata (districts).

**Equal Allocation:** In this approach, the sample is allocated equally to each stratum. The strata sizes vary considerably and equal allocation will provide not so efficient estimates at higher levels as the districts will not get due representation in the sample. This allocation is therefore, not a suitable alternative.

**Proportional Allocation:** This allocation will provide sample sizes in proportion to strata sizes. This is a good alternative for provincial and national level estimates. Estimates for larger districts should be good enough, but smaller districts will have poor estimates.

**Compromise Allocation:** In this approach we try to get a balance between producing reliable district level estimates and reliable national level estimates. Sometimes a "square root" allocation is used, in which the sample is allocated in proportion to $x^{1/2}$, where x is the measure of size. A more general allocation plan is the "power allocation" in which the sample is allocated in proportion to $x^{\lambda}$, where $\lambda$ can take values between zero and 1. A suitable value of $\lambda$ may be obtained by obtaining the design effects for national level estimates and also keeping in mind the requirement of getting reasonable district level estimates. Normally, $\lambda$ =0.4 or 0.5 is considered good enough in many situations.

**Optimum Allocation:** In this method variance of the stratified estimator is minimized with respect to a given cost. Let us consider a simple cost function

$$C = c_0 + \sum_{h=1}^{L} c_h n_h$$

where $c_0$ is the overhead cost, $c_h$ is the cost of observing study variable y for each unit selected in the sample from h$^{th}$ stratum, h=1,…., L. After optimization, a fixed cost – minimum variance allocation is given by

$$n_h = \frac{(C - c_0)W_h S_h \big/ \sqrt{c_h}}{\sum_{h=1}^{L} W_h S_h \sqrt{c_h}}$$

If the cost per unit is same for all the strata then the variance is minimized with respect to the restriction

$n = \sum_{h=1}^{L} n_h$ and $n_h$ is given by

$$n_h = n \frac{W_h S_h}{\sum_{h=1}^{L} W_h S_h}$$

$$= n \frac{N_h S_h}{\sum_{h=1}^{L} N_h S_h}$$

This allocation is also known as **Neyman's optimum allocation.**

### *4.5 Cluster sampling and Multi-stage sampling*

Consider the situation of agricultural censuses, in which agricultural holdings are the sampling units. In case, a list of all the holdings in a stratum (say district) is not available, a sample of holdings can not be selected. Even if the list is available, a sample of holdings straightaway selected from the entire stratum will be scattered all over the stratum. This will involve lot of travel expenditure.

A list of EAs is usually available. Each EA is a group or cluster of households. If EAs are selected and all the agricultural households in the selected households are enumerated, then a considerably reduced number of EAs will account for the same number of agricultural households to be selected inn the sample. The spread of selected households will be limited to the selected EAs only, thereby reducing the travel expenditure.

The cluster sampling consists of forming suitable clusters of contiguous population units and completely enumerating all the units in a sample of clusters, selected according to a suitable sampling scheme.

In terms of efficiency, cluster sampling is advantageous if clusters are heterogeneous with respect to study variable. In this respect, cluster sampling is converse of stratified sampling in the sense that both constitute of groups of units but strata should be homogeneous whereas clusters should be heterogeneous.

Multi stage sampling is a natural extension of cluster sampling. If the clusters are not completely enumerated, but units are further selected within selected clusters then it is called two stage sampling. Thus, in agricultural census example, if EAs are selected and then within each selected EA, agricultural holdings are selected, the sampling is done in two stages. The selection may be extended to more than two stages and the procedure is termed as multistage sampling. The sampling units at the first stage are called first stage units (FSUs) or also primary sampling units (PSUs). The units at the second stage are termed as second stage units or secondary sampling units (SSUs).

An important feature of multi-stage sampling is that at different stages, samples are selected independently and different methods of selection may be used at different stages. For example, in two-stage sampling, SRSWOR method may be followed at both the stages. Particular cases of special interest are when PPS with replacement or PPS systematic sampling is followed at first stage and SRSWOR are followed at the second stage. The later case, i.e. when PPS systematic sampling is followed at the first stage and SRSWOR at the second stage, is quite common in agricultural censuses and surveys. In this case, EAs are selected at the first stage with measures of sizes as the number of agricultural households in EAs and a given number of agricultural households are selected at the second stage in each of the selected EAs. This approach yields an EPSEM method of selection for each house holds. However, EPSEM nature of the selected sample is sometimes vitiated slightly due to differences in the size measure used in the selection process and actual number of agricultural households at the time of field work, when second stage selection of agricultural households actually takes place.

### *4.6 Multivariate Probability Proportional to Size (MPPS) Sampling*

In PPS sampling, samples are selected with probability proportional to a size measure. The size measure is normally some auxiliary variable, which is highly correlated with the study characteristics. If there are several characteristics of interest, there may be a number of variables which may be correlated to the study variables. However, for sample selection with PPS, only one variable may be used. This variable could be a combination of auxiliary variables in order to generate a probability measure for selection. Multivariate approaches for generating a common index which could be used for selection purposes are sometimes used. The situation of several study characteristics of interest is very common in agricultural censuses and surveys. The characteristics related to different themes of supplementary modules are simple examples of multiple characteristics of interest.

An approach of MPPS as used in Censuses and surveys used in China (Steiner (2000)) is described below.

Define N= Number of units in the population

$N_k$= Number of units in the population having $k^{th}$ characteristic (k=1, …, K)

$n_k$= Number of units to be selected for the $k^{th}$ characteristic).

$X_{ik}$= value of the $k^{th}$ auxiliary variable for $i^{th}$ population unit.

$$X_k = \sum_{i=1}^{N_k} X_{ik}$$

$$p_{ik} = X_{ik} / X_k$$
$$\pi_{ik} = n_k p_{ik}$$
$$\pi_i = Max(\pi_{i1}, \pi_{i2}, ....., \pi_{iK})$$

In MPPS procedure, for selecting i[th] unit, select a random number $r_i$ (say) between 0 to1. If $r_i \le \pi_i$, then i[th] unit is selected, otherwise rejected. Continue this procedure independently for all the N units in the population. Essentially, the procedure is a Bernoulli's trial experiment with $\pi_i$ as the probability of selection for i[th] unit.

The procedure ensures that the individual selection probabilities for different characteristics are taken into account and maximum one is taken as the selection probability for i[th] unit. The probability $\pi_i$ serves as an index value based on all the auxiliary characteristics.

For estimation purposes, $w_i = \frac{1}{\pi_i}$ serves as the basic weight.

**5. Determination of Sample Size**

Determination of sample size is one of the initial questions which a survey statistician has to face while planning any sample survey. Cost and variance are the prime considerations while working out the sample size requirement. In random sampling, sampling variances are generally expressed as a function of sample size it reduces with increase in sample sizes. Cost of the survey is an increasing function of the sample size. Thus, increasing the sample size reduces the variance but it increases the cost. For a desirable sampling size, a balance is needed between cost and variance.

The principal steps involved in the choice of a sample size are as follows:

1. There must be some statement concerning the desired limits of error. In other words some statement is needed as to what is the tolerable margin of error in the estimates. This statement has to come from the persons, who wish to use the results.
2. Some statement that connects the sample size n with the desired precision of the sample must be found. One of the advantages of probability sampling is that sampling variances which measure the precision can be expressed in terms of n.
3. Sampling variances are population parameters and it contains some parametric values which need to be estimated in order to give specific results. For example, in simple random sampling, the sampling variance is a function of n but it has also got mean squares i.e. $S^2$.
4. Finally, the chosen value of n must be appraised to see whether it is consistent with the resources available to take the sample.

We consider the case of simple random sampling for quantitative character y, to demonstrate the steps needed for determining the sample size. Let r be the margin of relative error to be tolerated in estimating the population mean $\bar{Y}$. An unbiased estimator of population mean $\bar{Y}$ is sample mean $\bar{y}$. We want

$$\Pr\left(\left|\frac{\bar{y}-\bar{Y}}{\bar{Y}}\right| \geq r\right) = \Pr\left(\left|\bar{y}-\bar{Y}\right| \geq r\bar{Y}\right) = \alpha ,$$

where $\alpha$ is a small probability. We assume that $\bar{y}$ is normally distributed. Also, the standard error of $\bar{y}$ is

$$\sigma_{\bar{y}} = \sqrt{\frac{N-n}{N}}\frac{S}{\sqrt{n}}$$

Hence

$$r\bar{Y} = t\sigma_{\bar{y}} = t\sqrt{\frac{N-n}{N}}\frac{S}{\sqrt{n}}$$

Solving for n gives

$$n = \left(\frac{tS}{r\bar{Y}}\right)^2 \Bigg/ \left[1+\frac{1}{N}\left(\frac{tS}{r\bar{Y}}\right)^2\right]$$

Here, $\frac{S}{\bar{Y}}$ is the coefficient of variation which is a fairly stable quantity. In order to calculate n, we need an approximate idea about this coefficient of variation.

As a first approximation, we take

$$n_0 = \left(\frac{tS}{r\bar{Y}}\right)^2 = \frac{1}{C}\left(\frac{S}{\bar{Y}}\right)^2$$

If $n_0/N$ is appreciable, we compute n as

$$n = \frac{n_0}{1+(n_0/N)}$$

In case of **qualitative characteristics if a proportion P** is to be estimated and p is the sample proportion, the sample size is given by

$$n = \frac{\left(t^2PQ/d^2\right)}{1+\left(\frac{t^2PQ}{d^2}-1\right)\Big/N}$$

If N is large, a first approximation is

$$n_0 = \frac{t^2PQ}{d^2}$$

and $$n = \frac{n_0}{1+(n_0/N)}$$

**Design Effect (Deff) and its role in sample size determination for complex designs**

What has been described above is a procedure for determining the sample sizes in simple random sampling without replacement. In actual practice designs are much more complex.

Kish (1965) described Deff as ratio of the variance of the estimate obtained from the (more complex) sample to the variance of the estimate obtained from a simple random sample of the same number of units. The sample size as obtained for simple random sampling is multiplied by Deff in order to get the required sample size for the complex design. The concept was initially given in the context of cluster sampling. In cluster sampling with equal clusters, the design effect is given by $\{1+(M-1)\rho\}$, where M is the cluster size and $\rho$ is the intra-class correlation. In actual practice, the design effect is worked out from previous surveys and is used to determine the required sample size for the current survey. If the complex design is more efficient than the simple random sampling, value of Deff will be less than one and the required sample size will be smaller than the one obtained for simple random sampling. On the other hand if Deff is more than one, the required sample size will be more than the one obtained on the basis of simple random sampling.

## Technical Session 5B

### Practical Exercises on Sampling and Estimation

**Learning Objectives**

*(1) To allocate a given sample size of Enumeration Areas (EAs) to strata (districts), following different methods of allocation.*

*(2) To select samples for the allocated sample sizes in given strata (Districts – rural/urban)*

(Data source-Mozambique Census of Agriculture 2009-10)

**Activities**

1) The data from Mozambique Agricultural Census 2009-10 (CAP-II) for one Province is given to each Group (Seven groups of 4 participants each). Sample sizes for each stratum will be given to each group. Participants are expected to allocate the sample sizes to different strata according to compromise allocation for $\lambda = 0, 0.4, 0.5$ (The rationale and the method of allocations is already described in the class).
2) The same data will be used for selection of allocated samples for given districts according to PPS systematic sampling.
3) The samples selected for different groups will be pooled and overall interpretation for different allocation approaches to be made.

## 6. Estimation Procedures

One of the main objectives of conducting sample surveys is to estimates population parameters of interest. Quite often, the interest lies in estimating parameters like population mean/total, sampling variances etc. Keeping in view the parameter of interest, estimators are chosen satisfying desirable properties like unbiased-ness, efficiency etc. For every sampling design, the estimation procedure invariably includes estimator of the parameter and estimators for sampling variance, which is a measure of the precision of the estimator.

Let us consider the estimators of population mean $\bar{Y}$ or population total $Y = N\bar{Y}$ and estimators of sampling variances in case of some of the prevalent sampling designs.

**Simple random sampling (SRS):**

For both with replacement (WR) as well as without replacement (WOR) cases, sample mean $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$ is an unbiased estimator of population mean $\bar{Y}$ .

Estimator of sampling variance in case of SRS WR is given by

$\hat{V}(\bar{y}) = \frac{s^2}{n}$; where $s^2$ is the sample mean square given by $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2$ .

In case of SRSWOR, estimator of variance is given by

$$\hat{V}(\bar{y}) = \left(\frac{1}{n} - \frac{1}{N}\right)s^2$$

**Systematic sampling:**

For linear systematic sampling, with N being a multiple of n, systematic sampling is an EOSEM procedure and sample mean $\bar{y}$ is an unbiased estimator of population mean $\bar{Y}$ .

Unbiased estimation of variance is not possible in this case. However some approximations are available. One such approximation is as follows:

$$\hat{V}(\bar{y}) = \frac{1}{2}\left(\frac{1}{n} - \frac{1}{N}\right)\frac{1}{n-1}\sum_{i=1}^{n-1}\left(y_{i+1} - y_i\right)^2$$

However, if it is assumed that population is randomly distributed, then the same expression as used in case of simple random sampling may be used , i.e.

$$\hat{V}(\bar{y}) = \left(\frac{1}{n} - \frac{1}{N}\right)s^2$$

**Probability proportional to size (with replacement) sampling:**

For PPSWR, estimator of population total is given by

$$\hat{Y}_{pps} = \frac{1}{n}\sum_{i=1}^{n}\frac{y_i}{p_i},$$

where $p_i$'s are the initial probabilities of selection. Unbiased estimator of the sampling variance is given by,

$$\hat{V}(\hat{Y}_{pps}) = \frac{1}{n(n-1)} \sum_{i=1}^{n} \left( \frac{y_i}{p_i} - \hat{Y}_{pps} \right)^2$$

**Varying probability sampling (without replacement):**

Most common estimator of population total Y in case of PPS WOR schemes is due to Horvitz and Thompson and is given as follows

$$\hat{Y}_{HT} = \sum_{i=1}^{n} \frac{y_i}{\pi_i}$$

where $\pi_i$ is the selection probability or the probability of inclusion of $i^{th}$ population unit in the sample. Calculation of inclusion probabilities in general PPSWOR schemes is quite complicated and efforts have been made to either suggest estimators which do not require calculation of inclusion probabilities or to suggest varying probability without replacement schemes in which selection probabilities are proportional to size measures used for selection. These schemes are known as IPPS or $\pi ps$ schemes. One of the IPPS schemes is PPS-systematic sampling, which has already been described in the section 5.3. In IPPS schemes, $\pi_i = np_i$. Thus,

$$\hat{Y}_{HT} = \frac{1}{n} \sum_{i=1}^{n} \frac{y_i}{p_i}$$

which is of the same form as the estimator in PPSWR case. As in the case of systematic sampling, in this case also unbiased estimation of sampling variance is not possible. However, variances are estimated under some approximations and assumptions. Quite often, software packages for estimation of variances for complex sample surveys are used. Some of the methods used are sample re-use procedures, which are quite computer intensive methods. Simple expressions for estimators of variances are not available.

**Stratified sampling:**

In stratified sampling, samples are selected independently within each stratum. The estimation procedure depends on the method of sampling used within each stratum. Here, we consider the estimation procedure when SRSWOR method has been used within strata. An unbiased estimator of population mean is given as

$$\hat{\bar{Y}}_{st} = \sum_{h=1}^{L} W_h \bar{y}_h$$

An unbiased estimator of sampling variance is given by

$$\hat{V}(\hat{\bar{Y}}_{st}) = \sum_{h=1}^{L} W_h^{\,2} \left( \frac{1}{n_h} - \frac{1}{N_h} \right) s_h^{\,2}$$

The estimators will depend on the method of sampling used within strata. If it is PPS selection, the formulae will change accordingly.

**Cluster sampling:**
We consider the case of equal clusters of size M each. Let n clusters be selected from N clusters using SRSWOR. Define

$Y_{ij}$ = value of the character under study for j$^{th}$ unit in the i$^{th}$ cluster
$Y_{i.}$ = total for the i$^{th}$ cluster
$Y_{..}$ = total of the y-values for all the units in the population
$\overline{Y}_i$ = per unit i$^{th}$ cluster mean
$y_{i.}$ = i$^{th}$ sample cluster total
$\overline{y}_i = \frac{1}{M}\sum_{i=1}^{M} y_{ij}$ = per unit i$^{th}$ sample cluster mean
$\overline{y}_c = \frac{1}{n}\sum_{i=1}^{n} y_{i.}$ mean per cluster in the sample
$\overline{Y} = \frac{1}{NM}\sum_{i=1}^{N}\sum_{j=1}^{M} y_{ij}$
$\overline{Y}_c = \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{M} y_{ij} = M\overline{Y}$ = Population mean per cluster

An unbiased estimator of population mean is

$$\hat{\overline{Y}}_{cl} = \frac{1}{n}\sum_{i=1}^{n} \overline{y}_i$$

Variance of this estimator is given by

$$V(\hat{\overline{Y}}_{cl}) = (\frac{1}{n} - \frac{1}{N})\frac{1}{N-1}\sum_{i=1}^{N}(\overline{Y}_i - \overline{Y})^2$$

An estimator of variance is given by

$$\hat{V}(\hat{\overline{Y}}_{cl}) = (\frac{1}{n} - \frac{1}{N})\frac{1}{n-1}\sum_{i=1}^{n}(\overline{y}_i - \hat{\overline{Y}}_{cl})^2$$

For unequal clusters also estimation procedure is available, but several alternative estimators are considered depending upon whether population total is known or not.

An alternative form for the $V(\hat{\overline{Y}}_{cl})$ is approximately given as

$$V(\hat{\overline{Y}}_{cl}) = \frac{S^2}{n}\{1 + (M-1)\rho\}$$

where $\rho$ is the intra class correlation. In fact this very form of the variance leads to the well known form of Design Effect as $\{1 + (M-1)\rho\}$.

**Multi stage sampling**
Consider the case of two stage sampling of unequal PSUs where selection at both the stages is done with SRSWOR. Define the following

N = number of PSUs in the population
n = number of PSUs selected in the sample
$M_i$ = number of SSUs in the $i^{th}$ PSU
$m_i$ = number of SSUs selected in the $i^{th}$ PSU
$Y_{ij}$ = value of the study variable y for the $(ij)^{th}$ SSU
$y_{ij}$ = value for the study variable in the $j^{th}$ selected SSU in the $i^{th}$ selected PSU
$Y_i$ = total of y values in the $i^{th}$ PSU
Y = total of y values in the entire population
$\bar{Y}_i$ = Mean per SSU in the $i^{th}$ PSU
$\bar{y}_i$ = mean per SSU as obtained in the sample

An unbiased estimator of the population total Y is given by

$$\hat{Y}_{ts} = \frac{N}{n}\sum_{i=1}^{n}\frac{M_i}{m_i}\sum_{j=1}^{m_i} y_{ij} = \frac{N}{n}\sum_{i=1}^{n} M_i \bar{y}_i$$

Variance of this estimator is given by

$$V\left(\hat{Y}_{ts}\right) = N^2\left(\frac{1}{n}-\frac{1}{N}\right)S_{bt}^2 + \frac{N}{n}\sum_{i=1}^{N} M_i^2\left(\frac{1}{m_i}-\frac{1}{M_i}\right)S_i^2$$

where $S_{bt}^2 = \frac{1}{N-1}\sum_{i=1}^{N}\left(Y_i - \frac{1}{N}\sum_{i=1}^{N} Y_i\right)^2$ and

$$S_i^2 = \frac{1}{M_i - 1}\sum_{j=1}^{M_i}\left(Y_{ij} - \frac{1}{M_i}\sum_{j=1}^{M_i} Y_{ij}\right)^2$$

Estimator of variance is given by

$$\hat{V}\left(\hat{Y}_{ts}\right) = N^2\left(\frac{1}{n}-\frac{1}{N}\right)s_b^{\ 2} + \frac{N}{n}\sum_{i=1}^{n} M_i^2\left(\frac{1}{m_i}-\frac{1}{M_i}\right)s_i^2$$

where $s_{bt}^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(\hat{Y}_i - \frac{1}{n}\sum_{i=1}^{n}\hat{Y}_i\right)^2$ ; $\hat{Y}_i = \frac{1}{m_i}\sum_{j=1}^{m_i} y_{ij}$

and $s_i^2 = \frac{1}{m_i - 1}\sum_{j=1}^{m_i}\left(y_{ij} - \frac{1}{m_i}\sum_{j=1}^{m_i} y_{ij}\right)^2$

In case of equal clusters where $M_i$=M and $m_i$=m, if mean is to be estimated the above formulae reduce to a simpler form as follows:

$$\hat{\bar{Y}} = \frac{1}{nm}\sum_{i=1}^{n}\sum_{j=1}^{m} y_{ij} = \sum_{i=1}^{n} \bar{y}_i$$

$$V\left(\hat{\bar{Y}}\right) = \left(\frac{1}{n} - \frac{1}{N}\right) S_b^2 + \frac{1}{n}\left(\frac{1}{m} - \frac{1}{M}\right) \bar{S}_w^2$$

where $S_b^2 = \frac{1}{N-1}\sum_{i=1}^{N}\left(\bar{Y}_i - \bar{Y}\right)^2$

and $\bar{S}_w^2 = \frac{1}{N}\sum_{i=1}^{N} S_i^2$ ; with $S_i^2 = \frac{1}{M-1}\sum_{j=1}^{M}\left(Y_{ij} - \bar{Y}_i\right)^2$

In the variance formula as given above, the two components denote the contributions towards the total sampling variance due to PSUs and SSUs respectively. This splitting the variance into parts representing different stages of selection is very helpful in optimizing the sample sizes for two stages.

Estimator of variance is given by

$$\hat{V}\left(\hat{\bar{Y}}\right) = \left(\frac{1}{n} - \frac{1}{N}\right) s_b^2 + \frac{1}{N}\left(\frac{1}{m} - \frac{1}{M}\right) \bar{s}_w^2$$

where

$$s_b^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(\bar{Y}_i - \bar{Y}\right)^2$$

$$\bar{s}_w^2 = \frac{1}{n}\sum_{i=1}^{n} s_i^2 \text{ ; with } s_i^2 = \frac{1}{m-1}\sum_{j=1}^{m}\left(y_{ij} - \bar{y}_i\right)^2$$

If N is large enough to ignore terms of $O\left(\frac{1}{N}\right)$, we obtain a simple expression for estimator of variance as

$$\hat{V}\left(\hat{\bar{Y}}\right) = \frac{s_b^2}{n}$$

**Role of sample weights in estimating population totals/means**

It may be observed from the above discussions that sampling weights have got important roles to play in estimation of various parameters. Quite often we are interested in parameters like totals and means which are linear in nature. The estimates for such parameters are also linear in nature with sample observations suitably weighted with appropriate sampling weights. In agricultural censuses, since sampling is done for small and medium agricultural households only, the weighting procedure is considered only for such households. Large farms and institutional holdings are anyway completely enumerated. Corresponding weights for such holdings will be one only. The weighting procedure is essentially based on following three types of weights:

1) Base weights
2) Non response adjustments
3) Post-stratification adjustments

**Base weights**

It may be observed that in varying probability sampling without replacement, Horvitz-Thompson estimator is given as:

$$\hat{Y}_{HT} = \sum_{i=1}^{n} \frac{y_i}{\pi_i} = \sum_{i=1}^{n} w_i y_i$$

where the weights are inverse of selection probabilities of the ultimate units. In agricultural censuses, operational holdings are the units of observation. Since the selection probabilities are associated with the units of selection, which are agricultural households in this case, the agricultural households associated with the holding provide the base weights for the holdings. If there is one to one correspondence between agricultural households and holdings, the selection probabilities of the holdings are straightaway the selection probabilities of the agricultural households. There is no problem for multiple-holding households, as the selection probabilities for such holdings are straight forward. In the cases where one operational holding corresponds to several agricultural households, pro-rata adjustments may be done by considering pseudo-holdings corresponding to each household. However, such cases are likely to be very rare.

In two stage sampling, the selection probability of a SSU is the product of selection probability of corresponding PSU and the conditional selection probability of SSU for the given PSU. In the present case, EAs are PSUs which are selected with pps systematic sampling and agricultural households are SSUs which are selected with equal probability sampling.

Let $\pi_i$ be the probability of selection for i$^{th}$ PSU (i.e. EA) and $\pi_{j|i}$ be the conditional probability for selecting j$^{th}$ SSU (household) in i$^{th}$ PSU, then the probability of selection for j$^{th}$ SSU in i$^{th}$ PSU is given by $\pi_{ij} = \pi_i \pi_{j|i}$. In this case,

$\pi_i = n\frac{X_i}{X}$; $X_i$ is the measure of size (number of Agricultural households in i$^{th}$ EA as per 2007/08 PHC) and $X$ is the sum of $X_i$ in the specific stratum to which i$^{th}$ EA belongs. Also, $\pi_{j|i} = \frac{m}{M_i}$ where $M_i$ is the number of agricultural households in i$^{th}$ EA as observed at the time of field work for preparing the frame and m is the number of households selected in each EA.

Thus, $\pi_{ij} = \frac{nmX_i}{XM_i}$

In case when $X_i = M_i$, $\pi_{ij} = \frac{nm}{X}$ and the sample design is EPSEM. However, when $X_i \neq M_i$, the design is no more EPSEM and the base should be calculated carefully.

In general, the Base weights for each household in the i$^{th}$ EA is $\frac{XM_i}{nmX_i}$

**Non-response adjustment**

Invariably, there is some amount of non-response in every survey, which disturbs the weights. Therefore there is a need for adjusting for non-response. Normally, the non-

response adjustments are done within each EA. The adjustment factor is (m/r), where m is the number of sampled holdings while r is the number of responding households.

**Post-stratification adjustment**

Sometimes it is felt desirable that the estimated totals for certain characteristics (auxiliary variables) in some population groups (which may as well be post-strata) are in conformity with the known totals for these groups. Some characteristics from PHC 2007 for which information is also collected in CAP-II may serve as a suitable variable for this adjustment. For example, number of households in a district may be known from PHC. An estimate for this characteristic may also be developed from the survey. The weights may be adjusted in such a way that the estimated value is equal to the known value fro PHC. This type of adjustment provides a check on the face value of the estimates with respect to known characteristics. Since the auxiliary characteristic is also correlated to the main study variable, the adjustment is also expected to provide more reliability to the estimates.

The final weights are the product of Base weight, non-response adjustment and the post-stratification adjustment.

# SESSION 6

## ESTIMATION OF SAMPLING ERRORS

**7. Variance Estimation**

As described in the previous section, estimation of variance should be an integral part of estimation procedure. Estimation of variance provides a measure of precision for the estimates. It provides a level of confidence to the estimates developed.

*Coefficient of Variation:* Another convenient method for measuring precision of an estimator for a parameter (say, $\bar{Y}$ ) is Coefficient of Variation (CV)**,** which is defined as follows

$$C.V. = \left( \frac{\sqrt{sampling\ \mathrm{var}iance}}{Parameter} \right)$$

and is estimated as

$$Est(CV) = \left( \frac{Est(SE)}{Estimate} \right)$$ ; where SE is the Standard Error.

Since it is a unit free measure, it is often used to compare the precision levels of estimators in different populations.

*Confidence Intervals:* Another useful concept associated with precision levels is the concept of Confidence Intervals. The concept of confidence limits and confidence interval is closely linked to interval estimation. A point estimate is a single value given as the estimate of a population parameter that is of interest, for example the mean of some quantity. An interval estimate specifies instead a range within which the parameter is estimated to lie. A confidence interval (CI) can be used to describe how reliable survey results are. For a given point estimate, 90% or 95% confidence intervals can be generated depending upon whether the level of confidence is 10% or 5%. When mean is to be estimated, in case of simple random sampling, sample mean $\bar{y}$ is a point estimate for population mean $\bar{Y}$ and the confidence limits are $\bar{y} \pm t_{\alpha} SE(\bar{y})$, where $t_{\alpha}$ is the t-value with $(1-\alpha)$ per cent level of confidence. At a given level of confidence, and all other things being equal, a result with a smaller CI is more reliable than a result with a larger CI. A major factor determining the width of a confidence interval is the size of the sample used in the estimation procedure.

Confidence intervals are closely related to statistical significance testing. In many situations, if the point estimate of a parameter is *X*, with confidence interval [*a*,*b*] at confidence level *P*, then any value outside the interval [*a*,*b*] will be significantly different from *X* at significance level α = 1 − *P*, under the same distributional assumptions that were made to generate the confidence interval.

**Variance Estimation in Complex Surveys**

In the estimation procedures corresponding to different sampling designs as described above, formulae for estimates of variances for estimated means/totals are provided. In the case of linear estimates it is simple. However, in more complex survey situations, it is not

always possible to express the estimated variances in terms of simple formulae. Even in more familiar situations of estimating variances for ratio and regression estimators, which are non linear in nature, expressing variance estimators, similar to linear estimators is not feasible.

Several alternate methods for estimating variances in complex survey situations are available. Some of these methods are

1) Linearization (Taylor's series)
2) Random Group Methods
3) Balanced Repeated Replication (BRR)
4) Re-sampling techniques
   - Jackknife, Bootstrap

**Taylor's Series Linearization Method**

In this method, non-linear statistics are approximated to linear form using Taylor's series expansion. This involves expressing the estimate in terms of a Taylor's series expansion, and then approximating the variance of the estimate by the variance of the first-order or linear part of the Taylor series expansion. This method requires the assumption that all higher-order terms are of negligible size. If this assumption is correct, then the variance approximation works well. In this linearization approach to variance estimation, a separate formula for the linearized estimate must be developed for each type of estimator. We are already familiar with this approach in a simple form in case of ratio estimator.

**Random Group Methods**

This concept is based on the concept of replicating the survey design. The earliest form of this method is available in the concept of Interpenetrating samples. However it is usually not possible to replicate the survey. In such cases, survey can be divided into R groups so that each group forms a miniature version of the survey. Based on each of the R groups estimates can be developed for the parameter of interest $\theta$,(say). Let $\hat{\theta}_r$ be the estimate based on $r^{th}$ sample. Considering the groups as independent, an unbiased estimate of variance of $\hat{\theta} = \frac{1}{R}\sum_{r=1}^{R}\hat{\theta}_r$ is given as

$$\hat{V}\left(\hat{\theta}\right) = \frac{1}{R(R-1)}\sum_{r=1}^{R}\left(\hat{\theta}_r - \hat{\theta}\right)^2$$

Advantages of this method are that it is easy to calculate and it is a general method in which complex functions can be tackled easily. However, the assumption of independent samples may be somewhat restrictive, if samples are not selected independently.

**Balanced Repeated Replications (BRR) methods**

Consider that there are H strata with two units selected per stratum. There are $2^H$ ways to pick 1 from each stratum. Each combination could be treated as a sample. Pick R samples. Which samples should we include? Following steps may be followed:

- Assign each value either 1 or –1 within the stratum
- Select samples that are orthogonal to one another to create balance
- One can use the design matrix for a fraction factorial
- Specify a vector $\alpha_r$ of 1,-1 values for each stratum

An estimator of variance based on BRR method is given by

$$\hat{V}_{BRR}\left(\hat{\theta}\right) = \frac{1}{R(R-1)} \sum_{r=1}^{R} \left(\hat{\theta}(\alpha_r) - \hat{\theta}\right)^2$$

where $\hat{\theta} = \frac{1}{R} \sum_{r=1}^{R} \hat{\theta}(\alpha_r)$

Some of the advantages in the method is that they are relatively few computations ant it is asymptotically equivalent to linearization methods for smooth functions of population totals.

**Jack-knife Method**

The method was initially developed (Quenouille (1949)) in the context of reducing the bias of ratio estimator. The procedure was to randomly divide the sample (SRS) in two parts. Each part could provide an estimate of the estimator (ratio estimator in this case). The third estimate could be developed from the whole sample. These three estimators could be combined linearly such that first order term in the bias expression vanishes. If units are dropped individually, the corresponding statistics was found (conjectured first by Tukey (1958)) to be uncorrelated. This property has been exploited for variance estimation.

Let $\hat{\theta}^i$ be the estimator of θ after omitting the i$^{th}$ observation. Define

$$\tilde{\theta}^i = n\hat{\theta} - (n-1)\hat{\theta}^i$$

The Jackknife estimate is given by $\hat{\theta}_J = \frac{1}{n} \sum \tilde{\theta}^i$

Jackknife estimator of the variance is given by

$$\hat{V}_J(\hat{\theta}_J) = \frac{1}{n(n-1)} \sum_{i=1}^{n} (\tilde{\theta}^i - \hat{\theta}_J)^2$$

**Bootstrap Method**

This is also a re-sample technique, in which large numbers of samples are selected by equal probability sampling with replacement from the main sample. Similar to the estimate for the main sample, independent estimates are prepared for each sample. Estimate of variance is obtained from the repeated samples. An advantage of this method is that estimates of variance for complex statistics like quantiles and median can be obtained.

All these variance estimation techniques are highly computer intensive. Most of the survey data analysis packages utilize one of these methods. Some of these packages are as follows:

- OSIRIS – BRR, Jackknife
- SAS – Linearization
- Stata – Linearization
- SUDAAN – Linearization, Bootstrap, Jackknife
- WesVar – BRR, JackKnife, Bootstrap

# SESSION 7

## EVALUATION AND TREATMENT OF NON-SAMPLING ERRORS

## 8. Non Sampling Errors (NSEs)

In the previous discussions on survey design and estimation methodology, the focus was on sampling errors only. There are, however, other sources of variation in surveys caused by non-sampling errors. All survey data are subject to error from various sources. The difference in the true value of the parameter and survey results is an error due to one reason or the other. The sampling variance and mean square errors are measures of error due to sampling. All other types of errors from various sources are termed as non-sampling errors. Non-sampling errors arise mainly due to misleading definitions and concepts, inadequate frames, unsatisfactory questionnaires, defective methods of data collection, tabulation, coding, incomplete coverage of sample units etc. Sampling errors arise solely as a result of drawing a probability sample rather than conducting a complete enumeration. Non-sampling errors, on the other hand, are mainly associated to data collection and processing procedures.

### *8.1 Types of Non-sampling Errors*

Non-sampling errors arise due to various causes right from initial stage when the survey is being planned and designed to the final stage when data are processed and analyzed. Some of the factors contributing towards Non-sampling error are as follows:

1) Data specification being inadequate and/or inconsistent with respect to objectives of the survey.
2) Duplication or omission of units due to imprecise definition of the boundaries of area units, incomplete or wrong identification particulars of units or faulty methods of enumeration.
3) Inappropriate methods of interview, observation or measurement using ambiguous questionnaires, definitions or instructions.
4) Lack of trained and experienced field enumerators including lack of good quality field supervision
5) Inadequate scrutiny of the basic data.
6) Errors in data processing operations such as coding, keying, verification, tabulation etc.
7) Errors during presentation and publication of tabulated results.

Five prominent components of NSEs are known as:

1. Specification errors,
2. Coverage errors,
3. Measurement or response errors,
4. Non-response errors and
5. Processing error.

These types or error are briefly discussed below:

#### *8.1.1 Specification errors*

This occurs when the concept implied by the question is different from the underlying construct that should be measured. A simple question such as whether a household is an agricultural household can be subject to different interpretations. A person may be doing agriculture as an own account holder, he may be involved in agricultural activities as a part time activity. The meaning of the questions must be conveyed in an unambiguous way and must be properly understood by the respondent. Unless the right screening and filter questions are included in the questionnaire, the answers may not fully bring out the message behind the question.

*8.1.2 Coverage errors*

In most area surveys primary sampling units comprise clusters of geographic units generally called enumeration areas (EAs). It is not uncommon that the demarcation of EAs is not properly carried out during census mapping. Thus households may be omitted or duplicated in the second stage frame. Updating of EA boundaries before the conduct of agricultural census becomes very important. Cartography of EAs is normally available from the population censuses, but updating of the selected EAs is an essential part of the cartography for agricultural censuses. Otherwise, exclusion of sample units in some EAs and duplication of units in other EAs are highly probable. Frame imperfections can bias the estimates in several ways: If units are not represented in the frame but should have been part of the frame, this results in zero probability of: selection for those units omitted from the frame. This leads to under-coverage. On the other hand if some units are duplicated; this results in over-coverage with such units having larger probabilities of selection.

It is important to note that sometimes there is a deliberate and explicit exclusion of sections of a larger population from survey population. Survey objectives and practical difficulties determine such deliberate exclusions. For example, when we define the agricultural households by putting certain cut-offs, some households are deliberately excluded. When computing non-coverage rates, members of the group deliberately and explicitly excluded should not be counted either in the survey population or under non-coverage. In this regard defining the survey population should be part of the clearly stated essential survey conditions. Non-coverage is often associated with problems of incomplete and faulty frames. If the flames are not updated or old frames are used as a device to save time or money, it may lead to serious bias.

The most effective way to reduce coverage error is to improve the frame by excluding erroneous units and duplicates and updating the frame through filed work to identify units missing from the frame. It is also important to undertake a good mapping exercise during the preparatory stages of a population and housing census. However, the frame prepared during the census should be updated periodically. It is also imperative to put in place procedures that will ensure the coverage of all selected sample units.

*8.1.3 Measurement errors*

These errors arise from the fact that what is observed or measured departs from the actual values of sample units. These errors centre on the sustentative content of the survey such as definition of survey objectives, their transformation into able questions, and the obtaining, recording, coding and processing of responses. These errors concern the accuracy of measurement at the level of individual units. When we get responses from the selected units through a questionnaire and there the responses are different than the true values, these

errors are called response errors. Inadequate instructions to field staff and inadequate training normally lead to response errors.
Mathematical treatment of measurement errors is available in the form of linear response error models (Refer Cochran, W. G. (1977)). Such models have also been used in the treatment of interpenetrating net-work of sub-sampling which is used for estimating the enumerators' effect. The mathematical details are not given here.

*8.1.4 Non-response errors*

Non-response refers to the failure to measure some of the sample units. Thus failure to obtain observations on some units selected for the sample. It is instructive to think of the sample population as split into two strata, one consisting of all sample units for which measurements can be obtained and the second for which no measurements could be obtained.
In most cases non-response is not evenly spread across the sample units but is heavily concentrated among subgroups. As a result of differential non-response, the distribution of the achieved sample across the subgroups will deviate from that of the selected sample. This deviation is likely to give rise to non-response bias if the survey variables are also related to the subgroups. While non-response can not be completely eliminated in practice, it could be overcome to a great extent by persuasion or by some other methods. One way of dealing this problem was due to Hansen and Hurwitz (1946). In this method the population was conceived as divided in two strata – respondents and non respondents. From the non respondents, a sub-sample is selected and special efforts are made to get response from these units. An estimation procedure is developed on the basis of suitably pooling the results of respondent and non-respondent groups. Yet another technique was developed by Politz and Simon (1949) for reducing the bias without call backs by asking to the respondent as to how many times he was at home during previous week.
There are two types of non-responses: unit non-response and item non-response. Unit non-response implies that no information is obtained from certain sample units. This may be because respondents refuse to participate in the survey when contacted or they cannot be contacted. Item non-response refers to a situation where for some units the information collected is incomplete. Item non-response is therefore, evidenced by gaps in the data records for responding sample units. Reasons may be due to refusals, omissions by enumerators and incapacity.

*Causes of non-response*

Respondents to provide information can cause non-response error if they are being not- at home or by sample units not being accessible. This introduces errors in the survey results because sample units excluded may have different characteristics from the sample units for which information was collected. Refusal by a prospective respondent to take part in a survey may be influenced by many factors, among them, lack of motivation, shortage of time, sensitivities of the study to certain questions, etc.
Errors arise from the exclusion of some of the units in the sample. This may not be a serious problem if the characteristics of the non-responding units are similar to those of the responding units. But such similarity is not common in practice.
With specific reference to item non-response, questions in the survey may be perceived by the respondent as being embarrassing, sensitive or/and irrelevant to the stated objective. The enumerator may skip a question or ignore recording an answer. In addition, a response

may be rejected during editing. For sensitive questions a technique of randomized response is available.

In personal interview surveys, the enumerator can play an important role in maximizing response from respondents. The way interviewers introduce themselves, what they say about the survey, the identity they carry, and the courtesy they show to respondents matter. In most surveys the enumerator is the only link between the survey organization and respondent. It is for this reason that enumerators and their supervisors should be carefully selected, well trained and motivated. Close supervision of enumerator's work and feedback on achieved response rate is of paramount importance.

*8.1.5 Processing errors*

Processing errors comprise:

— Editing errors.
— Coding errors.
— Data entry errors.
— Programming errors etc.

The above errors arise during the data processing stage. For example in coding open ended answers related to economic characteristics, coders may deviate from the laid out procedures in coding manuals, and therefore assign wrong codes to occupations. In addition, the weighting procedures may be wrongly applied during the processing stage, etc.

### *8.2 Interpenetrating sub-sampling*

It is worthwhile to mention about this technique which was initially developed (Mahalanobis (1946)) in the context of study of correlated errors. In this technique a random sample of n units is divided at random into k sub-samples, each sub sample containing m=n/k units. The field work and processing of the sample are planned so that there is no correlation between the errors of measurement of any two units of in two different sub-samples. For instance, suppose that the correlation with which we have to deal arises solely from biases of the enumerators. If each of k enumerators is assigned to a different sub-sample and if there is no correlation between errors of measurement for different interviewers, we have an example of this technique. With a suitable model it is possible to estimate the relative amount which the correlated component (in this case due to interviewer's effect) of the response variance contributes to the total variance.

The technique has also been very helpful in estimation of variances for complex statistics.

### *8.3 Evaluation of non-sampling errors*

<u>*Consistency checks*</u>

In designing the survey instruments (questionnaires), care should be taken to include certain items of information that will serve as a check on the quality of the data to be collected. If the additional items of information are easy to obtain, they may be canvassed for all units covered in the survey, otherwise, they may be canvassed only for a sub-sample of units.

It is also desirable to follow some external consistency checks on salient results thorough comparable data sources. It is important for validity as well as acceptability of the estimates.

*Sample check/verification*

One way of assessing and controlling non-sampling errors in surveys is to independently duplicate the work at the different stages of operation with a view to facilitating the detection and rectification of errors. For practical reasons the duplicate checking can only be carried out on a sample of the work by using a smaller group of well-trained and experienced staff If the s is properly designed and if the checking operation is efficiently carried out, it would be possible, not only to detect the presence of non-sampling errors, but also to get an idea of their magnitude. If it were possible to completely check the survey work, the quality of the final results could be considerably improved. With the sample check, rectification work can only be carried out on the sample checked. This difficulty can be overcome by dividing the output at different stages of the survey, e.g. filled in schedules, coded schedules, computation sheets, etc., into lots and checking samples from each lot. In this case, when the error rate in a particular lot is more than the specified level, the whole lot may check and corrected fir the errors, thereby improving the quality of the final results.

*Post-survey checks*

An important sample check, which may be used to assess non-sampling errors, consists of selecting a sub-sample, or a sample in the case of a census, and re-enumerating it by using better trained and more experienced staff than those employed for the main investigation.

Usually the check-survey is designed to facilitate the assessment of both coverage and content errors. For this purpose, it is first desirable to re-enumerate all the units in the sample at the high stages, e.g. EAs and villages, with the view of detecting coverage errors and then to re survey only a sample of ultimate units ensuring proper representation for different parts of the population which have special significance from the point of view of non-sampling errors.

# SESSION 8

## USE OF AGRICULTURAL CENSUS AS BASIS FOR ANNUAL CROP FORECAST AND PRODUCTION SURVEYS

# Agricultural Census & Sample Surveys of U.S. Agriculture

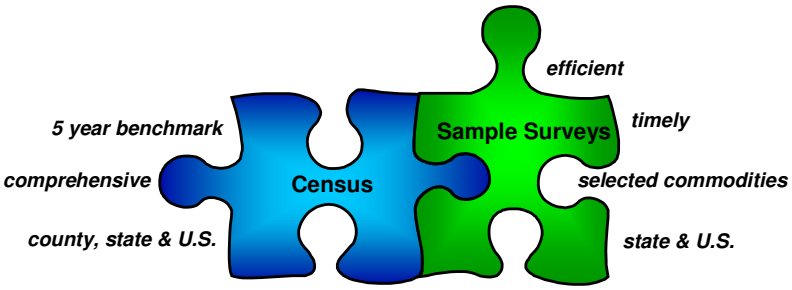
presented by
Kevin L. Barnes
National Agricultural Statistics Service
U.S. Department of Agriculture

FAO/NASS/GSS Technical Workshop on Sampling
19-23 April 2010
ACCRA, Ghana

## Discussion Topics

- NASS view of census and sample surveys
- How NASS uses frames for census and sample surveys
- NASS survey program for major crops
- Examples of census based follow-on surveys conducted by NASS
- Challenges for census based surveys

efficiently use limited resources
to satisfy needs of data users






## Census of Agriculture

- uniform, comprehensive statistics on farms & farm productivity for every county in the U.S. every 5 years

## Agricultural Census

all farms

total area & land use
irrigation
land in government programs
field & forage crops
fruits, nuts & berries
vegetables & melons
horticultural specialties

livestock & poultry
animal specialties
aquaculture
production contracts
gross value of sales
direct sales to consumers

government loans
government program payments
farm-related income
grain storage capacity
operator characteristics
farm organization

fertilizer & chemical use
farm production expenses
inventory & value of machinery & equipment
market value of land & buildings
farm labor

~ years ending in “2” & “7” ~

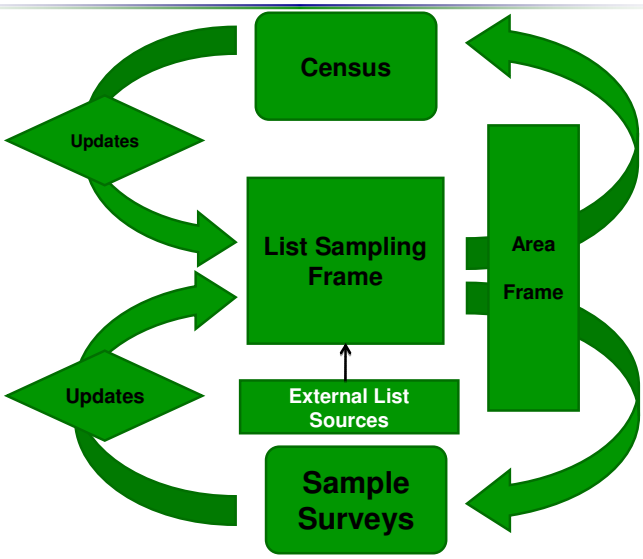
*inventories as of December 31 ~ production, sales & other information for calendar year*

## U.S. Census of Agriculture

NASS maintains a list of farmers and ranchers from which the Census Mail List (CML) is compiled.

- Goal: Build as complete a list as possible containing all farms that meet the NASS definition of a farm.
- The area frame accounts for farms not on the list frame.

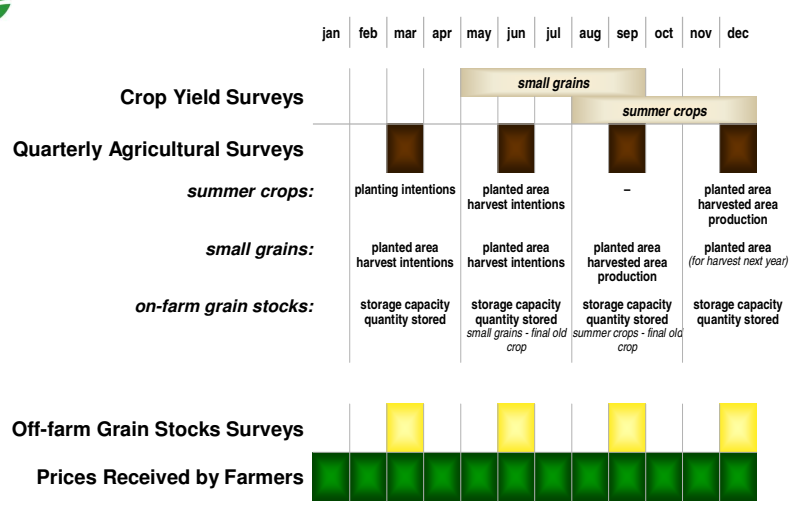
## U.S. Census of Agriculture




## U.S. Census of Agriculture

| Central Data Name | Tgeoid | Value | Unit | Date Recorded | Source | Source Name | Sequence |
|---|---|---|---|---|---|---|---|
| TOTAL FARMS | 108 | 10.0 | ACRES | 12/01/2007 | 202 | CENSUS MAIL LIST | 25 |
| LAND IN FARMS | 111 | 10.0 | ACRES | 12/01/2007 | 202 | CENSUS MAIL LIST | 99 |
| CALCULATED FARM VALUE SALES | 115 | 2,250,000 | DOLLARS | 12/01/2007 | 202 | CENSUS MAIL LIST | 99 |
| CALCULATION TYPE OF FARM | 180 | 15.00 | LI - LIST INDICATOR | 01/01/2010 | 27 | PROFILER CALCULATION | 70 |
| CENSUS MODEL WEIGHTS (DECIMAL) | 242 | 0.5029 | WEIGHT/PCT - PERMANENT | 01/01/2007 | 5 | LIST FRAME MASTER | 242 |
| CENSUS TOTAL VALUE OF PRODUCTS | 243 | 2,250,000 | DOLLARS | 12/01/2007 | 202 | CENSUS MAIL LIST | 243 |
| CENSUS TENURE CODE | 254 | 1 | LI - LIST INDICATOR | 12/01/2007 | 202 | CENSUS MAIL LIST | 99 |
| CENSUS FARM TYPE | 257 | 6 | LI - LIST INDICATOR | 12/01/2007 | 202 | CENSUS MAIL LIST | 257 |
| CENSUS CONGRESS_STATE_CODE | 270 | 41 | LI - LIST INDICATOR | 12/01/2007 | 202 | CENSUS MAIL LIST | 99 |
| CENSUS CONGRESS_DISTRICT_NUMBER | 271 | 1.0000 | WEIGHT/PCT - PERMANENT | 12/01/2007 | 202 | CENSUS MAIL LIST | 270 |
| PERMANENT WORKLOAD PRIORE ID | 274 | 40.0 | ACRES | 01/01/2007 | 202 | CENSUS MAIL LIST | 274 |
| TOTAL CROPLAND | 280 | 10.0 | ACRES | 12/01/2007 | 202 | CENSUS MAIL LIST | 60 |
| TOTAL IRRIGATED CROPLAND | 281 | 10.0 | ACRES | 12/01/2007 | 202 | CENSUS MAIL LIST | 395 |
| FARM WORKERS HIRED | 885 | 20 | NUMBER | 12/01/2007 | 202 | CENSUS MAIL LIST | 885 |
| MARKETS HOG, FEED, SHORES, BY | 970 | 2,250,000 | DOLLARS | 12/01/2007 | 202 | CENSUS MAIL LIST | 99 |
| ANY MARKETED PRODUCTS INDICATOR | 939 | 1 | LI - LIST INDICATOR | 12/01/2007 | 202 | CENSUS MAIL LIST | 939 |
| TOTAL MARKETED PRODUCTS | 925 | 2,250,000 | DOLLARS | 12/01/2007 | 202 | CENSUS MAIL LIST | 925 |

## NASS Survey Program for Major Crops




## Census Based Follow-on Surveys

- 2008 Farm and Ranch Irrigation Survey (FRIS)
- 2008 Organic Production Survey (OPS)
- 2009 Census of Horticulture
- 2010 Tenure, Ownership & Transfer of Agricultural Land (TOTAL)
- 2006 Aquaculture Census
- 2009 On-Farm Renewable Energy Production Survey

## 2008 Farm and Ranch Irrigation Survey

- Conducted every 5 years
- Probability survey of all farms
- Sample of horticulture operations that reported irrigation on the census
- National sample size of 33,085 farms

## 2008 Farm and Ranch Irrigation Survey

Purpose: To provide data relating to on-farm irrigation activities for use in preparing a wide variety of water-related local programs, economic models, legislative initiatives, market analyses, and feasibility studies.

### Census form screening question

**SECTION 3** **IRRIGATION**

1. Were any of the acres on this operation irrigated by sprinklers, flooding, ditches or furrows, drip or trickle irrigation, etc. in 2007?

1065 1 ☐ **Yes** - Complete this section 3 ☐ **No** - Go to SECTION 4 below

**Farm & Ranch Irrigation Survey**

sample of operations reporting irrigation on the agricultural census

irrigated area
area, yield & quantity of water applied by crop
methods of on-farm water distribution
quantity of water used by source
number & description of irrigation wells & pumps
energy use
irrigation facilities & equipment costs
irrigation maintenance & repair costs
irrigation & water management practices
improvements to irrigation systems
recycled water use and systems

~years ending in “4” & “9” ~

**2008 Farm and Ranch Irrigation Survey**

## Results

- 207,000 farms irrigating
- 54.9 million acres irrigated
- 9,843 farms use recycled water
- 78,846 farms implemented changes in equipment or management
  - 46% reduced energy costs
  - 59% reduced water applied

## 2008 Organic Production Survey

First in-depth survey of organic farming in the United States.

Target Population:

- All operations with positive information reported in the organic section of the 2007 Census of Agriculture who met the standards of the USDA National Organics Program.

Total sample size was approximately 29,000

## 2008 Organic Production Survey

Purpose: To collect data from operators of farms that were either USDA-certified organic, were making the transition to organic production or were exempt from certification because of sales totaling less than $5000.

### Census form screening questions

**SECTION 22** ORGANIC AGRICULTURE

1. Did this operation produce organic products (according to the National Organic Standards) for sale in 2007?

1248 1 ☐ Yes - Complete this section 3 ☐ No - Go to item 7 below

all USDA certified organic farms

total land area
area harvested by crop
acres Irrigated
field & forage crops
fruits, nuts & berries
vegetables & melons

livestock & poultry
animal specialties
production and marketing practices
gross value of sales
direct sales to consumers
value added sales

fertilizer & chemical use
farm production expenses

2008 Organic Production Survey

## Results

- 14,540 organic farms and ranches in the United States, comprising of 4.1 million acres of land.
- $3.16 billion in total sales.
  - $1.94 billion in crop sales and $1.22 billion in sales of livestock, poultry and their products.

## 2009 Census of Horticultural Specialties

Ninth Census of Horticultural Specialties

Population:
All operations identified in the 2007 Census of Agriculture with sales of $10,000 or more of horticultural specialty crops.

Includes producers of floriculture, nursery, and other specialty crops, such as sod, and food crops produced under glass or other protection, transplants for commercial production, and seeds.

Conducted in conjunction with the Annual Commercial Floriculture Survey to reduce respondent burden and minimize data collection expenses.

## 2009 Census of Horticultural Specialties

Purpose: To obtain a comprehensive and detailed picture of the horticultural sector of the economy.

### Census form screening question

**SECTION 9** **NURSERY, GREENHOUSE, FLORICULTURE, SOD, MUSHROOMS, VEGETABLE SEEDS, AND PROPAGATIVE MATERIALS**

1. Were any nursery, floriculture, or greenhouse crops, including ornamental plants, flowers, mushrooms, aquatic plants, sod, food crops under protection, vegetable seeds, flower seeds, or other propagative materials, grown for sale on this operation in 2007? Include crops produced under contract and food crops grown in greenhouses, caves, and high tunnels where crops were always covered. Exclude personal or home use crops. Report food crops temporarily covered for early germination, frost protection, etc. in SECTION 10, 11, or 12.

1032 1 ☐ **Yes** - Complete this section 3 ☐ **No** - Go to SECTION 10

## Horticultural Census

establishments growing & selling
$10,000 or more of horticultural specialties

**units, wholesale & total sales:**

bedding plants
potted flowering plants
cut flowers
cut cultivated greens
foliage plants
deciduous trees
unfinished plant materials

**area, wholesale & total sales:**

nursery plants
turfgrass
bulbs
greenhouse fruits & vegetables
vegetable seeds
aquatic plants
Christmas trees
tobacco transplants

**area, production, wholesale & total sales:**

mushrooms
flower seeds
short term woody plants

**value of land & buildings**
**value of machinery & equipment**
**area & structure of greenhouses**
**area irrigated**

**total retail & wholesale sales**
**markets**
**production expenses**
**number of hired workers & wages**

~ years ending in "8" ~

## 2009 On-Farm Renewable Energy Production Survey

First national renewable energy production survey conducted by NASS

Population:

- All operations that selected "yes" to the energy question in the 'practices' section in the 2007 Census of Agriculture.

Sample size is approximately 16,000 operations.

## 2009 On-Farm Renewable Energy Production Survey

Purpose: To provide information covering electricity produced using wind turbines, solar panels, and manure/methane digesters.

### Census form screening question

**SECTION 32 PRACTICES**

1. At any time during 2007, did this operation -
   - a. Have Internet access? . . . . . 1603 1 ☐ Yes 3 ☐ No
   - b. Have high speed Internet access? . . . . . 1604 1 ☐ Yes 3 ☐ No
   - c. Use more than 5,000 gallons of water in any one day for any purpose? . . . 1721 1 ☐ Yes 3 ☐ No
   - d. Receive irrigation water supplied by a U. S. Bureau of Reclamation project or facility? Include reclamation water delivered by a local district. . . . . 1722 1 ☐ Yes 3 ☐ No
   - e. Generate energy or electricity on the farm using wind or solar technology, methane digester, etc.? . . . . . 1723 1 ☐ Yes 3 ☐ No

## 2010 Tenure, Ownership & Transfer of Agricultural Land Survey

| | |
|---|---|
| **Target Population:** | • Included the universe of land owners and farm operators |
| **Targeted Coverage:** | • 85%-95% of U.S. farm cash receipts |
| **Information to be collected for farm operators and their landlords** | |
| **Multiple Samples Approach:** | • Multiple frame sample for farm operators.<br>• Landlords who are not farm operators are accounted for by area sample only. |

## 2010 Tenure, Ownership & Transfer of Agricultural Land Survey

### sample of farm operators & their landlords

land owned, rented in & rented out
rent by method of payment
quantity & value of land acquisitions & sales
land use
market value of land & buildings
production expenses
capital expenditures
farm assets
farm income
farm debt
farm household characteristics
characteristics of farm owners
landlord participation in management decisions

~ periodic as follow-on to agricultural census ~

## Challenges For Census Based Surveys

Rapid changes in the agricultural sector can result in the census frame being outdated.

Sampling and stratification rely on the quality and accuracy of list frame control data.

Changes to the frame can cause difficulty in maintaining the integrity of the frame.

It can sometimes be a challenge to accurately define the sample population.