FAO STATISTICAL DEVELOPMENT SERIES 7

Multiple frame agricultural surveys

Volume 1

Current surveys based on area and list sampling methods

FAO STATISTICAL DEVELOPMENT SERIES 7

Multiple frame agricultural surveys

Volume 1
Current surveys based on area and list sampling methods

FOOD AND AGRICULTURE ORGANIZATION OF THE UNITED NATIONS
Rome, 1996

The designations employed and the presentation of material in this publication do not imply the expression of any opinion whatsoever on the part of the Food and Agriculture Organization of the United Nations concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries.

M-77
ISBN 92-5-103806-6

FOREWORD

Article 1 of the FAO Constitution states 'The Organization shall collect, analyse, interpret and disseminate information relating to nutrition, food and agriculture'. Much of the required information assembled by FAO for the agricultural sector, such as crop production, livestock inventories, and basic social and economic data is obtained from National Agricultural Statistical Systems. Data is collected through periodic, national multi-purpose, agriculture data collection programmes called 'Current Agricultural Surveys'.

The Statistics Division of FAO is charged with cooperating with member countries to improve consistency and quality of data and help develop and improve agriculture statistics. This publication "Multiple Frame Agricultural Surveys", Volume 1, has been developed to provide countries with a manual on new technology and survey methods to improve the national agricultural data collection programmes.

This project began with the Statistical Development Service conducting an International Expert Consultation on Multiple Frame Agricultural Surveys using a draft document prepared by Consultant, Mr. Montie Wallace, as a focus for discussion. In attendance were experts from ten countries and several international organizations. In addition, experts in several countries were contracted to prepare case studies on multiple frame surveys which will be published in Volume II.

FAO wishes to acknowledge the contributions of the staff of the Statistical Development Service and in particular that of Mr. A. González-Villalobos, Senior Officer, Statistical Development Service who coordinated the activities and co-authored with Messrs. Montie Wallace, Fred Vogel, Jim Cotter and Rick Kestle of the U.S. Department of Agriculture, National Agricultural Statistics Service. FAO is grateful to all who have provided input and would hope that member countries find that this manual leads to improvements in the conduct of their Current Agricultural Surveys.

The Director
Statistics Division

CONTENTS

INTRODUCTION

With the global shift towards market economies, the need for timely and reliable agricultural information has become more important than ever before in the decision making process at international and national levels. Much of the required information for the agricultural sector, such as crop production estimates, livestock inventories and basic social and economic data pertaining to the sector, is obtained through *Current Agricultural Surveys,* which are periodic (annual or seasonal), national (or large-scale), multiple-purpose, agricultural data collection programmes. The establishment and development of such survey programmes is, therefore, a fundamental component of the agricultural information system.

From one source or another, annual estimates are available for most types of agricultural commodities in almost every country. Moreover, several estimates may exist for the same commodity. Often these estimates are based on nothing more than an educated, or worse an uneducated, guess. There is usually no way to judge their accuracy because they are not based on any procedure that allows for statistical analysis or evaluation.

The reasons why current agricultural statistics are poor, for many countries are many. Lack of political support for data collection, the high cost of agricultural surveys, the shortage of requisite skills and the failure to identify the most appropriate methods are among the most commonly identified. Timely and reliable national statistics of a country's agricultural sector can only be provided by the establishment of an adequate, periodic, national agricultural survey based on probability sampling methods.

The *Current Agricultural Surveys* considered are based on probability sampling and estimation methods, which are the only ones that could provide timely and reliable basic data of a country's agricultural sector.

This manual describes the sample design, organization and implementation procedures of current agricultural surveys based on *Multiple Frame Probability Sampling Methods.*

Procedures are presented in the context of the considerations and steps needed to develop and maintain a current agricultural data collection programme in countries where experience with area frame and multiple frame sampling methods is lacking.

The intention is to introduce the subject in a straightforward and practical manner, and describe list and area frame construction and sample selection methods that require a minimum of resources and specialized staff taking into account the constraints faced by developing countries. In order to illustrate with specific examples the use of multiple frame methods in a range of countries and conditions, volume II will include summaries of multiple frame and area frame agricultural surveys conducted in the Albania, Brazil, Canada, the Czech Republic, France, Honduras, Italy, Morocco, Nicaragua, Pakistan, Spain and the United States. Briefer references will also be included to the area sample surveys undertaken in a number of other countries.

The survey designs proposed consider the world-wide increasing availability in recent years of satellite images, computerized area measurement and scale-transfer instruments, and even geographic information systems and hand-held global positioning systems, as tools for the application of the proposed multiple frame agricultural survey methods in developing countries. They may be the most practical way for a country to produce the required annual basic data for the agricultural sector.

The agricultural survey methods for developing countries presented are derived from those developed to conduct the multiple frame agricultural survey of the United States, which constitute a record and experience with multiple frame agricultural surveys of primary importance; and take into account the experience of applying area and multiple frame sampling methods in many countries in North Africa, Asia, North, Central and South America and Western Europe.

The Multiple Frame Agricultural Survey Design Considered

This manual describes the overall survey design of a *Current Agricultural Survey Programme*: a periodic (annual or seasonal), national (or large-scale), multiple-purpose, agricultural probability sample survey designed in order to obtain timely and reliable basic data for the agricultural sector.

Survey estimates are required, it is assumed, for most of the following *variables*: crop areas (prepared, planted and harvested), crop yields (forecast and achieved), crop production, livestock inventories, grain stocks, farming systems, cost of production, farm expenditures and social and economic characteristics of the agricultural holdings. The determination of the *variables* to be studied and their required *level of accuracy* should be clearly established at the outset of the survey planning since this has a direct bearing on the overall sampling design and, in particular, to the type of multiple frame estimators used, the questionnaire design and data collection procedures.

It is assumed also that, in terms of the timeliness of the survey design discussed, survey data should be collected, tabulated and published within a two month period. However, in the case of cost of production surveys or other special surveys that include variables that require a complex questionnaire and data collection procedure, the period to obtain the final results can be allowed to be longer.

The multiple frame sampling methods described combine a probability sample of land areas called segments, selected from an area frame, with a complementary short list of special agricultural holdings to be completely enumerated during the survey field data collection. The multiple frame estimates combine estimates from the area sample with estimates obtained from the list of special agricultural holdings.

Area sample component of the multiple frame survey

The area sample design consists of a stratified probability sample of *segments*, with a replicated selection procedure. The area sampling frame considers the territory divided into a number of land-use strata defined by proportion of cultivated land, predominance of certain crops or other land-use characteristics. The strata and sample segments must have *identifiable physical boundaries* (roads, paths, rivers, etc.) that can be located both in the field and on the cartographic materials used for their identification (satellite images, mosaics of aerial photos and maps). For the replicated sample selection, each land-use stratum is completely subdivided, by similarity of agricultural characteristics or following a geographic distribution criteria, into areas with equal number of segments called *substrata* or *zones,* which provides a further level of stratification. In each stratum, the area sample consists of a number of independent replicates. Each replicate is formed by one segment in each substrata, with equal probability of selection within the stratum. The area sample design can also be considered as a stratified, cluster sample of *tracts*, a tract consisting of the part of a holding (or non agricultural areas) included in the segment. Area sample surveys with segments that coincide with the land of agricultural holdings or with square segments are illustrated in several case studies included in volume II.

List sample component of the multiple frame survey

The complementary list of special holdings ensures the inclusion of those holdings which make a significant contribution to the total estimate of some important survey variables. Such list of special holdings may consist, for instance, of those holdings with the largest total area, those with the largest area for a given crop, those with the largest number of livestock and poultry, those with the largest revenues, those with the largest number of agricultural workers, those corresponding to a localized production and those concentrating on highly specialized types of production: these relatively small categories may not receive appropriate representation in an area sample with the consequence that the precision of the estimates for certain variables may be significantly affected. The technical difficulties of adding a short list of special agricultural holdings (the area frame component) to an area sample design are relatively minor.

Data collection procedures for the multiple frame survey

The area sample component involves an annual (or seasonal) field data collection carried out by enumerators that complete a questionnaire for each *tract* included in each selected sample segment. The enumerators collect the data for each *tract* by personal interviews with the holder or other respondent who can provide information on the *tract.* The data collection, in addition to completion of a questionnaire, often involves *identification and measurement of agricultural areas.* For each sample segment, the enumerator use an aerial photo enlargement (or a map or scale drawing), that includes the boundaries of the segment. This is called the *segment photo.* For each *tract* within a given sample *segment* the enumerator delineates on the *segment photo* the boundaries of the *tract* and the boundaries of all *fields* included in the *tract.* The enumerator verifies the crops planted and other uses of land for each field, information provided also by the holder. During the interview, the enumerator may also use a transparent grid on the segment photo to verify, approximately, the reported area of fields. Such identified agricultural areas in each sample segment can later be *measured* in the office using a computerized measurement instrument or a planimeter. The checking of area estimates made by holders and/or enumerators, provide a very important feature concerning data reliability. The enumerators also collect information on the complementary list frame of special holdings applying a questionnaire to each such holding.

Since the technical difficulties of adding a short list of special agricultural holdings (the list frame component) to an area sample design are relatively minor, most of the manual discuss the design and implementation of an area sample survey with segments that have identifiable physical boundaries, as described. Area sample surveys with segments that coincide with the land of agricultural holdings or with square segments are illustrated in several case studies included in volume II.

The decision to base a Current Agricultural Survey on the proposed multiple frame sampling methods should take into account local conditions and resources and also alternative survey methods. For this reason, **Part I** of this volume refers to the concepts and definitions required to describe and compare the main types of survey designs used to conduct such current agricultural surveys. **Part II** describes the sample design, organization and implementation procedures of agricultural area sample surveys with segments that have identifiable physical boundaries, that is, the area sample component of the multiple frame survey. **Part III** describes the preparation of list frame component of the survey which is the list of special agricultural holdings, and the multiple frame estimators.

PART I: Current Agricultural Surveys

Chapter 1 refers to the initial considerations required for the establishment of a *current agricultural survey programme* based on probability sample methods that could provide timely and reliable basic data for the agricultural sector; and Chapter 2 describes the main types of current agricultural survey designs. The chapter also discusses the problem of choosing an appropriate current agricultural survey design and provides a number of comparisons between multiple frame and other alternative survey methods.

PART II: Area Sample Designs with Segments that have Identifiable Physical Boundaries

Chapter 3 introduces the general characteristics of the survey design; the basic concepts, definitions and procedures used in area sampling and the criteria for determination of the main parameters used for area frame construction and sample selection: for example the total sample size and frame size, the number and type of land-use strata, the minimum size of compact blocks within strata, the allocation of the sample to strata, the target size of segments and primary sampling units as well as the permitted range of actual size in each land-use stratum.

Chapter 4 presents the survey estimation methods: the formulae for the direct expansion and variance calculations for the closed, weighted and open segment estimators along with their associated strengths, weaknesses and applications. Radio estimators and their variance formulae, both for within- survey and between-surveys, are also presented in addition to their strengths, weaknesses and applications.

Chapter 5 covers in more detail the resources required for the implementation and maintenance of the area sample survey, including personnel, equipment and materials. It refers to the cartographic materials (maps, satellite images and aerial photographs) and instruments needed in order to identify, transfer and measure the area units involved in area frame construction and sample selection.

Chapter 6 deals with the area frame construction procedures, that is, land-use stratification and subdivision of the *strata* into *Primary Sampling Units* (defined as areas with physical boundaries used to determine the exact size of the area frame and to facilitate the subdivision of strata into substrata and the sample selection). It covers, also, the use of different combinations of cartographic materials (satellite images, maps and aerial photographs), area measurement and scale-transfer instruments and equipment. In particular, it includes a section on the current use of satellite images for area frame construction.

Chapter 7 deals with the area sample selection of segments, that is, with the procedures for selecting the random (or systematic) sample replicates. It includes also a point sampling procedure for selecting segments in strata that contain few easily recognizable physical boundaries. And deals with the preparation of segment photo enlargements or scale drawings to assist data collection.

Chapter 8 describes the survey preparations needed for data collection, including preparation of the survey questionnaire and other forms, timing, hiring and training of supervisors and enumerators, materials and equipment for field staff. And covers the survey field data collection procedures used by enumerators to conduct direct interviews with holders with the assistance of photo enlargements of the segment.

Chapter 9 summarizes issues concerned with the survey data processing and analysis, including questionnaire editing and review, data entry, and the preparation of preliminary tabulations of the final results.

PART III: The Multiple Frame Survey Design

Chapter 10 discusses the preparation of the complementary list frame of special agricultural holdings to be enumerated during the field data collection.

Chapter 11 presents the multiple frame estimation procedures and provides some actual examples of survey results.

Chapter 12 features a selection of topics that require special attention and knowledge in the design and implementation of a multiple frame agricultural survey. It also presents some concluding views and basic recommendations for the planning and implementation of periodic, multiple frame agricultural surveys in developing countries.

The multiple frame agricultural survey methods described represent, in several ways, an improvement on those methods based exclusively on a list sample of agricultural holdings or holders' addresses currently used in most countries as the basis for their current agricultural surveys. Multiple frame methods result in grater precision of estimates of agricultural areas, the main crop areas and other key variables of all multiple-purpose agricultural surveys, since the area sample component involves a practical procedure for the objetive measurement of agricultural areas on aerial photos. In addition, the area sample component provides the means for selecting probability samples of fields needed to conduct crop cutting yield surveys, that provide objective crop production and crop forecasting estimates.

PART I. CURRENT AGRICULTURAL SURVEYS

This first part refers to the initial considerations and definitions required for the establishment of *Current Agricultural Survey Programme,* that is, a national (or large-scale) and periodic (annual or seasonal) agricultural data collection programme based on probability sample methods to be conducted in order to obtain timely and reliable basic data for the agricultural sector.

It includes also the concepts and definitions required to describe the main types of *Current Agricultural Survey* designs, and provides comparisons between the alternative methods and in particular with the multiple frame survey design described in detail in Parts II and III of this manual.

CHAPTER 1

CURRENT AGRICULTURAL SURVEY OBJECTIVES

1.1 Existing Knowledge and Available Agricultural Data

For the establishment of a current agricultural survey programme it is indispensable first to review existing estimates and the state of knowledge of, at least, the following characteristics of the agricultural sector:

* Total production for the most important agricultural commodities; areas under crops and livestock and poultry inventories.
* Number of agricultural holdings and their geographic distribution.
* Geographic distribution of crops and livestock.
* Number and percentages of holdings producing individual commodities of interest.
* Average size of holding and size ranges.
* Average number of parcels per holding and average distances between parcels.
* Extent of holders living on or close to their holdings, or in villages.
* Average distances that holders must travel to reach their holdings.
* Number of holders grazing livestock on public or communal land without any lease or way to relate them to the land. Number of holders with no land.
* Cultivation methods- multiple cropping, mixed or associated cropping, single crop per field etc.
* Crop Calendar - crops grown, planting and harvesting dates per year.
* Weather conditions and communications.
* Marketing channels for important crops and livestock.
* Sources of agricultural data (surveys and censuses, administrative records, export records, tax lists, purchase records of controlled crops, farmers' associations', etc.).
* Administrative and political subdivisions of the country.
* Types of cartographic materials (maps, satellite images and aerial photography) that exist for the country and source from which they can be obtained.
* Agricultural zones with different proportion of cultivated land; and zones were the land is difficult to divide into small areas with stable and recognizable physical boundaries.

In addition of obtaining the best possible information and estimates of the above items, in order to plan the survey the following factors should also be taken into account:

* The holders' capacity to provide accurate data.
* The capacity of staff to organize and conduct the survey programme.
* Prioritisation of the needs of data users.

1.2 Survey Objectives and Survey Variables

Once the data requirements are identified and a decision is made that existing data are inadequate, the first step in planning an agricultural survey is to lay out the objectives of the investigation. The survey objectives should be as specific, clear-cut and unambiguous as possible and should include the required level of accuracy of the data since this has a direct bearing on the overall survey design.

In discussing the specification of objectives, questions of the type illustrated below should be explored in detail.

* What is expected from the agricultural survey?
* What is an agricultural holding? Which are the definitions of a holder and other respondents responsible to provide information on the holdings?
* Which agricultural variables should be investigated in the periodic agricultural survey and what is the required periodicity for the estimates of such variables over time?
* Which are the existing main estimates of agricultural commodities considered inadequate that should be investigated in a new agricultural survey?
* What are the required levels of accuracy for the estimates of the main crops, livestock and other main variables to be included in the agricultural survey?
* At what level will the data be summarized, i.e., country, state, region, watershed, etc.?

1.2.1 Agricultural Holdings, Holders and Respondents

In setting out the objectives of the agricultural survey, it is necessary, in particular, to define precisely some common terms used in describing the agricultural sector. For instance, the terms agricultural holding and agricultural holder must be carefully defined in such a way that they are practical to use.

An <u>Agricultural Holding</u> (or <u>Farm</u>) is 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. It includes land rented, land owned, and land being effectively used by the management under whatever type of other arrangement. Single management may be exercised by an individual, jointly by two or more individuals, or by a household, clan or tribe, or by a juridical person such as a corporation, religious organization, cooperative or government agency. The holding's land may consist of one or more separated parcels (simple compact blocks of land), located in one or more separate areas or in one or more territorial or administrative divisions, providing the parcels share the same production means utilized by the holding, such as labour, farm buildings, machinery or draught animals; cf. FAO (1995).

A <u>holding parcel</u> is any piece of land entirely surrounded by other land, water, road, forest, etc. not forming part of the holding. A parcel may consist of one or more fields adjacent to each other.

A field is a piece of land in a parcel separated from the rest of the parcel by easily recognizable demarcation lines, such as paths, fences, cadastral boundaries and/or hedges, and on which a specific variety of crop and planting date, or specific crop mixture is cultivated[1].

The following additional points relate to identifying a holding:

(a) holdings may have no significant land area, e.g. poultry hatcheries or holdings keeping livestock for which land is not an indispensable input for production;

(b) holdings consisting exclusively of the harvest of tree crops may be operated by persons who do not have any rights to agricultural use of the land on which the trees are grown.

The holder (farmer or operator) is a civil or juridical person who makes major decisions regarding resource use and exercises management control over the agricultural holding. The holder decides which crops to plant; when to plant; when to harvest, when to sell livestock, where to sell livestock and how many to sell etc. The holder has technical and economic responsibility for the holding and may undertake all responsibilities directly, or delegate responsibilities related to day-to-day work management to a hired manager; FAO (1995).

A hired manager is a civil or juridical person who takes technical and administrative responsibility to manage a holding on a holder's behalf. Responsibilities are limited to making day-to-day decisions to operate the holding, including managing and supervising hired labour. Where the hired manager shares economic and financial responsibilities in addition to managing the holding, the hired manager is usually considered a holder or a joint holder; FAO (1995).

The respondent is the person from whom data are collected about the holding or part of a holding. The respondent for an agricultural survey should be the holder or the manager.

1.2.2 Survey Area and Reporting Units of Survey Variables

The agricultural survey objective is to obtain estimates of several variables of interest for the total survey area. The estimate of each variable for the survey area is a number obtained through an inference procedure based on the values of the variable in all or a sample of reporting units, which are usually the *holdings* or land areas called *tracts*. The estimates are numerical characteristics of the population of reporting units.

The survey area of interest will be the total of the country or a given Province or State, or other primary level of the administrative or political subdivisions of the territory.

---

1 A field as defined here corresponds to a plot in FAO (1995).

For the national, periodic and multiple-purpose agricultural surveys considered, the most common survey variables are the following:

* Planted and harvested area, area intended for harvest, potential and actual crop yield of each crop or variety of crop, crop production and number of trees;
* Livestock and poultry inventories (e.g. number, type, age, sex, breed and use);
* Production of milk, eggs, honey and seeds;
* Number and types of farming methods and agricultural inputs including labour, type and quantity of seeds, fertilizers and pesticides, source of irrigation water, drainage, extent of shifting cultivation, stocks; machinery, equipment and agricultural buildings;
* Number and types of agricultural holdings (e.g. number, location, legal status, land tenure);
* Costs of production and value of sales;
* Population involved in agriculture (e.g. basic demographic characteristics of the holder, holder's household members working in the holding, hired workers on the holding, days of work, etc.).

The reporting unit of a survey variable, is the unit to which the information collected on the variable refers. Each survey variable should be well defined on the set of reporting units. The most common type of reporting unit is the *agricultural holding*, since agricultural characteristics are naturally defined for the holdings. But the holding is not the only possible reporting unit of a variable. For a given survey variable, the reporting unit may be a land area called a *tract*, as will be defined in what follows.

For agricultural surveys, it is often convenient to consider the total survey area (a Province, for instance), completely subdivided into non-overlapping land areas called segments, and that each segment is subdivided into non-overlapping *tracts*. A tract is defined as the land area of a holding inside a segment, or the land area of a segment that does not belong to any holding. In this case, the total survey area is completely subdivided into non-overlapping tracts, a tract being part of a holding or a non-holding area.

All of the above survey variables are defined in the holdings, that is, a value of the variable for each holding can be known. However, some of those survey variables can also be defined on tracts, as it is the case of the planted area of a crop, for example.

For the agricultural surveys considered, a value of a survey variable must be defined for each holding of the survey area, or it must be defined for each tract of the survey area.

Therefore, the reporting unit of a variable can be a holding or a tract of the survey area.

Some agricultural surveys consider variables with both types of reporting units.

1.2.3 Basic Types of Agricultural Surveys

As already mentioned, the agricultural surveys consider an inference procedure to estimate the value of each variable for the survey area from the values of the variable in *all* or a *sample* of reporting units (holdings or tracts) should be established. Consequently, the agricultural surveys can be classified into *censuses* or *sample surveys.*

An agricultural census is a survey in which the value of each variable for the survey area is obtained from the values of the variable in *all* reporting units, that are usually the *holdings. The primary objective of agricultural censuses is to provide a detailed classification of the agricultural structure of the country*[2].

An agricultural sample survey is an agricultural survey for which the inference procedure to estimate each survey variable for the total survey area is based on the values of the variable obtained from a *sample* of reporting units. Questionnaires are completed for each of a sample of reporting units. *An agricultural sample survey is usually conducted to measure the performance of the agricultural structure.*

An agricultural sample survey for which the inference procedure to obtain the estimates of the survey variables is based on probability sampling and estimation methods is called a probability sample survey, a term due to Deming. For a probability sample survey the statistical precision of the estimates can be established.

On the other hand, an agricultural sample survey for which the inference procedure to obtain the estimates of the survey variables is *not* based on probability sampling and estimation methods is called a nonprobability or subjective sample survey.

Therefore, the agricultural data collection programmes considered can be classified as follows:

Agricultural Censuses	
	Subjective Sample Surveys
Agricultural Sample surveys	
	Probability Sample Surveys

2 The term agricultural census is used by FAO in a broader sense to designate an agricultural survey with the above primary objective, but conducted on a complete or sample enumeration basis; cf. FAO (1995). However, for the purpose of classifying surveys in this manual the above standard definition of an agricultural census is used.

This manual will describe in detail only probability sample surveys, since these are the only surveys for which the required levels of accuracy for the estimates can be established. However, some brief considerations of censuses and nonprobability sample surveys are also provided.

1.2.3.1 Complete Enumeration Surveys (Censuses)

An *agricultural census* provides estimates for each holding and, therefore, aggregate data for the smallest administrative, political or statistical subdivisions of the country and for classifications of holdings by size or other subgroups of interest. On the other hand, by definition, an *agricultural sample survey* does not provide data from each holding and it can only provide reliable estimates for a relatively small number of political or administrative subdivisions (or other subgroups of holdings) for which it was designed to cover, usually at a significant level of aggregation.

With a large population of holdings, a complete enumeration of all holdings is an expensive and time consuming operation subject to considerable errors, including errors of omission, duplication, recording errors, data processing errors, etc. These errors derive mainly from the difficulty of properly supervising such a large operation and controlling the quality of the work. A census is always incomplete and never completely accurate, but, if efficiently conducted, it can serve also as a reference for planning and judging results of sample surveys. A sample survey often produces more accurate results than a census for a large population of holdings, because it makes possible a more careful supervision of the field work and processing of results.

An agricultural census is not a practical means of providing current (annual or seasonal) agricultural data for a large population of holdings. Even if the resources to collect the data were available, the large quantity of data to be processed usually implies that census results are released to the public two to five years after the census data collection period.

1.2.3.2 Nonprobability (Subjective) Sample Surveys

As already defined, a sample survey in which the statistical precision of the estimates cannot be calculated is called a subjective (or nonprobability) sample survey. Subjective surveys are used in cases when statistically accurate data is not required or when there are no resources for its production.

Most estimates in most countries are still based on subjective agricultural sample surveys. In fact, if all else fails, and it sometimes does, the opinions of experts (rural agricultural agents), the use of windshield surveys, administrative records, rapid rural appraisals and subjective samples of holdings or holders' addresses are widely used to provide estimates. The accuracy of such estimates depends on the knowledge and ability of the person that puts them together and there is no way to estimate the possible error that may be contained therein. Nonprobability surveys can be useful if the objective of the survey is to produce only simple ratio estimates. They are also used when good check data exists to establish a regression procedure.

Subjective surveys based on administrative registers are quite common. Estimates for some items in many countries are derived from summing area and production of crops, livestock and other items that are reported to a government agency by each holder in order to obtain permission for marketing, or to receive a subsidy, or simply because the report is required by law. This source of information should be examined closely. First of all, it may be significantly incomplete because a great part of the production may be marketed informally as contraband with only a minimum reported to maintain eligibility. Or, on the contrary, exaggerated production may be reported to obtain a larger subsidy. *There are many reasons why holders may not be inclined to report accurately on their holdings.*

It should also be mentioned that in most countries that formerly had a centralized economic system, national agricultural statistics were based on complete censuses or complete registers, and probability sample surveys were seldom utilized. But in practice, in many cases, such supposedly complete censuses and registers are more properly described as being subjective sample surveys.

1.2.3.3 Probability Sample Surveys

Agricultural sample surveys which use probability sampling methods that allow calculation of the statistical precision of the estimates are called probability sample surveys. In such surveys the statistical precision of the estimates have a precise mathematical meaning.

The planning of an agricultural probability sample survey requires the following additional specifications in order to define the probability model on which the estimates of the variables for the total survey area (usually a Province or the total of the country) from the selected sample are based:

* The survey population (or sampled population), which is the set of units to be sampled, called sampling units or enumeration units. The sampling units may be different from the reporting units of the survey variables, which are the *holdings* or the *tracts* of the survey area. The most commonly used sampling sampling units are *holdings* (or *holders addresses)* or *segments* which are land areas.

* A probability selection procedure such that each possible sample of sampling units of a given size has a non-zero probability of selection.

* Rules of association between the *sampling units* and the *reporting units*. A sampling unit must either be associated with one and only one reporting unit or there must be a known rule to associate sampling units with a group of reporting units.

* The survey variables should be defined in each sampling unit as a function of its values in the group of associated reporting units.

* The estimator for each survey variable. An estimator is a random variable: a numerical function defined for each possible sample of sampling units. The value of the estimator for the *selected sample* provides the *estimate* of

the survey variable. The estimate of each variable for the survey area is based on the values of the variable in a sample of *sampling units* associated with the *reporting units* of the variables (the holdings or the tracts of the survey area).

* The variance of the above estimators, which are also random variables, provide the precision of the survey estimates. For a given variable, the value of the variance of the corresponding random variable for the *selected sample* provides the sampling error of the estimate. The level of accuracy or desired degree of precision of the estimates should be established.

The target survey population is the population it is desired to survey: because of practical constraints, the survey population, which is the population actually sampled (the set of sampling units), and for which inferences are valid for obtaining the survey estimates, may be different from the target population.

1.2.4 Sample Design and Survey Design of Probability Sample Surveys

The **sample design** of a sample survey refers to the techniques for selecting a probability sample and the methods to obtain the estimates of the survey variables from the selected sample.

Sample designs involve, therefore, different types of sampling units, rules to assign probabilities of selection to sampling units, sampling fractions, possible stratification and clustering procedures and different types of estimation methods. Some designs involve several sampling selection stages in which case, for each sampling stage, the sampling units, probabilities of selection and sampling method have to be established to obtain the final survey estimates.

For a given sampling stage of a probability sampling design, a sampling frame is the *total set of sampling units with their probabilities of selection*, that is, the list of sampling units from which the sample is selected together with their probabilities of selection. A frame is needed and has to be constructed for each sampling selection stage and a non-zero probability of selection has to be assigned to each sampling unit of the frame.

A few of the most common terms used for the description of agricultural probability sample designs are listed below for easy reference.

In stratified sampling, the survey population is subdivided into non-overlapping sets called strata. Each stratum is treated as a separate population.

Primary Sampling Unit (PSU) is the term used for designating the sampling unit at the first stage of selection in multiple-stage sampling.

Cluster sampling is the term used for sampling plans in which the sampling units are groups (clusters) of population units.

Equal Probability Selection Method (EPSEM) refers to a sample selection in which every sampling unit has the same probability of being selected for the sample.

Sample Selection with Probability Proportional to Size Measure (PPS) is a sampling procedure in which the probability of selection of a sampling unit is proportional to its assigned size called measure of size.

In replicated sampling the total sample is made up of a set of replicate subsamples, called replicates, with the identical sample design.

The Coefficient of Variation (CV) of an estimate is obtained by dividing the standard error of the estimate by the estimate itself, expressed as a percentage. It is an indication of the precision of a sample estimate.

The Design Effect (Deff) of an estimator, a term due to Kish, is the ratio of the variance of the estimator for the given sample design to the variance of a corresponding simple random sample of the same size.

The Mean Square Error (MSE) of an estimate is equal to the variance of the estimate plus the square of the bias inherent in the survey procedures.

The difference between a sample result and the result from a complete count taken under the same conditions is measured by the precision or the reliability of the sample result.

The difference between the sample result and the true value is called the accuracy of the sample survey.

It is the *accuracy* of the survey that is most important; but it is the *precision* that is possible to measure in most instances.

Coverage error refers to omission and duplication of reporting units, including incorrect determination of the land area corresponding to the reporting units.

The overall **survey design** of a probability sample survey refers to the definitions and the established methods and procedures concerning all phases needed for conducting the survey: the sample design, the selection and training of personnel, the logistics involved in the management of the field force and the distribution and receipt of survey questionnaires and forms, and the procedures for data collection, processing and analysis.

For the general theory of survey sampling the reader can refer to the classical books by Hansen, Hurwitz and Madow (1953), Deming (1960), Kish (1965) and Cochran (1977).

CHAPTER 2

ALTERNATIVE CURRENT AGRICULTURAL SURVEY DESIGNS

2.1 Sampling Frames for Agricultural Probability Sample Surveys

There are two basic types of sampling frames used for the *last-stage* of selection of an agricultural probability sample survey: area frames and list frames.

The sampling units of an area frame are *land areas* called *segments,* and the selection probability of each segment is *proportional to its area measure*. The sampling units of a list frame are usually the holdings or holders addresses.

Therefore, there are two basic types of sample designs in terms of the *last-stage sampling unit* and the *rules to assign their probabilities of selection*, namely area sample designs and list sample designs, also known as area frame sample surveys and list frame sample surveys.

Multiple frame surveys are those agricultural probability sample designs that combine more than one sample design to obtain the survey estimates, combining area frame designs with list frame designs. A multiple frame survey usually consists of an area sample component and a list sample component.

Consequently, agricultural probability sample surveys can be classified as follows:

	Area sample surveys
Agricultural Probability Sample Surveys	List sample surveys
	Multiple frame surveys

For the current agricultural probability sample surveys considered, it will be assumed that most data will be obtained by enumerators through personal interviews in the field. Indeed, other means of periodic data collection, for example the use of self-administered questionnaires, mail, telephone inquiries, or extraction of information from registers are usually not feasible in developing countries.

The field data collection of an agricultural survey may include or not, in the selected sampling units, expert observation or identification and objective measurement of crop areas or other agricultural characteristics of special interest.

2.2 List Frame Sample Surveys

List sample designs are the most commonly used sampling procedures for agricultural surveys.

As defined above, a probability sample that is not an area sample is called a list sample. The usual reporting unit is the *holding* and the commonly used last-stage sampling unit is the *holding or the holder address.*

A list frame is a frame used for the last selection stage of a list sample design. Therefore, list frames are often formed by holdings or holders addresses.

Frame Construction and Selection Procedures of List Sample Designs

In many countries, country-wide surveys are conducted using samples from large list frames of holdings or holders addresses. These list sampling frames come from previous agricultural, housing or population censuses, from an accumulation of lists prepared by political or administrative subdivisions, from farmer's associations, land tax records, cadasters or from other sources. There is usually information on holdings size, crops grown, livestock held and other characteristics for each holding. Some of this information allows to stratify the frames, which greatly improves sampling efficiency.

Large list frames are often incomplete or inaccurate. They contain an unknown amount of duplication, and their accuracy degenerates rapidly over time. If the list is a few years old, many of the names will no longer represent a holding due to sales, deaths, and abandonment, and new holdings will not be represented. It is also common in list building to unintentionally include more than one name that could be associated with a specific holding. Joint holders are a problem in list building. In short, it can be seen that it is difficult for a list of holding addresses to fulfil the requirements for an accurate probability sampling frame, particularly if it is a large list. To be truly effective, a large list of holdings must be continuously and systematically updated which is expensive, time-consuming and requires a large staff. When the list frame is inaccurate the sample cannot be assumed to be a probability sample and should be considered a subjective survey. This is the case in many countries that use list sample surveys.

Some European countries are able to utilize large, country-wide list frames effectively because of administrative procedures that result in the frequent registration of all holdings. These countries have the advantage of being able to use mail and telephone for data gathering from the list. Otherwise, holdings selected at random, or even systematically, from such a list will be widely spaced over the countryside, greatly adding to the cost of field enumeration.

List sample designs often include at least one stratum of special holdings that is completely enumerated, or sampled with a high sampling fraction. Such special holdings are defined as those which contribute to a significant proportion of the total estimated value of important survey variables. Examples are the *largest holdings, holdings with the largest area for a given crop or with the largest livestock herd, highly specialized holdings or those corresponding to a localized production.* Such list of special holdings is fairly easy to update since the holdings involved are usually visible and well known. Information can be obtained from extension agents, producers' associations, banks, tax records, agricultural censuses and from government agencies that control and/or purchase production of certain crops and other agricultural commodities. The preparation of such list of special holdings should include the accumulation of data on each holding such as holding size, crops grown, type of livestock held and inventory for stratification purposes if the list is to be sampled.

In most countries, a reliable country-wide list frame of holdings or holders addresses does not exist, and so a rough approximation is used on which to base a multiple-stage list sample survey. In such common cases, a sample of holders is selected by first selecting a sample of villages (clusters) with probability proportional to their total population (or housing units) since such information is usually available in most countries and approximates the number of holders. Some additional information about the villages such as farm population at some point in time and primary agricultural activity of the holdings may be available allowing for some rudimentary stratification. Other small administrative subdivisions such as districts or subdivisions of census enumeration areas (EAs) from the latest population census or from the latest agricultural census can be used instead of villages as Primary Sampling Units (PSUs) which is a common form of cluster sampling. All holdings are listed within the selected PSUs and a sample of holdings (as represented by holders) is chosen in the second and final stage.

Data Collection Procedures for List Sample Designs

During data collection, usually the enumerator completes a questionnaire for each selected holding by conducting an interview with the holder. Often, the enumerators also measure the area of the holding and its fields for a subsample of holdings. The objective measurement of agricultural areas is often required since in many countries such basic data is not known (or an accurate answer cannot be obtained) from the holders.

The above types of list sample designs are those frequently used in developing countries.

2.3 Area Frame Sample Surveys

An area sample survey is a probability sample survey in which the final stage sampling units are land areas called *segments*, and the selection probabilities are proportional to their area measures.

The measures of size used to select the segments (sampling units) are defined as a function of their area measurements. The usual type of measure of size is the total area.

The *segments* should not overlap and must cover the entire survey area. The term segment is also used to refer to the piece of land associated with the sampling unit or to the group of reporting units associated with the piece of land.

Segments Considered for Area Sample Designs of Agricultural Surveys

> *Types of Segments Used for Agricultural Area Sample Surveys*
>
> I. *segments with identifiable physical boundaries,* terrain features such as roads, rivers, canals, railroads, etc. that are readily found and provide an unambiguous identification of the segment; or
> II. *square segments*, that is, segments defined by straight lines forming squares whose end points are established by map coordinates. In this case, grid sampling procedures are used.
> III. *segments that coincide with the land of agricultural holdings*. In this case, point sampling procedures are used.

The land of each segment is subdivided into non-overlapping tracts. A tract is the part of a holding which lies within the boundaries of the segment, or the land of the segment not part of any agricultural holding. A tract is determined by the boundaries of a *segment* and by the *holdings* with land in the segment. The definition of tract is not necessary for area sample designs in which the segments coincide with the land of agricultural holdings (type III above).

A tract within a segment is often divided into a number of fields which have recognizable boundaries and in which the land use is different.

A holding is composed of one or more tracts. All land under the operating arrangement, non-agricultural land, farmstead, barns, corrals, pasture, ponds, etc., are included in the tract. A questionnaire should be completed for each tract, except for those tracts that consist of wasteland, water, and non-holding land, that will merely be listed on a control sheet.

Stratification of Area Sample Designs

Most area sample surveys consider a subdivision of the frame into land-use strata. Stratum boundaries must consist of physical terrain features (roads, paths, rivers, etc.) that can be located in the ground.

The land-use strata are defined by proportion of cultivated land, predominance of certain crops, special agricultural practices, average size of cultivated fields, agro-urban areas, or other land-use characteristics. Unless otherwise stated, when referring to an area sample the word *strata* will be used to denote *land-use strata.*

Most area sample designs consider replicated sampling selection methods within substrata (*zones*) of the land-use strata, each with an equal number of segments. This provides a further stratification which is applied in order to improve the efficiency of the design.

2.3.1 Area Sample Designs with Segments that have Identifiable Permanent Physical Boundaries (Type I)

Sample Selection Methods

A sample design of this classical type considers segments of nearly equal size in each land-use stratum. The area sample design usually consists of a stratified two-stage sample design using a random replicated sampling selection procedure. Stratified single-stage designs using a systematic replicated sampling selection procedure have also been used. Three sampling stages are used also for special strata in the above designs.

Each stratum is completely subdivided into non-overlapping Primary Sampling Units (PSUs) -Counting Units or Frame Units-, which are areas with recognizable physical boundaries formed by segments. The measure of size assigned to each PSU is equal to the number of segments it contains. Within each stratum, PSUs are ordered by similarity and then selected with probability proportional to size measure (PPS sampling). Then, within a selected PSU, segments are selected with equal probability (EPSEM).

The PSUs defined in such a way can be formed quickly and provide a means of identifying and *counting* all segments in each stratum and in the total frame without actually mapping each segment.

Once the PSUs are defined and their areas are measured then the exact size (number of target segments) of each stratum and of the total frame will be known. The PSUs are ordered by increasing order of similarity of agricultural characteristics, ordering that will ensure also a geographic distribution of the sample. In each stratum the PSUs facilitate the location of the sample of segments selected using a systematic or random selection of segments with equal probabilities (EPSEM). For identifying a selected segment it is only necessary to partition the corresponding PSU into a number of segments equal to its assigned number of segments.

For one-stage area sample designs, the PSUs as defined above are also known as counting units, and are also constructed to calculate the size of each stratum and of the total frame, introduce a further stratification given by their order, avoid mapping each segment, and facilitate the probability sampling of segments. For simplicity on notation, in any case CU's or frame units will be called PSUs.

In area sampling, for practical purposes, the ordered list of PSUs within each stratum, with their assigned measures of size -which are the number of segments in each PSU- is called the area sampling frame, even for one-stage area sample designs. As mentioned, the frame of PSUs allows to select the probability sample without actually listing all segments.

Preparing an area frame and selecting a sample of segments for this classical type of area sample designs with segments that have identifiable physical boundaries is a demanding job but some of the work can be effectively done by applying the procedures to be described in this manual.

Estimation Methods

In an area sample design of this type, it is indispensable to distinguish between the sampling unit (segment), and the reporting unit (holding or tract) for each variable in order to define the corresponding estimator.

Since a given survey variable is defined in the reporting units (holdings or tracts), and the sampling units are the segments, in order to define the estimator, rules should be established to define a value of the variable in each segment as a function of its values in a group of the associated holdings or tracts.

Data Collection Procedures

As described in the Introduction, the field data collection for each *tract* of a segment is carried out by enumerators that complete a questionnaire through personal interviews with the agricultural holder. The data collection often involves also *identification and measurement of agricultural areas* undertaken with the help of an aerial photo enlargement (or a map or scale drawing) of the segment that is used to measure the identified agricultural areas.

2.3.2 Area Sample Designs with Square Segments (Type II)

Sample Selection and Estimation Methods

The usual sample design considered is a stratified sample of square segments selected with equal probability within each stratum. Square segments of equal area (the cells of a grid overlaid on the strata) are selected within each stratum.

As discussed in the description of the area samples of the Czech Republic and Spain, provided in volume II, the area frame construction and sample selection is basically simpler than the area frame construction and sample selection for an area sample design with segments that have recognizable physical boundaries.

The estimation methods for an area sample design with square segments are basically analogous to ones described above for an area sample design with segments that have identifiable physical boundaries

Data Collection Procedures

The field data collection for an area sample design with square segments is analogous to the data collection described above of an area sample design with segments that have permanent physical boundaries (Type I). However, reliable data collection for multiple-purpose area sample design with square segments is more difficult (and may be even impossible in practice) because of the difficulty of obtaining from holders reliable data *for the tracts* of a segment that cannot be observed on the ground.

2.3.3 Area Sample Designs with Segments that coincide with the Land of Agricultural Holdings (Type III)- Point Sampling of Holdings

Sample Selection

The sample design can be considered a stratified sample of holdings selected with probability proportional to their areas. In this case, a grid is overlaid on the strata and a sample of points is selected. Then, the selected points are identified on the ground and the corresponding holding is selected for the area sample. The area frame construction and sample selection is simpler than the area frame construction and sample selection for an area sample design with segments that have recognizable physical boundaries.

Estimation Methods

The estimation methods for this type of area sample designs are simpler than those needed for the other two types of area sample designs. In fact, a one-to-one correspondence between reporting units (the holdings) and last-stage sampling units (the land of the holdings) is considered.

Data Collection Procedures

If objective measurement of areas is required (as is often the case in many countries), it can be accomplished by the enumerators by measuring the area of the holding and its fields during data collection, for a subsample of holdings -as it is undertaken for list sample surveys-. This procedure is more cumbersome and less precise than the area measurements undertaken in the office with the help of aerial photos, as commonly used with the other types of area sample designs.

A data collection procedure for the area sample design of Nicaragua, which does not involve objective measurement of areas, is described in volume II.

2.3.4 Development of Area Sampling Designs and Use of Geographic Information Systems (GIS)

A distinguishing characteristic of area sample survey methods is that they have incorporated important technological advances in computer data processing to a larger extent than list sample methods. In fact, area sample methods can utilize satellite imagery or even digital satellite data as part of Geographic Information Systems (GIS), hand-held Geographic Positioning Systems (GPS) and generally a variety of automated procedures and techniques for sample selection and data analysis.

Geographic Information Systems are used in many countries for a variety of purposes related with agricultural statistics. A GIS called CASS (Computer-Aided Stratification and Sampling) is currently used in the United States for area frame construction and sample selection. These procedures are highly automated and the material requests, mainly mapping requests are difficult to meet in most countries. The methods are important because they are adopted at present in the biggest and oldest area frame project in the world and because they may represent the trend of the evolution of area frame construction and sample selection for manual to computerized procedures. In other words, these computerized methods may be partially or totally adapted to different conditions and requirements in other countries. As already mentioned, this manual does not present the use of GIS for area frame construction and sample selection, describing instead methods that require a minimum of resources and specialized staff taking into account the constraints faced by developing countries. However, in order to illustrate the use of GIS for area frame construction and sample selection, summaries of the area sample designs of the Czech Republic, Spain and the United States will be included in volume II.

National or large-scale agricultural area sample surveys have being applied and developed for many years in a large number of countries involving a diversity of conditions and requirements. The first nation-wide area sample was implemented some fifty years ago; cf. King and Jessen (1945) and since then, most important advances in procedures have taken place in connection with the development of the periodic national area frame agricultural survey in the United States and other countries.

General, classical presentations of area sampling methods as originated and used in the United States can be found in Houseman (1975), Vogel (1986), Fecso, Tortora and Vogel (1986), Cotter and Nealon (1987) and Cotter and Tomczak (1994).

2.4 Multiple Frame Sample Surveys

An agricultural multiple frame sample survey is a probability sample survey that combines the two basic types of probability sampling methods, an area sample component with a list sample component.

Multiple frame estimators combine area sample with list sample estimators for each survey variable. If the list frame is sampled, the results obtained from the selected holdings are expanded and added to the area sample estimate. If the list frame is completely enumerated, the results obtained from the list are added to the area sample estimate with no additional contribution to the overall variance.

For multiple frame surveys, all holdings in the list frame must be removed from the area frame, an operation that requires special attention and resources. In other words, all tracts in the selected sample of segments, that correspond to holdings in the list frame, should not be considered to obtain the area sample estimates.

A list frame of special holdings is a necessary addition to an area sample in order to provide adequate estimates for important agricultural variables that have a highly skewed frequency distribution. As it is known, a number of important agricultural variables concentrate a significant proportion of the total estimate in a small proportion of the holdings. For each of these variables, the list sample should account for the skewness of its distribution. As a result, the corresponding multiple frame estimates will be more precise than the area sample estimates.

The list of special holdings, that should be updated annually prior to the survey field data collection, can be formed (as already mentioned for the case of list sample designs) by those holdings with the largest area for a given crop, those with the largest number of livestock, or those that may not be duly represented in an area sample. The annual or seasonal data collection for the list of special holdings can be done by direct interview with holders or managers, or by leaving the questionnaires with the holder and collecting it afterwards, or data collection may be even done by using the mail if the list sample is relatively short and therefore easy to control.

The list frame of a multiple frame survey can be a *large*, nationwide list of holdings. The preparation and updating of such frame requires a heavy investment in computer hardware and software and a very controlled field operation for its use in combination with the area sample. In the USA and Canada, for example, the multiple frame survey designs combine a large, nationwide list sample with an area sample. However, such type of multiple frame design, although the most efficient, is not feasible in most developing countries.

On the other hand, the most practical multiple frame methods for developing countries are those with a relatively *short* list of special holdings, to be completely enumerated, used as a complement to the area sample. This is the case, for example, of the multiple frame survey design described in the following chapters of this manual.

2.5 Comparison of Alternative Agricultural Survey Designs

The choice of an appropriate current agricultural survey design can best be made by considering the comparative advantages, disadvantages and requirements among *multiple frame survey designs*, *area sample designs* and *list sample designs*.

A distinction is made between surveys that require or not objective measurement of agricultural areas. It should be mentioned that in many countries or large areas of developing countries the areas reported by farmers are considered not reliable and therefore objective measurements of areas are required to obtain reliable estimates. This need arises partly on account of all kinds of arbitrary local units of measurements that are in use in different parts of the same country, and partly because of the general tendency among farmers to under report their areas and production.

It will be assumed that the list sample design is of the common type already described in section 2.2.

Multiple Frame Designs vis-a-vis Area Sample Designs

* A multiple frame design that combines an area sample with a relatively short list of special holdings which make a significant contribution to the total estimate of some important survey variables is preferable to the area sample design alone, since it can provide more accurate estimates of such variables, and because the extra work involved for its design and implementation will generally not be significant.

Multiple Frame Designs vis-a vis List Sample Designs

When referring to an *area sample* it will be assumed here to be the area sample component of a multiple frame design that includes also a list frame of special holdings; and also that, unless otherwise stated, the segments have identifiable physical boundaries.

Some of the advantages, disadvantages and requirements of such multiple frame designs of current agricultural survey programmes as compared with list sample designs are the following:

Advantages:

* Coverage Errors Related to Complete Coverage. The area frame provides *complete coverage* of the population of reporting units (holdings or tracts). The target population coincides with the population of interest. Coverage errors are a major problem in list sampling, but not in area sampling provided the rules of association linking reporting units with selected segments are performed correctly. Therefore, probability area sample estimates are not biased concerning coverage errors. This cannot be achieved with a list sample survey since, in practice, a complete list of holdings, valid as of the date of the survey data collection period, cannot be established.

* Coverage Errors Related to Repeated Use of the Frame. If an annual or seasonal agricultural survey is to be implemented, it is worth noticing that an area frame is generally far more durable than a list frame of holdings. An area frame does not become outdated for many years (say 5 to 15 years) unless the population extends into areas not covered by the frame. Arrangements must be made to keep the frame up-to-date in areas where the proportion of cultivated land (the stratification criteria used) is changing rapidly as in areas of new urbanization, or to cover new agricultural areas. Changes in land use, or in the number and location of holdings may reduce the precision of the area sample estimates but they do not introduce bias.

* Precision of the Estimates. An area sample design with segments that have recognizable physical boundaries or one with square segments, obtain *more precise estimates of agricultural areas* (a key variable studied in all agricultural surveys) than a list sample. In fact, by definition, in area sampling the probabilities of selection and the sampling expansion factors are proportional to agricultural areas.

* Non-sampling Errors and Objective Measurement of Areas. If objective measurement of agricultural areas is required, for the area sample design non-sampling errors associated with area measurements are reduced by the use of aerial photographs of the selected segments that clearly indicate the holdings and the fields and is used to check reported area of fields and total area of the holding itself. The main crop areas are identified and delineated on aerial photographs or detailed maps during the field data collection and then *measured* in the office. The holder is more inclined to be truthful when confronted with questions about specific portions of his holding that are also being observed by

the enumerator at that moment. The area sample design involves a more convenient and accurate procedure for the objective measurements of agricultural areas, which is an important advantage for providing accurate area estimates. For list sample designs the measurement of areas is usually undertaken only in a subsample of holdings during data collection. This procedure is generally slow and cumbersome to apply, and impractical when dealing with holdings formed by parcels a long distance from one another.

* Basis for a Crop Cutting Yield Survey. An area sample design with segments that have recognizable physical boundaries or one with square segments, provides the means for selecting a statistical sample of fields needed to conduct *crop cutting yield surveys* in order to estimate crop production when crops mature and make crop forecasting yield estimates by measuring plant characteristics at certain stages of growth during the crop year. Such a sample of fields, obtained from the area sample, can also be used by trained enumerators to make eye-estimates of crop yields. In developing countries, holders are often not able to report reliable estimates of crop yield and production and there is the problem of local measures which often vary from village to village or even from holder to holder. Crop-cutting methods are recommended to collect objective estimates of yield although they are time consuming, expensive and require strict training and supervision of enumerators for their adequate implementation.

* Sample Size. When comparing the sample size of a multiple frame design with a list sample design a larger sample may be necessary due to the between cluster and within cluster variances. The probabilities of selection of the PSUs, and of the initial stages of the sampling design, cannot often be properly established for an efficient sample design.

* Boundaries of PSUs. For a list sample design, it is often not easy (sometimes impossible) to establish the boundaries of the PSUs whether they are villages, enumeration areas or administrative subdivisions.

* Changing System of Holdings. The area frame construction is independent of agricultural holdings. This makes area sampling an appropriate solution for example, for countries or areas where the system of holdings and other agricultural infrastructure has been altered drastically and will be in a state of flux well into the future, and for which the knowledge of agricultural production is more important than obtaining results by categories of holdings (e.g. countries of Central and Eastern Europe).

* Data Collection Costs. Area sample surveys are cost effective on a per holding basis since each segment contains a group or cluster of *tracts* (reporting units). If a list sample is not a cluster sample of holdings, the holdings to be surveyed by an enumerator could be separated by a significant distance consequently increasing survey costs.

Disadvantages and Requirements

* Cartographic Requirements. Constructing an area frame requires cartographic material upon which accurate identification of areas and land measurements can be made. Accurate maps and preferably satellite images or a mosaic of aerial photography, as well as instruments for area measurement and for transferring maps to different scales (planimeters and preferably computerized equipment) are needed for preparing a stratified area frame and select an area sample. Aerial photographic enlargements, or accurate drawings of a known scale for each selected segment are a great advantage if objective measurement of areas is required. In this case, a tremendous advantage, but not a requirement, is the ability to take aerial photographs for selected land areas. The increasing availability in recent years of satellite images, computerized area measurement and scale-transfer instruments and Geographic Information Systems greatly facilitate the application of area sample methods. In addition, hand held Global Positioning Systems (GPS) provide a practical tool to locate and/or define segments. In particular, the availability of satellite images has effectively removed one of the previous problems for area frame construction which was lack of aerial photography and maps for stratification purposes. Such a variety of cartographic materials and instruments are not required for list sample surveys.

* Lack of permanent physical boundaries. For an area sample design with segments that have recognizable physical boundaries or with square segments, the lack of permanent boundaries in the maps, satellite images and aerial photos constitute a serious problem. In tropical areas such as West Africa, for example, because of the climatic conditions and shifting cultivation systems, boundaries change more frequently or get covered by bush and are not visible on the cartographic materials.

* Proximity of the Holder or Respondent to the Holding. It may not be feasible or even possible to use an area sample in some countries due to difficult terrain, lack of funding or due to certain social mores of the rural population. In a number of countries, or large regions of countries, holders live in villages often some distance from their holding. In this case, a survey based on an area sample of segments is difficult to implement. For a given selected segment far from a village, it will be problematic to identify and interview the holders with holdings partially or totally included in the segment. And even when they can be located, the holders will tend to report for their entire holding and not for the part of the holding included in the segment. For these situations, surveys without objective measurements of agricultural areas are used. If an area sample design with segments that have recognizable physical boundaries, or one with square segments is used, then the stratification can be modified to place emphasis on the villages, and the area of the segment is not taken into account except to define the holders who reside within the segment (open segment estimation procedure).

* Sample Selection Costs. There are higher costs involved *at the start* of the survey programme, for selection of an area sample of segments with recognizable physical boundaries, than those needed for a list sample. However,

for a current agricultural survey programme, a list frame of holdings needs to be updated frequently in order to control a proper covering and obtain reliable estimates; and therefore the resources needed for an alternative area sample may be easily justified.

* Technical Expertise. The implementation of a sustainable agricultural survey programme based on area sampling methods requires a dedicated, highly qualified office staff that is willing to accept the tedium and precise attention to detail indispensable in the construction and maintenance of the area frame. It requires highly trained statisticians for data analysis, for the interpretation of results and corresponding necessary adjustments and refinement of survey procedures. The frame and the sample will be no more durable than the staff. This type of staff is not easy to hire and keep motivated in civil service conditions and low agricultural budgets of a developing country. An area frame can be built and an area sample can be selected with technical assistance from various sources, but its continued existence and usefulness will depend on strong government support. The implementation of a an area sample design require more technical expertise than the implementation of a list sample design. Where it is possible to construct an accurate list frame (a list frame is often incomplete, biased and outdated), the advantage of a list sample design over an area sample design concerning simplicity of implementation is partly due to the fact that the estimation methods are simpler, since there is usually a one-to-one correspondence between sampling and reporting units. However, for an area sample with segments that coincide with the area of the holdings, the technical expertise required is approximately similar and there is also a one-to-one correspondence between sampling and reporting units.

PART II: AREA SAMPLE DESIGNS WITH SEGMENTS THAT HAVE IDENTIFIABLE PHYSICAL BOUNDARIES

This Part II describes the sample design, organization and implementation procedures of agricultural area sample surveys with segments that have identifiable physical boundaries. The field data collection, undertaken by enumerators through direct interview with the holders, is assisted with aerial photos of the selected segments used to identify and measure agricultural areas, which constitutes a very important feature for data reliability.

The subject is introduced in a straightforward and practical manner, and describe area frame construction and sample selection methods that require a minimum of resources and specialized staff taking into account the constraints faced by developing countries.

CHAPTER 3

GENERAL CHARACTERISTICS OF THE SURVEY DESIGN

This chapter provides the concepts, definitions and procedures used for the design of an agricultural area sample survey with the type considered. It includes the criteria for determination of the main parameters used, for example, it refers to the total sample size and frame size, the number and type of land-use strata, the minimum size of compact blocks within strata, the allocation of the sample to strata, the target size of segments and PSUs as well as their allowed ranges of actual size.

The terms and definitions used frequently throughout the text concerning the design of agricultural area sample surveys are included here. A few of the terms that were used earlier to invite curiosity are explained here if it was not done adequately before.

A summary description of area frame construction and sample selection will now be given in order to facilitate the presentation of the above topics.

3.1 Overview of the Area Sample Survey Design

To prepare an area frame, the first requirement is the availability of up-to-date cartographic material (maps, satellite images, aerial photos) covering *all* the land to be included in the frame. The resolution or detail of the material must be sufficient to allow a stratification according to proportion of land cultivated, predominance of certain crops or other uses of land, etc., and the subsequent subdivision of these land-use *strata* into areas called *primary sampling units* (PSUs) which must have also recognizable permanent physical boundaries. Six to ten land-use strata is the normal range constructed by using satellite images (if available) aerial photographs and maps. PSUs are usually constructed on photography and or satellite images that show the boundaries of the strata. They are transferred to maps for measurement. Each PSU must be measured and assigned a target number of *segments*. The number of segments assigned to each PSU are summed to provide the total number of segments in each stratum and summed again to provide the total in the frame. Then a two-stage probability sample of segments is selected from each stratum using a replicated selection procedure. Each sample segment is constructed on small mosaics of aerial photography on which the boundary of the corresponding PSU have been transferred. Finally, the selected sample segments are located on aerial photo enlargements used to control field data collection.

Area sample surveys usually involve a single-stage or a two or three-stage sampling design.

For a one-stage or multiple-stage sampling design, the Primary Sampling Units (PSUs) are also called Counting Units (CU's) or Frame Units.

3.2 Criteria for Area Frame Construction and Determination of the Main Parameters of the Survey Design

The area sample frame considered corresponds to a Province of the country, or to a geographic area of interest for which reliable survey estimates are required.

3.2.1 Frame Boundaries

The area sample of a country generally consists of a set of independent area samples of the provinces (or states), the first level of the political subdivision of the country. However, in the United States frames are constructed county by county, which is the second level of political subdivision, although estimates are produced at the state level. This improves the work in countries where there are accurate maps and enough information to accomplish the delimitation at that level. A frame constructed at the county or department level can more easily provide special frames and is more easily adaptable to changes in regionalization. Such a frame can also be coordinated more precisely with census data. This level of delimitation has also been accomplished in Brazil and few other countries (cf. Volume II).

In other countries, however, independent area frames are constructed for each province, and this is the case considered in this manual. It is important in this regard that provincial boundaries are established on physical terrain features. The first lines to be drawn on the photography or images are those showing provincial boundaries and the *frontier of the country*. National boundaries must be placed on the images or photography exactly as shown on officially accepted maps, arbitrary straight lines and all. Internal provincial boundaries should be varied slightly, if necessary, in order to follow physical terrain features. Estimates must be published for the official province.

An area frame can be constructed for almost any size area and special frames will be discussed later. It should be recalled that a small area will probably require a similar size of sample for a certain level of precision as that required for estimates with the same CV at the regional level. For a periodic (annual or seasonal) survey, the smallest practical level of estimation is probably the province or department due to the extra expense required to collect data for smaller political units. Departments or provinces can be combined to form regions if the survey budget is severely restricted. It can be a problem if regions consist of parts of provinces. If this is the case and the region is the level for which estimates are required, strata can still be formed within provinces but respecting regional boundaries which must also be drawn on the frame materials. Any changes anticipated in the country's regions should be ascertained before drawing in regional borders. A change in the regions after the frame is constructed will mean considerable work to fit the frame and sample to the new order.

3.2.2 Land-Use Strata Definitions

The criteria for stratification should be directly associated with the information required. Stratification in a multiple-purpose sample survey should reduce the sampling variance thereby improving the precision of the most important estimates. Stratification usually produces a satisfactory level of sampling variance with a smaller sample than would be necessary if there was no stratification. In grouping similar units into a stratum the aim is to have the lowest possible variation between units within a stratum and the highest possible variation between strata.

Stratification also enables a concentration of resources. Land that is not likely to have much or any agriculture, such as deserts, virgin forests, national parks, military reservations, and central urban areas, can be separated from the areas of primary interest and sampled at a low rate, or not at all. Meanwhile, survey efforts can be concentrated on the portion of the frame area that contains the cultivated land and pasture. Stratification can also be used in cases where different segment sizes need to be used.

The optimal stratification would usually be different for each variable to be investigated. In practice, for a large-scale, multiple-purpose agricultural survey, an optimal stratification for all variables is difficult to achieve. Thus, the main stratification criteria for the cultivated land and pastures of an area frame is based on a single characteristic, namely proportion of land cultivated; a characteristic directly related to the majority of the survey variables to be estimated. Additional strata, usually substrata, are formed by using special site or crop specific information.

Therefore, the most commonly used characteristic for defining the *strata* is proportion of land cultivated. All cultivated land planted to annual or temporal crops and land under permanent crops such as cacao, citrus, apples, grapes, coffee, etc. is classified as being used intensively. Areas where crop fields are seen to be intermingled with pasture, are classed as less intensive. Areas where there is only an occasional crop field are classed as extensive agriculture. Pasture land or range land will be the most extensive land use category. Non agricultural land consisting of deserts, heavy forests, national parks, military reserves, central areas of large cities, etc., will be in non-agricultural strata. An agro-urban stratum is often formed around cities where there is a high population density but with some cultivated fields. Bodies of water such as lakes, lagoons and large rivers would be delineated as a separate water stratum. To include a river in the water stratum, the usual definition is that it must be at least 1.5 kilometres wide. Upper areas of the river that are narrower are included with whatever stratum it flows through. In marking off water, it is useful to take into account land that will be covered by water when large dams are built in the future. This land becomes a substratum sampled for agriculture until the land is actually covered. At that time, the area is added into the water stratum and no longer sampled.

The area frame stratum definitions used in Honduras (cf. Volume II) are shown in the following example:

Stratum Number	Stratum Definition
1	61% to 100% cultivated
2	31% to 60% cultivated
3	0% to 30% cultivated
4	Mostly forest land
5	Non-agricultural land
6.1........................	Urban Areas
6.2........................	Agro-urban areas
7	Permanent water
8	Projected water

Photo-mosaic or mosaic is a term used to describe a composite of contact prints of the same scale laid together with a degree of overlap, like tiles on a roof, to form a "picture" of a large area, often equivalent to the area covered by a specific 1:50,000 scale topographic map. Mosaics are constructed in order to visualize an area of a size appropriate for stratification and construction of PSUs. If appropriate satellite images are available they can substitute with great advantage for a photo-mosaic. A proper stratification requires, in particular, a synoptic view of the total area. The area covered by one contact print, even one at a scale of 1:50,000, is insufficient for making the decisions required for stratification.

A stratum block is a compact portion or "block" of a stratum that is demarcated within or alongside another stratum.

Photography and images of the area to be stratified should be studied carefully before defining strata. The group that will prepare the stratification can learn how the different strata and physical features that will be used for boundaries appear on the photography or image by taking them to the field for visual verification. Note should be taken of natural differences between areas that are easily detected, and observe their characteristics. It is likely that these observed differences in land use can form the strata to be delimited in the construction of the area frame. A flight in a light plane over the area to be stratified by the statistician and the construction crew supervisor is a worthwhile investment. Flying can also be used to clarify questions about areas that are difficult to reach by ground travel.

The point of diminishing returns in forming strata is reached when about seven to ten major land-use strata are defined. Well formed strata reduce sampling variance, but the point is reached when the gain from adding another can be less than the problem of delineating and using an additional stratum.

A color is assigned to each stratum and the color is used to delineate that stratum on all the frame materials i.e. maps, photography, and images.

Each stratum will be sampled independently. The water stratum will not be sampled and the non-agricultural strata will usually not be sampled unless there is some doubt about what is really there, or if there is a requirement to gather environmental or land use data.

Additional land-use strata can be formed when it is necessary to take into account a special crop or some area that is important. Examples are: (a) a truck farming region that supplies a significant portion of the country's vegetables on a continuous basis, thus requiring a specialized and frequent survey; (b) zones of production of important permanent crops such as citrus, coffee and cacao. Substrata are effective only where the crop or class of crops are truly localized. It might be thought desirable to form a substratum for coffee, but if coffee is found in scattered small fields throughout the intensive stratum, a lot of work might be undertaken only leading to the realization that the new coffee stratum and the previously formed intensive stratum are practically identical.

Therefore, land-use strata with physical boundaries are defined and delineated as a function of the combination of some of the following characteristics:

(a) Intervals of percentage of cultivated land;
(b) Predominance of a particular crop or combination of crops;
(c) Pasture land or range land;
(d) Average size of cultivated fields;
(e) Special sites of agricultural activity;
(f) Areas covered with water;
(g) Urban Areas;
(h) Agro-urban areas;
(i) Other non agricultural land consisting of deserts, heavy forests, high mountains, national parks, military reserves, etc.

3.2.3 Strata Boundaries and Minimum Block Size

Boundaries of regions, provinces and strata *must* follow physical features that actually exist on the ground. This is true because any line could become part of a segment boundary and if an enumerator cannot find the boundary on the ground, it is not an acceptable one. The only exception will be national frontiers. All lines should be drawn to one side of and almost touching the physical feature that is being used for the boundary. Placing the line immediately on top of the feature would obscure it in many cases and require erasing to verify the work.

Minimum Block Size. Compact pieces of a stratum outlined on a photograph, map or image are referred to as stratum blocks. Bits and pieces of other strata will almost always be found within any given stratum. The question is "how large do these pieces have to be before we take them into account and outline them as a portion of the appropriate stratum ?" One possible answer to this question is the calculation of how much the variance of the estimate is likely to increase if the odd portion were to be selected in a sample from the stratum that surrounds it. For example, the expansion factor for a pasture land stratum may be large due to a minimum sample having been selected. If a chunk of intense cultivation fell in the pasture land sample, the size of the chunk of cultivation could have a significant effect on the overall estimate of cultivated land due to the large expansion factor for the pasture land stratum. An *absolute minimum size* for a stratum block can be set at the equivalent of one segment. However, for practical reasons, stratum blocks should be made as large as possible.

Acceptable Strata Boundaries. The most permanent and easily discernible physical features make the best boundaries. The more permanent the boundaries that are used, the longer the frame will last. The following is a listing of acceptable terrain features to be used as boundaries, ranked by priority:

(1) Paved Highways;
(2) Secondary All-Weather Roads;
(3) Local Farm-to-Market Roads;
(4) Railroads;
(5) Permanent Rivers and Streams;
(6) Permanent Drainage and Irrigation Canals;
(7) Intermittent Streams and Rivers or Prominent Watercourses that carry water during and immediately following rains;
(8) Field Boundaries;
(9) Internal Farm Roads and Trails.

As a last resort, lines of trees and small gullies can be used but only after a thorough search for more permanent and well defined features has failed. Local farm-to-market roads sometimes suffer drastic realignments due to washouts and/or drifting sand. This would also be true for internal farm roads and trails. Remember that each line drawn has the potential of being a segment boundary. Every effort should be made to use only the first five in the above list for stratum boundaries. Point to point boundaries can be used as a last resort in non agricultural strata and in pasture land or heavily wooded strata where no other boundaries are visible. If livestock can graze across segment boundaries, they can be prorated to the land inside the segment provided the livestock are ranging in fenced pastures. This requires an appropriate questionnaire design.

During data collection, as already mentioned, the enumerators indicate on the segment photo enlargement, or scale drawing of each selected segment, the tracts and the fields of each tract, in which the land use is different. The tracts are allocated letters and the fields numbered consecutively within each tract.

A field of a tract is a block of land, recognized by the holder, that the holder usually treats as a unit. It may be all planted to the same crop or the same mix of crops in any one season, or it may be subdivided into plantings of different crops in which case, for enumeration, each area planted to a different crop is treated as a different field. The crops may be changed for the next planting but the outermost boundaries or "edges" are maintained. Fields are usually recognizable on photography and even on imagery where fields are large. The term as defined here may not be applicable in some countries. When applicable, *the strata, PSUs or segment boundaries should not cut across fields.* Foot paths can sometimes be well defined and be seen to angle across a field whose squared-off boundaries are equally visible. In this case, the foot path or trail is not a good boundary. Pipelines and power lines are often distinctly visible on photography and are usually shown on maps but they are not acceptable as boundaries because fields can be (and often are) established over the pipeline or under the power line and they do not limit the movement of livestock.

3.2.4 Target Segment Size in each Strata

In area sampling, segment size can refer to the measure of land area of a segment (its total area), to its area of cultivated land or, in special strata, to an area of specific use, such as area irrigated or its area under vineyards. In other words, segments size can refer to its *measure of size* given by a variable closely related to the survey variables of interest. Segment size may refer also to the size of the cluster (number of *tracts* in the segment, or the number of associated holdings - reporting units-).

Sample size depends on the method used for associating reporting units (holdings or tracts) with segments. It will be assumed that the weighted, closed or open segment procedures are applied.

The determination of the optimum target segment size in a given stratum involves the study of a number of factors, for example: the comparison of sampling variance for different cluster sizes and the percentage of holdings reporting non-zero values for the different survey variables, cost considerations, the enumerator daily workload, coverage problems associated with segment boundaries, the availability of physical boundaries on maps, data from previous surveys.

In a given stratum, a sample of 20 segments that include 100 holdings each (a total of 2,000 holdings) will generally have a larger sampling variance but less cost than a sample of 200 segments with 10 holdings each. This is because adjacent holdings having similar environment tend to be alike and large clusters of holdings are inefficient.

In order to minimize the sampling variance, since sampling variance depends on the variation of segment size (number of associated holdings), it follows that *segments should be of approximate equal size, and an appropriate ratio estimator should be used.* Since sampling variance also depends on the average segment size, it is in principle convenient to define the segments *as small as possible.* Because segments with one reporting unit per segment cannot be defined in practice.

The previous considerations regarding the optimum segment size refer to the minimization of the sampling variance. However, the problem is to minimize the mean square error (MSE), which is a combination of the sampling variance and the bias. Bias associated with ambiguities with segment boundaries tend to be larger for smaller segments. Indeed, for square or circular segments, for instance, the area is a function of the *square* of the perimeter.

The availability on aerial photography of easily identifiable physical boundaries for segments in the aerial photography, and also the average workload that the enumerator can complete in one day are the practical factors that often determine the optimum average segment size in a given land-used stratum.

In general, experience has shown that the optimum average size of segment is *smaller* than the minimum practical size that is feasible to define using the available aerial photography.

For a fixed cost, the optimum average segment size is difficult to determine especially when estimates are required for many variables and domains of study, which is the case of a multiple-purpose survey.

It is desired then that segments in a given stratum will have approximately equal, small size (area), called <u>target segment size</u>. Available acceptable segment boundaries will make it necessary to allow the actual segment size to vary within rational limits. A variation of plus or minus 25% has been a general practice. If the target size of a segment in a given stratum is 1 square kilometre, the segments as formed would be allowed to vary in size from 0.75 to 1.25 square kilometres. Care must be taken not to have a tendency toward the high side or the low side in forming the segments. Expansion factors will be based on the target segment size in each stratum.

As a starting guideline, the target segment size in a given stratum can be set so that it is equal, on the average, to a day's work for an enumerator. The amount of work is determined, for the most part, by the number of questionnaires the enumerator will have to complete for the segment and how long it takes to finish each interview. A *questionnaire* must be filled for each agricultural *tract* in the segment. A suggested starting number would be to try to form segments with 10 to 15 tracts. Making this determination is not as difficult as it might seem. For one thing, intensely cultivated land with large numbers of small parcels is easily detected on the images or photography, and it may be known (from the census) that the average size of holding in that area is around 1.5 hectares. From this an acceptable target segment size for that area could be one quarter of a square kilometre (25 hectares). Target segment size - 25 ha.- divided by holding size, say 1.5 ha. = 16.66 or 17 holdings or tracts anticipated within the segment. This might be slightly more than the desired number but thought to be manageable. Knowledge of the average size of holdings and percentages of holdings by size in the various areas of interest and strata is therefore important. In particular, an approximation of the number of holdings per segment should be obtained. Initial surveys in some countries have been seriously delayed by selection of segments containing too many holdings.

Since strata with different percentages of cultivated land are used, the target size of the segment (equal in each strata) is directly related to the characteristics of the stratum. Each stratum will have, most probably, a different target segment size. Small segments are usually created in a stratum with a high proportion of cultivated land containing many acceptable boundaries. Large segments are usually created in a range land or pasture stratum with little cultivated land and few boundaries.

Segments in range or pasture land can be very large and highly variable in size and are often selected by point sampling. Segments in the urban stratum should be no larger than a city block.

3.2.5 Sample Size and Allocation to Strata

The sample size in each stratum, the target segment size in each stratum and the total sample size, can be established based upon considerations of estimated coefficients of variation (CV) of the main variables, non-sampling errors, costs and time available for the survey. A good starting point, when reliable estimates of some of these parameters are lacking, is to refer to the experiences in other countries. For this purpose, the reader is referred to the design of agricultural sample surveys in a large number of countries included in Volume II. In any case, after conducting a large pilot survey and the first round of the survey, these parameters should be review.

Formulas for calculating sample size needed for a certain desired CV as well as procedures for properly distributing the sample among the various strata in a stratified sample for a *single* variable are to be found in well known sampling texts. The calculations are straightforward if variances are known; perhaps from a previous similar survey. That, usually, is not the case. With a new area sample survey, the expected variances can only be estimated.

A major problem for optimal sample size estimation is caused by the large number of agricultural variables that is desirable to include in the same survey. However, there are now a number of computer programs that calculate sample size and optimal allocation to strata for multivariate situations where variances are known or can be estimated. A number of sampling texts discuss ways of dealing with the absence of certain knowledge about expected variances.

Although optimum allocation of the sample to strata is the ambition, a number of countries have used proportional allocation due to its simplicity.

Assuming that the survey area is a province, experience has shown that an area sample size of some 300 to 400 segments often allow to estimate major crop and livestock variables with a CV of less than 10 %. Starting with a sample of this size would provide reliable estimates for major items. The sample can be distributed among strata proportional to some characteristic related to an important variable to be estimated. If cultivated area and area of crops are important, the sample can be distributed according to the proportion of holdings or cultivated land believed to be in each stratum. If livestock variables are to be estimated, another sample distribution should be made proportional to the estimated presence of the livestock items in each stratum. The two distributions will undoubtedly be different, at least in some strata. The final sample distribution will be a compromise between the two. A sample this size probably will not provide the desired CV for some items but many of the estimates will still be useful. However, on completion of a large pilot survey (say 200 to 300 segments) survey, solid information can be used to increase or decrease sample size overall and in each stratum; this should certainly be undertaken. The use of a replicated sample procedure (cf. §3.2.8) makes it easy to adjust sample size.

A pilot survey of even 50 segments may not provide variances of sufficient precision to calculate sample size for principal crops and livestock items in all strata nor would it provide usable estimates.

3.2.6 Subdivision of Strata into PSUs. Size of PSUs, Strata and Total Frame Size

Each *stratum* should be completely partitioned into non-overlapping areas with physical boundaries called *primary sampling units* (PSUs) or *counting units* (CU's). The *PSUs* are subdivisions of the *strata* and preferably also of the administrative subdivisions of the Province. The PSUs may even be subdivisions of agricultural or population census Enumeration Areas (EAs), provided the EAs were defined with recognizable physical boundaries. This would allow for additional agricultural or demographic information to be used. But in practice this approach is seldom followed because the census EAs do not have easily identifiable permanent physical boundaries and because this would imply a quite laborious undertaking.

PSUs are defined to facilitate the frame construction and sample selection and, in particular, to calculate the exact size (*number of segments*) of the frame and individual *strata,* whilst avoiding the need to identify all segments in the frame. Once the PSUs are defined and their areas are measured then the exact size (*number of segments*) of the frame and *strata* will be known.

Strata and PSUs are usually identified and delineated on satellite images or a mosaic of aerial photography and then transferred to maps and measured. Sample segments are usually identified and delineated on mosaics of aerial photos since these requires a higher resolution.

In each stratum the PSUs are used to *facilitate the location* of the *selected* segments.

The *target size* of the PSUs (from 4 to 10 segments) is set based on the target segment size for the stratum and the availability of good boundaries. The target size must be well below any sampling interval if systematic sampling is used. If the target segment size for a stratum is 1 square kilometre, PSUs could range in size from 4 to 10 square kilometres. If the target segment size is 0.5 square kilometre, PSU target size would range from 2 to 5 square kilometres and so on. The range allowed in PSU size should be from 2 to 15 segments, but the extremes of this range should be avoided: possibly the lower more than the upper. PSU boundaries must be physical terrain features as described earlier. PSUs should be formed in a manner that allows each one to contain a representative mix of the stratum characteristics. If a stratum is defined as having 50% to 75% cultivated land, every effort should be made to form each PSU so that it contained 50% to 75% cultivated land and *not* less than 50% *nor* more than 75% cultivated land.

A transparent grid will be used to aid in the formation of PSUs of the proper size. Finally, *PSUs will be digitized and measured using a computerized system or measured with a planimeter, and a number of segments assigned to each PSU.*

The total population of segments is then determined by adding the assigned number of segments for all the PSUs which, in turn, cover the entire frame area.

3.2.7 Ordering of PSUs within Strata

With the objective of organizing the work involved in the delineation and measurement of PSUs within each strata, and to facilitate the eventual systematic selection, the PSUs are ordered (by giving a sequential number) in a serpentine fashion within the districts or other small administrative subdivision for which agricultural information can be obtained. Once this initial serpentine ordering of PSUs is completed, PSUs can be re-ordered by similarity of agricultural characteristics. This ordering implies the grouping of PSUs according to some known or detectable characteristic, regardless of their geographic location within the land-use stratum. It implies a further stratification of the frame. For example, it may be possible to identify all or most of the PSUs that produce wheat within the land-use stratum. Those PSUs having similar characteristic can be grouped in a new listing having no relation to their physical location within the land-use stratum. Worthwhile gains can be made in this way if the information used to regroup the PSUs is up-to-date and relatively accurate. The characteristic on which the grouping is based must be distributed evenly among the potential segments in the PSU.

The ordering of PSUs can be done, for instance, by first ordering the districts within the province using techniques of hierarchical clustering analysis Then, PSUs can be ordered within the ordered districts.

3.2.8 Replicated Sampling Selection Methods

The use of a replicated sample allows the overall sample size to be increased or decreased by adding or deleting replicates. It also facilitates the preparation of samples for special surveys and allows for rotation of a portion of the sample in successive survey rounds in order to avoid respondent fatigue.

Replicated sampling can be carried out with the use of either systematic or random sampling. A number of smaller independent samples (called *replicates*) are selected to represent the population or universe rather than selecting one large sample. Care must be taken to ensure that each of the replicas is an unbiased representation of the universe. For example, if it has been decided that a sample of 60 units is sufficient to properly represent a population, with replicated sampling, 4 independent samples (replicates) of 15 units each could be selected from the population instead of one single sample of 60 units. Other possibilities are: 5 replicates of 12 units each; 3 replicates of 20 units; and 10 replicates of 6 units each (this may be the best). At the time of sample selection, additional replicates are selected for sample rotation and management. Replicated sampling for periodic, large-scale, agricultural surveys is recommended for the following reasons:

* The existence of replicates makes sample management easier. The sample size can be readily increased or decreased by adding or taking out replicates in order to achieve a desired precision. Additional replicates or retired replicates can be used for special one time or repetitive surveys.

* Sample rotation to ease respondent burden caused by repeated interviewing is greatly facilitated. The number of replicates in the sample can be coordinated with the desired sample rotation. If the desired level of sample

rotation is 25% each year, four (or a multiple of 4) active replicates are selected. One replicate (or 2 when using 8 replicates etc.) would be retired each year and a new replicate (or replicates) rotated in, resulting in a complete rotation every four years. The replicates that are rotated out after less than four years of service can be rotated back in later in order to bear their share of the survey burden. *The number of replicates can be adjusted to coincide with a complete rotation every three years or every five years.* This procedure also ensures that a significant portion of the sample will be the same from one year to the next, so that a stable measure of change can be calculated by using ratios of current estimates to those for the previous year for units surveyed in both years. The fact that sample rotation is facilitated by the existence of replicates does not mean that rotation is required with a replicated sample. As long as respondent fatigue is not a problem, sample rotation can be delayed.

* Replicated sampling provides the means to test different survey procedures. For example, half the replicates could be assigned to a new approach; the other half would be surveyed using the established method; the test could be between versions of a questionnaire and/or interviewing techniques. The test of difference would be derived using the means and/or totals from each replicate for each approach.

* Quality control is assisted by a quick comparison of replicate totals. A significant deviation by one or more replicates from the others indicates the need for an additional detailed check of data entered for those replicates. It does not mean that the data are incorrect. It simply means that a check should be made before allowing the data to go on to final summary.

* Replicated sampling provides a simple unbiased method for estimating sample variance (even where systematic sampling is used) and population parameters (expansions) using replicate means or totals. This is particularly important where computer power is not available. However, the variance estimate for an individual stratum is unstable due to the reduced degrees of freedom, but it gains stability as strata are added.

Assume that a sample of n segments has been allocated to the land-use strata (size n_h in each stratum) and that the PSUs were measured and ordered within each land-use strata.

Let r_h denote the number of replicates in the land-use stratum h. The number of sample *replicates* in a land-use stratum is dictated by the degrees of freedom in the variance of the estimates and by the desired sample rotation procedure to reduce respondent burden caused by repeated interviewing. Suppose that 5 sample replicates (r_h=5) are selected in the stratum.

For selecting the desired number of replicates (each with $n_h/5$ segments) in a given land-use stratum, the stratum should be completely subdivided into non-overlapping substrata with an equal number of segments called also *zones,* in order to select 5 segments in each zone, one for each replicate.

For each land-use stratum (with N_h segments) , the *number* of *substrata* (L_h) is given by the size (number of segments) of each replicate ($L_h = n_h/5$).

At this point, the allocated sample size in the land-use strata, n_h , should be adjusted, so that n_h is proportional to r_h, the number of replicates. For example, if 5 sample replicates are selected, the n_h values, resulting from the allocation of the sample, are adjusted to multiples of 5.

Within each land-use stratum, the determination of *zones*, is done by following the established order of the PSUs. For that purpose, in each land-use stratum a list should be prepared with the accumulated measures of size of the ordered PSUs.

The sample selection within each stratum can be done by selecting a random sample in each zone (random replicated sampling) or systematically by selecting replicates with identical selection intervals equal to the size of a substrata (systematic replicated sampling). The mechanics of these two selection procedures are described in chapter 7. Although random replicated sampling within each *zone* is recommended, systematic sampling is often used due to its simplicity and because it ensures a geographical spread of the sample if the PSUs were ordered in a serpentine fashion, as explained earlier. Systematic sampling can cause a serious bias if the sampling interval should happen to coincide with a certain periodicity in the arrangement of the segments, but such periodicity is rare in area frame segment listings.

3.2.9 Special Sampling Frames

Special sampling frames for specific localities or survey purposes that are not defined in the basic frame can be prepared by first outlining the area of interest on the basic frame. The PSUs associated with the special area or the special characteristic are identified and listed on another control sheet, the segments are accumulated and the new frame is ready for sample selection. A specialized questionnaire and possibly a different segment size might be used for the special surveys.

3.3 Pilot Surveys

A pilot survey should always be conducted. This involves the selection of a small sample of 20 to 30 segments distributed over all the land-use strata. This pilot survey will provide experience and training for those who will be involved in the further development of the sampling system; will test field procedures and questionnaires; and the completed questionnaires can be used to test editing, data entry and data processing procedures. Do not expect a pilot of this size to provide usable estimates of variance or estimates for any of the variables. Even a large pilot survey of 50 or 60 segments distributed among strata cannot be expected to provide reliable variance estimates for strata. So, the recommended sequence is to conduct a small pilot survey for testing and training and then go to a first survey at a provincial or departmental level with enough sample units to generate usable estimates for important variables and their variances by stratum. With this information, the sample can be refined for the next survey and results can be presented to data users.

CHAPTER 4

ESTIMATION METHODS

This chapter presents the most common estimation methods used for area sample surveys with segments that have identifiable physical boundaries.

The three basic methods to associate segments with reporting units (holdings or tracts) to define the survey variables are described, namely: the closed, weighted and open segment methods.

It presents the formulae for the direct expansion and variance calculations for the closed, weighted and open segment estimators along with their associated strengths, weaknesses and applications.

Radio estimators and their variance formulae, both for within-survey and between-surveys, are also presented in addition to their strengths, weaknesses and applications.

4.1 Methods to Associate Segments with Reporting Units (Holdings or Tracts) to Define the Survey Variables

For the agricultural sample survey considered, to obtain the estimate for the survey area of interest of a given *variable* it is necessary to define the variable in each *sampling unit* as a function of its values in a group of *associated reporting units*. In other words, for a given survey variable (defined in the set of reporting units), such methods refer to the way the data of the reporting units (holdings or tracts) is aggregated to define the segment totals.

Three types of methods have been used to associate reporting units (holdings or tracts) to the sampling units (segments) and define the value of the survey variables for each segment: the closed segment, the open segment and the weighted segment methods.

Consequently, three general types of area sample estimators are considered, namely *weighted segment estimators, open segment estimators* and *closed segment estimators,* that correspond to the various methods that establish rules of association between the segments and the reporting units.

The type of estimator chosen for a survey variable depends on its reporting units (the way data pertaining to that variable has been collected). The weighted segment and open segment estimators can be used for variables for which the *holding* is the reporting unit, and the closed estimator corresponds to variables for which the *tract* is the reporting unit.

A direct expansion area sample estimator of a given survey variable denotes the estimator that expands the data of each segment by the inverse of the sampling fraction in each stratum.

Methods used in area sample surveys to define survey variables in each segment

Sample Unit	Reporting Unit	Method of Association
Segment	Holding	Weighted segment estimators
Segment	Holding	Open segment estimators
Segment	Tract	Closed segment estimators

Closed Segment Method

For the closed segment method, the value of a variable in a segment is simply the sum of its values in each of the tracts of the segment.

For a given variable and segment, data on the totality of a holding is not needed except for holdings totally included in the segment. The closed segment method requires the collection of data (by direct observation or by personal interview with the holders) corresponding to agricultural activities that are found physically within the boundaries of the segment. If information on land use is required, data is collected on land use within the boundaries of the segment. If a holder of a tract within a segment is not available for an interview, the area of planted crops in the tract can be check on the ground and identified in the aerial photograph, and their areas estimated. Effective use of the closed segment depends on having photographic enlargements of a known scale or scale drawings of the segment to control data collection. Therefore, response and coverage errors are relatively low.

The closed segment is the normally accepted method to estimate planted crop areas. With regard to information on livestock, the closed segment method is only appropriate for those livestock controlled in pastures and/or corrals within the boundaries of the segment at the time of the enumeration. For estimating livestock that can roam outside the segment, the closed segment method should not be used.

The closed segment method has the major advantage, as compared with the weighted and open segment methods, of being independent of the definition of a holding so that it eliminates the ambiguities in ascertaining the land contained within a holding. However, for most variables of a multiple-purpose survey, the closed segment method is not applicable since the tract is not the appropriate reporting unit. Some examples include: crop production and yields, planting intentions, crops harvested, farm labour, economic data such as sales or cost of production or, as already mentioned, livestock that can roam freely across segment boundaries.

Weighted Segment Method

For the weighted segment method, the variable in each *tract* is defined as the value of the variable in the holding multiplied by a factor equal to the ratio between the area of the tract divided by the area of the holding. Then, the value of a variable in a *segment* is the sum of the variable in each of its tracts.

For the weighted segment method data is collected for the entire holding associated with any land area in the sample segment. The data collected from each such holding is weighted by the proportion derived by dividing the area of the holding that is within the segment by the entire area of the holding (both inside and outside the segment).

The effect of this method is that data for each holding is prorated among the segments in which it is located. For instance, if 10% of a holding's area is in the segment, 10% of the total holding data recorded on the questionnaire will be assigned to the segment and thus to the summarization process.

The weighted segment estimator can be used for all survey variables, since the *holding* is the reporting unit required.

The weighted segment method do not require that the holder, the location of his residence or the holding headquarters are precisely established. But it is obviously necessary to define the holding and find an informant that can provide data for the *totality of the holding*. The use of the weighted segment method simplifies the problem of urban segments if the cultural system in the country is such that each holder can be related to a definite piece of agricultural land (holding). Where that is the case, the urban segments can simply be observed, and if there is no agricultural land or activity no enumeration is necessary since it is assumed that any holder that might live there will have the opportunity to be selected where his land is; and local custom would not include the area of a urban residential lot as part of the holding. This has an important implication for the construction of the area frame, since, in the above case, the use of the weighted segment method does not require to construct a frame for urban areas.

However, in countries where there are many holders who graze livestock on public or communal land without any lease or other means to relate them to the land, it will still be necessary to canvass city and urban segments if the weighted segment method is being used so as to record livestock belonging to those who profess to have no land.

Open Segment Method

The <u>open segment method</u> (or holding headquarters method) associates a segment to all holdings with headquarters included in the segment. For this purpose, clear rules have to be established to define a unique reference point for each holding, called the *headquarters*. There are several ways to do this, but the most common procedure is to define the headquarters as the dwelling (residence) of the holder.

If an open segment method were to be used, the holder must be *uniquely* identified which involves practical difficulties in cases where more than one person or household are involved in the operation of the holding. In addition, to use open segment estimators the urban areas have to be included in the area frame. Since in general a large number of holders live in urban areas, the urban areas have to sampled which adds great difficulties to the area sample method applied because it would be necessary to screen a large number of dwellings to find the holders. However, using the open segment method with segments in the urban areas is the only way that an area sample can capture information from holders that cannot be related to a piece of rural land.

Open segment estimators were the first type of estimators utilized in area sampling, and are still used in some countries. Open segment estimators are no longer used in the United States and were also substituted in other countries (Brazil and Italy, for example) due to the difficulty that arises from creating a one-to-one correspondence between holders and holdings and also locating the residence of the holders in urban areas. In the agricultural surveys conducted in the US, the open segment estimator has been substituted by the weighted segment estimator for variables for which the reporting units are the holdings; and by the closed segment estimator for variables for which the reporting units are the tracts. Therefore, the open segment method of estimation implies special complications for area frame construction and for conducting a survey involving objective measurement of areas.

4.2 Direct Expansion Area Sample Estimators

The estimation formulas correspond to a single-stage design for simplification purposes. It should be noted that for the area sample design considered, the total number of segments is exactly known for each strata and PSU, and the probability of selection of each segment (constant within a stratum) is equal to its conditional probability in two stage sampling.

4.2.1 Closed Segment Estimator

A direct expansion area frame estimate, for which non-response must be manually imputed, calculated by expanding tract level data for each holding with at least part of its land located in a sample area segment, or by expanding tract level data for the non-holding land of a sample segment.

$$Y_c = \sum_{h\in S}\sum_{j\in B_h}\sum_{k\in G_{hj}} e_{hjk}\sum_{m\in T_{hjk}} t_{hjkm}$$

where

Y_c	is the sample estimate of a total for the survey variable y,
S	is the set of all land-use strata,
B_h	is the set of all substrata in stratum h,
G_{hj}	is the set of all segments in substratum j of land-use stratum h,
T_{hjk}	is the set of all tracts in segment k of substratum j or land-use stratum h,
e_{hjk}	is the expansion factor for all tracts in segment k,
t_{hjkm}	is the tract value for the variable y associated with tract m.

***Strengths*:**

- Very precise estimates can be made, especially for variables correlated with farmland and easily associated with the tract, such as crop areas.
- Biases due to non-response can be decreased because tract level data can usually be observed fairly accurately.
- Data collection efforts required are the simplest and easiest to follow, with fewer errors due to coverage problems stemming from survey concepts and definitions and associating holding Headquarters within the segment.
- Reported data for the tract are less susceptible to reporting errors than holding level reported data and are less sensitive for the respondent to report.
- Duplication of data between segments is less likely to occur as when entire holding data is reported.

Weaknesses

- Can not be used to estimate certain survey variables that are not associated easily with the tract, such as economic items.
- Is not efficient in estimating rare items or items concentrated in small, geographic areas (rice, sugarcane, etc.) unless with an ad-hoc stratification.
- Data collection costs will be higher than for the open estimator, resulting in two to three times as many interviews.
- Response errors can occur through misidentification of segments and tracts due to obsolete maps and photos and respondent unfamiliarity with the tract reporting concept.

***Applications*:**

- The best application is for areas of major crops that are commonly distributed across the Province. It can also be used for some common livestock items such as cattle.

4.2.2 Weighted Segment Estimator

A direct expansion area sample estimate, for which non-response must be manually imputed, calculated by expanding total holding data for each holding with land found in the sampled area segment after prorating the data by the proportion of the total holding land found in the segment.

$$Y_w = \sum_{h \in S} \sum_{j \in B_h} \sum_{k \in G_{hj}} e_{hjk} \sum_{m \in T_{hjk}} w_{hjkm} y_{hjkm}$$

where

Y_w is the sample estimate of a total for variable y,
S is the set of all land-use strata,
B_h is the set of all zones in stratum h,
G_{hj} is the set of all segments in zone j of land-use stratum h,
T_{hjk} is the set of all tracts in segment k of zone j of land-use stratum h,
e_{hjk} is the expansion factor for all tracts in segment k,
w_{hjkm} is the weight used to prorate y_{hjkm} (usually area of tract m divided by the total holding area for tract m), and
y_{hjkm} is the value for the variable y associated with tract m.

***Strengths*:**

- Can be used to estimate all agricultural variables (characteristics) pertaining to the holdings.

- Usually the most precise of the three kinds of area estimates, especially for livestock and economic items.

***Weaknesses*:**

- Data collection costs are the highest of the three kinds of direct area sample estimators -entire holding data must be collected for all tracts.

- There is a tendency for holders to underreport total holding areas by omitting non-cropland areas - this biases the weighted estimates upward.

- Entire holding data for non-respondents can not usually be observed within the segment. Nonrespondents must have their holding level data manually imputed subjectively.

- Coverage of holdings in densely populated areas can be poor due to the difficulty of screening for holders, multiple holders in a single household, and small holdings.

- Entire holding data are susceptible to reporting errors if the respondent is not knowledgeable about the entire holding or misunderstands survey concepts relating to the entire holding.

***Applications*:**

- All agricultural variables, including crop areas, stocks, and production, livestock inventories and economic items.

For the surveys conducted in the United States by the National Agricultural Statistics Service (NASS) a modified weighted estimator is now in use. It attempts to moderate the undercoverage problem by universally deleting one-half acre -0,405 ha.- (corresponding to the residence area) from all holdings, thereby avoiding the need to screen densely populated areas for holders. Also, alternative weighting schemes involving some other variable than total holding area have been tested, but none have proven successful enough to replace holding area.

4.2.3 Open Segment Estimator

A direct expansion area frame estimate, for which nonresponse must be manually imputed, calculated by expanding total holding data for each holding in which the holder or Headquarters resides in the sampled area segment.

$$Y_o = \sum_{h \in S} \sum_{j \in B_h} \sum_{k \in G_{hj}} e_{hjk} \sum_{m \in T_{hjk}} a_{hjkm} y_{hjkm}$$

where

Y_o	is the sample estimate of a total for variable y,
S	is the set of all land-use strata,
B_h	is the set of all zones in stratum h,
G_{hj}	is the set of all segments in zone j of land-use stratum h,
T_{hjk}	is the set of all tracts in segment k of zone j of land-use stratum h,
e_{hjk}	is the expansion factor for all tracts in segment k,
a_{hjkm}	is 1 if the holding Headquarters for tract m in segment k is within the segment and zero otherwise, and
y_{hjkm}	is the entire holding value for variable y associated with tract m in segment k, of zone j and stratum h.

***Strengths*:**

- Can be used to estimate all agricultural variables pertaining to the holding.
- Data collection costs can be lowered as only holders resident in their holding need to be contacted.

***Weaknesses*:**

- Usually the least precise of the three kinds of area estimates.
- Entire holding data for nonrespondents can not usually be observed within the segment. This calls for subjective manual imputation of data, a difficult task.
- Entire holding data are susceptible to reporting errors if the respondent is not knowledgeable about the entire holding or misunderstands survey concepts relating to the entire holding.
- Usually results in an underestimation of the farm population due to the difficulty of screening for holders in densely populated areas.
- The holder is more likely to refuse information about his entire holding than for a part of his holding.

***Applications*:**

- All agricultural variables can be estimated, particularly those easily associated with the entire holding such as livestock inventories, grain stocks, number of holdings, gross value of sales, and other economic characteristics. As already mentioned, NASS discontinued use of this estimator, however, when weighted estimators proved to be more efficient for these variables.

4.3 Variances for the Direct Area Sample Estimators

The variance formulae for the three area sample estimators are nearly identical. Let the expanded tract level value for the variable Y be:

Y_{hjkm}^{e} $= e_{hjk}t_{hjkm}$ for closed estimates,

$= e_{hjk}a_{hjkm}t_{hjkm}$ for open estimates,

$= e_{hjk}w_{hjkm}t_{hjkm}$ for weighted estimates, and

n_{hj} = number of segments in G_{hj}.

Then the expanded value for item Y for a segment is:

$$Y_{hjk.}^{e} = \sum_{m \in T_{hjk}} Y_{hjkm}^{e}$$

And the variance is:

$$V = \sum_{h \in S} \sum_{j \in B_h} [(\frac{n_{hj}}{n_{hj}-1})(\sum_{k \in G_{hj}} (Y_{hjk.}^{e})^2 - \frac{(\sum_{k \in G_{hj}} Y_{hjk.}^{e})^2}{n_{hj}})]$$

4.4 Ratio Area Sample Estimators and their Variances

It has been pointed out that one of the advantages of replicated sampling is the ability to easily replace a portion of the sample each year, allowing the respondents in the replaced segments a respite. The portion of the sample rotated out each year is usually 20 to 25%. The direct expansion of the sample data including the new segments may or may not be a good indication of change from the previous year since it is difficult to judge how much of the indicated change may have arisen due to the change of segments in the sample. A ratio estimate using segments that were in the survey in both years provide a better estimate of the changes unless there is a low correlation between the values of the variable for the two years (this year and last year).

As an example, consider an estimate of the area planted to corn in a certain year. A direct expansion of area planted to corn using the closed segment estimator provides one figure for corn area. A more stable estimate of change is provided by calculating a ratio between the area planted to corn this year and the area planted last year in segments that were surveyed both years. This ratio, expressed as a percent (90%; 100%; 110%) is then applied to the estimate of corn planted for the previous year. The CV of the ratio is governed in large part by the size of the correlation between the land in corn this year and the land in corn last year. If the correlation is high, the variance will be lower; if the correlation is low, the variance will be high. However, the application of this ratio with it's own variance to the previous year's direct expansion, that also has variability, will usually result in a ratio estimate with a higher variance than the current direct expansion. If the ratio estimate is significantly different to, or if the ratio estimate indicates a different change from the previous year to the one obtained by the direct expansion, a careful review of the data from the new segments is indicated.

4.4.1 Within Survey Ratios

An area sample ratio estimate for which nonresponse must be manually imputed, and which measures the proportion of one variable to another variable from the same survey. Examples are yields for crops, milk production per cow, litter rates for hogs, etc. The numerator and denominator totals are calculated at the land-use stratum level and then summed over strata to produce the Province level ratio.

$$R^{w} = Y^{/} / X^{/}$$

where R_w is a sample estimate of a within survey ratio of variables Y and X, and $Y^{/}$ and $X^{/}$ are the appropriate direct expansion estimates (open, closed, or weighted), for items Y and X.

The variance of the within survey ratio estimate is :

$$V_w = R_w^2[(\frac{V_{yy}}{(Y^/)^2}) - 2(\frac{V_{yx}}{(Y^/X^/)}) + (\frac{V_{xx}}{(X^/)^2})]$$

where

V_{yy} = the variance estimator for $Y^{/}$,
V_{xx} = the variance estimator for $X^{/}$, and
V_{yx} = the covariance estimator for $Y^{/}$ and $X^{/}$, and

$$V_{yx} = \Sigma_{h\in S}\Sigma_{j\in B_h}[(\frac{n_{hj}}{n_{hj}-1})(\Sigma_{k\in G_{hj}}(Y_{hjk.}{}^{e})(X_{hjk.}{}^{e}) - \frac{(\Sigma_{k\in G_{hj}}Y_{hjk.}{}^{e})(\Sigma_{k\in G_{hj}}X_{hjk.}{}^{e})}{n_{hj}})]$$

Strengths:

- If the two variables are highly correlated (such as harvested and planted acres) the ratic estimate will probably have lower CV's than a direct expansion estimate.

***Weaknesses*:**

- For some variable pairs (such as harvested vs. planted acres, or milk production vs. number of cows milked) asked on the same questionnaire, there may be memory bias associated with one of the responses, or an intention is recorded rather than an actual value.

***Applications*:**

- Crop yields and planted-to-harvested area percentages, stocks-to-production percentages, livestock birth and death rates, percents of subgroups to totals, milk or wool production rates, and a variety of economic items such as percent of farms by size and category. USDA/NASS uses mainly closed estimators for numerators and denominators of crops ratios, and weighted estimators for numerators and denominators of livestock and economic items. Some closed estimators are still used in ratios for cattle items.

4.4.2 Between Survey Ratios

An area sample ratio estimate for which nonresponse must be manually imputed, and which measures the percent change in a characteristic from the previous (base) survey. Most examples are percent changes in totals (livestock inventory levels, crop planted areas, and number of holdings, etc.) The current and previous year total estimates are calculated at the land-use stratum level and then summed over strata to produce the Province/State level ratio. Only matching area frame segments may be used in the calculation -NASS rotates approximately 20 percent of its area frame replicates each year, so in most cases the between survey ratios are calculated using only 80 percent of the current year's segment data.

$$R_b = \frac{\sum_{h \in S} \sum_{j \in B_h} \left(\frac{n_{hj}}{n_{hj}}\right) \sum_{k \in G_{hj}} Y_{hjk.}^{e}}{\sum_{h \in S} \sum_{j \in B_h} \left(\frac{n_{hj}}{n_{hj}}\right) \sum_{k \in G_{hj}} X_{hjk.}^{e}}$$

where

R_b is a sample estimate of a between survey ratio of variable Y in the current year and variable X of the previous year (usually the items are the same, but Y and X are used in these equations for ease of notation),

$G_{hj'}$ is the set of all segments in substratum j of land-use stratum h in both the current year and previous year sample, and

$n_{hj'}$ is the number of segments in $G_{hj'}$.

The variance formula would be very similar to the variance for the within survey ratio. Variances for Y' and X' and the covariance for $Y'X'$ would be calculated using $n_{hj'}$ and $G_{hj'}$, and the $Y_{hjk.}^{e}$ and $X_{hjk.}^{e}$ terms would be replaced with $(n_{hj}/n_{hj'})\Sigma Y_{hjk.}^{e}$ and $(n_{hj}/n_{hj'})\Sigma X_{hjk.}^{e}$ respectively.

***Strengths*:**

- Compares current reported data from two different time periods, thus avoiding memory bias.
- If the two characteristics are highly correlated the ratio estimate will probably have lower CV's than a direct expansion estimate.
- Closed area estimates for major crops within segments across years can be highly correlated and are not negatively affected by changes in operating arrangements and respondents from year to year.

***Weaknesses*:**

- Weighted area estimates based on reported holding level data (such as for livestock and economic items) can have low correlations across years due to changes in holdings found in the segments and differences in respondents from one year to the next.
- May not be appropriate for commodities that growers go into or out of on a cyclical basis.

***Applications*:**

- Same as Within Survey Ratios.

For estimation methods in area sampling designs, the reader can refer especially to Kott (1990) and Nealon (1984).

CHAPTER 5

RESOURCES REQUIRED

This chapter describes the resources required for the establishment of the area sample survey with segments that have identifiable physical boundaries; in particular the cartographic requirements according to the types of area units used for the area frame construction.

Careful attention has to be paid to the various resources discussed below before starting an area frame construction. The quantities needed will depend, of course, on the size of the area frame to be constructed. It is advisable to select one province or region as a basis for calculating with reasonable precision the resources required: these can be extrapolated to cover whatever additional portion of the country is to be included.

5.1 Cartographic Materials: Maps, Satellite Images and Aerial Photos

Maps of different types and scales, satellite images and aerial photography are used for identifying and measuring areas for area frame construction and sample selection. Aerial photography and/or detailed maps are also used for identifying and measuring areas during and after the survey data collection.

Strata, *PSUs* and *segments* are the three main types of area units that will be delineated using the available cartographic materials. The boundaries of the Provinces and preferably their administrative or political subdivisions also require identification. Population or agricultural census *enumeration areas* (EAs) may be of use for area frame construction; and if so these also need to be identified on the cartographic materials.

One example of the several possibilities of the range of different types of cartography used for area frame construction and sample selection is as follows:

* Maps for survey planning purposes (scales 1:500,000)
* *Strata* and *PSUs* constructed on satellite images, ortho-photomaps or a mosaic of aerial photographs (scales 1:100,000)
* *Strata* and *PSUs* transferred to and measured on maps (scales 1:100,000 or 1:50,000)
* Selected *PSUs* transferred to aerial photo-mosaics, and selection of sample *segments* (scale 1: 25,000)
* Selected *segments* located on aerial photos for use during data collection (scale 1: 5,000)

Locating the source of such materials is only half the battle. For example, the Geographic Institute may be the source of cartographic needs but obtaining the active cooperation of those persons in the Institute that are sympathetic and helpful is essential and often takes time.

5.1.1 Maps

Maps are virtually indispensable: every effort must be made to obtain them for the area to be covered by the frame. Topographic maps on a scale of 1:100,000 (or 1:50,000) are usually available from a geographic institute or the military or an external source and should be obtained for the area in question regardless of what other materials may be available.

Maps are important because they are often the only completely true to scale representation of the area upon which accurate measurements can be made. Normal contact prints have significant distortion around the edges and the same is true for a standard satellite image if not ordered corrected for scale. Large scale (1:5,000 or 1:10,000) maps may have been prepared for urban areas and should also be obtained. In some cases, maps are not available for certain portions of the frame area. Ways of dealing with this problem will be discussed in Chapter 6 under frame construction. In addition to topographic maps, up-to-date road maps and smaller scale (1:250,000 or 1:500,000) maps will be needed for planning purposes. The exact scale that is most useful will depend on the size of the frame area. A master map showing the location of sample segments should be made for each province or political subdivision of the frame. Sets of 1:50,000 maps showing segment locations and road maps are useful to the supervisors and enumerators for planning their daily work.

5.1.2 Satellite Images

Satellite images provide the best tool for subdivision and measurement of the land into strata according to different proportions of cultivated land or other land-use characteristics. They are also used for identification and measurement of urban and suburban areas, surfaces covered with water and other non-agricultural areas. Images from the LANDSAT-V satellite and from the SPOT 3 satellite are available. Satellite images and satellite photos may be available from other satellites and in particular from Russian satellites whose data were previously classified or impossible to obtain. Quality and price should be carefully checked before purchasing. To be used as frame material rather than just reference material, quality images on a scale of 1:100,000 are needed.

Satellite sensors sense and transmit reflected energy in a number of different wavelengths or "bands". Some bands or combinations of bands (panchromatic or multispectral) -2,3 and 4- may be found to be better for identifying boundaries than others, whereas other bands or combinations (false color) may give a better picture of crops and other ground cover. Different combinations of bands should be tested under local conditions. Then, the images with the band or combination of bands that are found to be most useful can be ordered.

It is essential that the images contain little or no cloud cover. The image index will specify the percent of cloud cover. Alternate dates of images may be required if the cloud cover percentage exceeds 5 - 10%. The date and color composition of satellite images should be chosen appropriately.

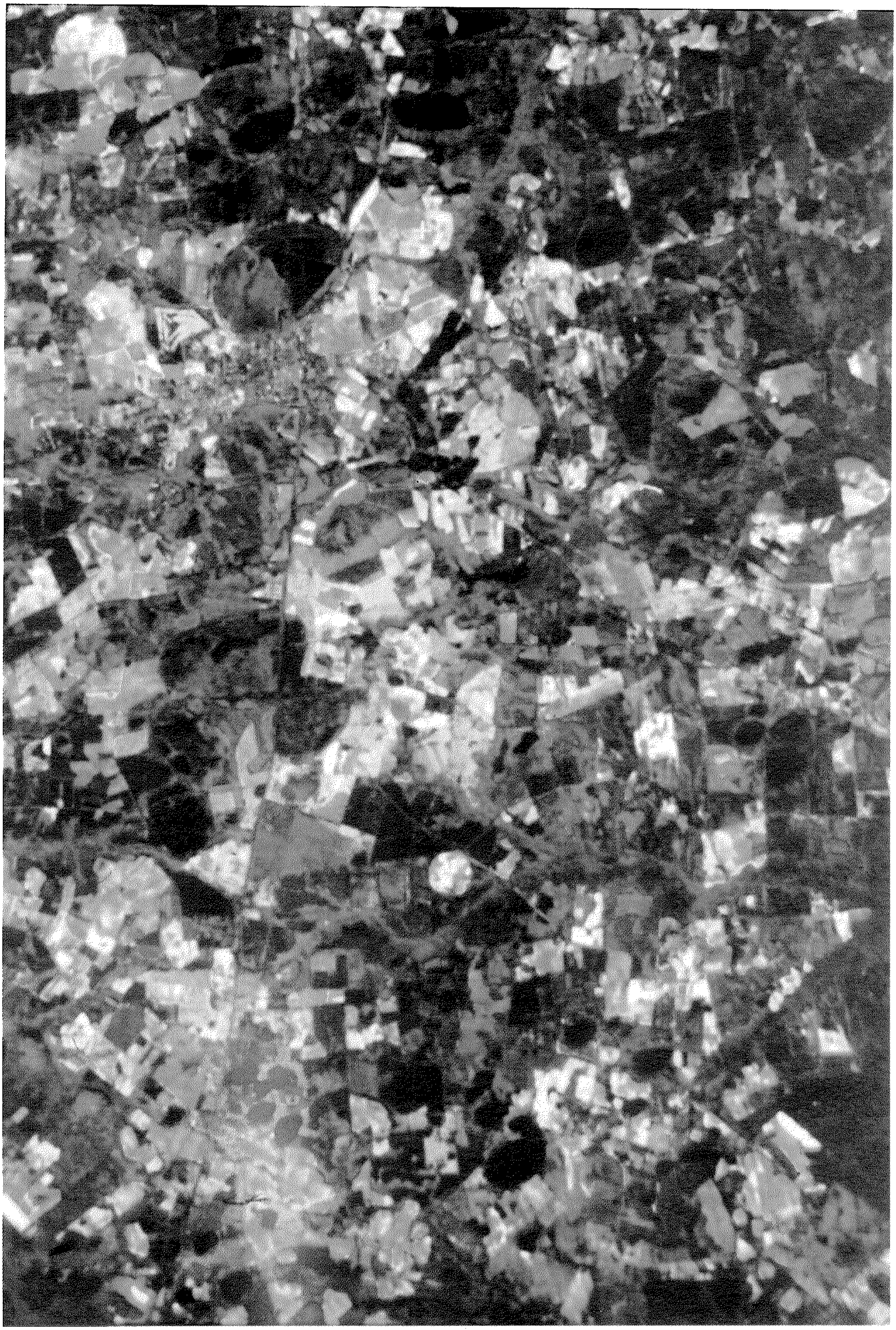

Figure 1. A TM/LANDSAT 5 scene

Satellite images can be obtained with varying degrees of geometric correction; up to four levels, horizontal and vertical. In areas where there are no extreme changes in elevation, a lower level correction or even no correction at all may be satisfactory but for areas with extreme changes and extensive terrace cultivation, the highest degree of correction possible is recommended.

A satellite image scene taken by the Thematic Mapper (TM) sensor in the LANDSAT satellite covering 31,450 square kilometres can be acquired and printed at a scale of 1:250,000. Prints of TM/LANDSAT subscenes at 1:100,000 scale are usually of a quality that can be used for stratification and formation of PSUs. These 1:100,000 scale scenes cover 10,000 square kilometres (100x100) and also cost about $3000; which shows how inexpensive they are in relations to the advantage they can provide. For this price, the positive and negative transparencies are provided giving an up-to-date view of the selected area at a price well under that for aerial photography.

It may be worthwhile to consider getting SPOT 3 satellite imagery, which has better resolution than the TM, especially for identifying areas where more detail is essential. However, it should be noted that SPOT data is significantly more expensive than TM/LANDSAT data.

Section 6.5 includes some further considerations on the use of satellite images for area frame construction and sample selection.

In Brazil, the area frame stratification used mostly prints of satellite imagery TM-LANDSAT V on 1:100,000 scale, color composition, channels 2, 3 and 4, or channels 3, 4 and 5, depending on the area; since those combinations were found to be the most suitable for stratification purposes. In the United States, for the stratification of the area frames updated since the last four years, digital satellite imagery TM-LANDSAT V, 1:100,000 scale, color composition, channels 2, 3 and 4 is generally used. In Pakistan, the stratification and formation of PSUs for the entire national frame was done on 382 SPOT images at a scale of 1:100,000. Each image covered an area of 60 km. x 60 km. at a cost of $350 each (in 1993). Further details of these examples are contained in Vol. II.

5.1.3 Aerial Photography

Aerial Photography can be obtained generally from the geographic institute or from the military. *Orthophotomaps* which are a type of map made using corrected (true to scale) photo mosaics in a scale of 1:25,000 or 1:50,000 are sometimes available and at least one set, regardless of age, should be obtained for the frame area. If the orthophotomaps are relatively up-to-date and of good quality, the frame can be built on this material. Satellite images can be used to determine changes in land use. However, in most countries, *standard aerial photography* for large area coverage is only available in what are called contact prints. A contact print is a photograph made directly from a negative without enlargement or reduction. They are about 25 cm. by 25 cm. in size. Any scale of contact print from 1: 20,000 to 1: 60,000 can be used. The ideal scale may be about 1: 30,000. A scale of 1:20,000 covers a very limited area with one print, whereas scales above 1: 40,000 require extensive use of magnifying glasses to find usable boundaries. In one frame building procedure to be described, contact prints are used to form simple mosaics, each mosaic coinciding with a topographic chart at 1:50,000 scale. The geographic institute or military will have what is known as stereographic coverage in which the area covered by each print overlaps by about 60% with adjoining prints in the flight line. A saving can be realized by purchasing only simple coverage that would require approximately half the existing prints for an area. It is advisable to go to the source and lay out the photos to ensure coverage rather than just ordering every alternate print. Contact prints should be ordered on double weight paper with a matte finish. The double weight allows them to withstand the handling to which they will be subjected and the matte finish is easier on the eyes and accepts marking better than the slick or gloss finish.

Current photography is the most desirable, but it is possible to work with photography as much as 20 years old *if* the area is well developed with few changes in land use or infrastructure in recent years. Even the most recent satellite images can be out of date in areas of new development or colonization.

5.2 Personnel

People make or break a project. A statistician with experience in area sampling to head the project plus a few junior assistant statisticians are essential. The leader will need a secretary. A person with extensive experience in cartography and photo interpretation is needed to head the frame construction and maintenance group. If the initial project starts with a province, about 6 to 8 frame construction workers should be employed. If the activity will be regional with intent to cover the country, a minimum of 15 should be employed in frame construction. The best type of people for frame construction are university students or recent university graduates. Recent high school graduates would also be acceptable and probably easier to incorporate into the agency structure. Such personnel should have good eyesight, intelligence and be amenable to learning new skills. An effective way of selection is to give a class in frame construction for about twice as many people as are needed. Those with the appropriate aptitudes and skills can be identified by observation and exam performance. Working with tenured civil servants can be successful but there must be a prior agreement that will allow the removal of individuals from the group as necessary.

Those who will head the field staff should participate in the frame construction from the beginning. Their knowledge of agriculture and rural conditions will be essential for frame construction. At the same time, understanding the construction and characteristics of the frame will lead to a more effective application of the frame and sample in surveys. If the first survey is to be conducted in a province or region with a sample of about 250 segments, some 24 enumerators should be hired for a period of six weeks. They could be formed into 6 teams of 4 enumerators each plus a supervisor. They would receive a full seven days of training and spend the rest of the time collecting information. In some cases it has been found convenient for enumerators to work in pairs during data collection in each segment, because of the type of work that involves the application of questionnaires and identification of boundaries in aerial photography.

5.3 Area Measurement and Scale-Transfer Instruments

Area Measurement Instruments. Precise cartographic measurements of the various components of the frame are necessary.

Grids, made of transparent plastic with squares that represent a known area on the ground when laid over a map or photograph with the scale for which the grid was made, are used constantly for estimating areas during all phases of area frame construction. Technicians can construct their own grids once the scales of the frame materials are known. Grids are used also by enumerators to verify on the field the land areas reported by holders during data collection.

Planimeters. In the past, the planimeter was the instrument most often used and the new digital planimeters with roller bars make the work as efficient and perhaps faster than some forms of digitizing. Although their reliability is high, it is prudent to buy at least two planimeters so one can be sent for repair while one is working. For a national project, the purchase of at least 6 planimeters is advisable. The digital planimeters can be set to read out in the desired units when measuring on a wide range of scales. Many of them can also store readings and perform mathematical calculations. Planimeters are quite inexpensive and very practical. They are useful even if computer methods for area measurement are available.

Computerized instruments for area measurements. Digitizing tables connected to a computer or a Computer Graphic System (that allows also the transfer of scales) are another means of measuring the components of an area frame. One of the important advantages of digitizing is that the records will be stored in the computer and can be manipulated for sample selection among other things. The digital record can be used to generate a copy of a frame map needed. Graphics display terminals and a plotter are needed to make full use of a digitizing system.

Scale Transfer Instruments are used to transfer boundaries of areas between satellite images and topographic charts and between topographic charts and aerial photos. This can be done with a computer graphic system (Intergraph System, for instance, are available in some countries) or with a Zoom Transfer Scope, optical equipment useful in visually overlaying and coordinating the scale between the different cartographic materials. Test before buying, since some instruments are cumbersome to use.

Geographic Information Systems. Computer assisted area frame construction and sample selection is currently a standard procedure in the United States. The system uses Thematic Mapper (TM) LANDSAT digital data and US. Geological Survey Digital Line Graph data instead of paper prints of satellite images and maps. It uses a computer screen to display cartography instead of paper-based cartography. However, aerial photos and detailed maps printed on paper are still used to some extent in the computerized procedures as summarized in Volume II. Sophisticated software and extensive experience with area frame construction make this possible. A geographic registration is made between satellite data and a digital map. One characteristic of the system is that the land-use stratification is performed visually on the computer screen by forming *PSUs* instead of *stratum blocks*. The size of the area that can be viewed at one time is small so that the formation of blocks is not feasible.

5.4 Computers and Vehicles

Computers. A country size operation should have at least 6 to 8 state-of-the-art PC's for summarizing survey data and maintaining frame records. Two PC's could suffice for a pilot area.

Vehicles. If frame construction starts with a province, the frame construction crew should have two four wheel drive vehicles available for field checks as needed. One vehicle for every four or five enumerators and their supervisor will be needed for efficient large scale survey operation. Regional supervisors and the national field supervisor each needs a vehicle.

5.5 Miscellaneous Equipment and Supplies

Miscellaneous equipment and materials will be needed to perform various tasks in frame construction. Test all the items for quality before buying in quantity. Make sure that there is a dependable supply in local stores of the brands selected or that their import as needed will not cause undue delays. Quantities needed will depend on the size of the project. The items listed are those that would be needed if the frame is constructed using a combination of photography, images and maps to include the construction of mosaics. If digitizers, plotters, zoom transfer scopes and computerized geographic information systems are available, many of the items will not be needed. Geographic Position Systems could be very useful to determine the location of the selected segments and for the verification of the strata boundaries.

Persons working on the frame should have:

* A supply of *common pencils* (HB) is needed for recording frame data and other office work, *erasers* and a pair of large *shears* along with two pair of smaller, less expensive *scissors*.

* A good supply of both block and pencil type *erasers* will be necessary. The soft white erasers made by Staedtler and other companies are of common use. A small supply of plastic "ink" erasers for erasing ink from film and pencil from paper with a minimum of damage are needed.

* *Pencil sharpeners*, placed in strategic locations are required. Electric models are nice but not necessary. Each worker needs a *pocket pencil sharpener* of the type that will sharpen common pencils and grease pencils.

* A *paper cutter* to assist trimming the margins from contact prints.

* A *metal meter stick* will be useful for drawing long straight lines and for cutting.

* *Short rulers* (10 to 20 cm.) for measuring on maps and photos and drawing short straight lines.

* Two or three *scaling rulers* will be needed for the accurate transfer of boundaries from photos and images to maps.

* *Magnifying glasses* are required to enlarge portions of the photography or image when searching for proper boundaries for PSUs and strata.

* It is important to have a *light table* for tracing and for viewing transparencies. There is a portable table top light table that would be sufficient if a larger model cannot be obtained.

* Several rolls of both one inch and half inch *drafting tape* will be needed to hold photographs, maps, and images in place on the work table. It is important to purchase drafting tape for this as other types of tape may damage the surface of the map or photograph.

* A small, *portable hand operated binder* will make binding the edges of maps, photographs, acetate, and images easier and faster. Special half inch tape will have to be purchased for use with the binder.

* *Paper masking tape* can be used to bind the edges of frame materials if a binder is not available. A rather large amount of tape may be required.

* *Coloured pencils* will be used for placing the components of the frame on maps. The number of different colours used will depend on the number of strata employed. Either permanent or erasable pencils may be used. Erasable pencils have an obvious advantage but their color is not quite as intense and they will fade more than the "permanent" types. These disadvantages are usually not enough to offset the ability to make changes without ruining maps. "Faber Castell Col-Erase" is an erasable brand of pencil that has been satisfactory. "Berol Eagle Verithin" and "Berol Prismacolor" are quality pencils of the "permanent" type. *Neither* the "permanent" *nor* the "erasable" type pencils are to be used on photography or images. The hard lead will damage the surface. Test pencils on your material before purchasing in quantity. There are types of "all media" *coloured pencils with softer leads* that can be used on photography and images without damage and they can also be erased. Two of the brands that have been used successfully are "Schwan-Stabilo All-Stabilo" and Staedtler Mars-Omnichrom". This type are good for photos and images and drafting film but are not to be used on maps. They are available in only a few colours.

* *China markers* can be used for outlining PSUs and strata on photography, images and acetate. "Berol Blaisdell" is one brand that has been used extensively. They are not to be used on maps. They are sensitive to heat and will smear on photos and acetate in hot weather. China markers would be used to provide additional colours that might not be available in the "all media" pencils or in the case that the "all media" pencils are not available.

* *Clear film*, such as 5 mil. clear acetate, and *drafting film* - mylar or polyester- will be needed in small quantities; perhaps one roll of each.

* A roll of quality *tracing paper* should be on hand.

* A *safety knife* for cutting cardboard backing, film, and paper should be obtained. The "Olfa-FL" with disposable blades has been used satisfactorily.

* Each worker need a *transparent plastic grid* for estimating areas on photos, maps and satellite images. It is convenient to make a grid for each of the scales being commonly used. Squares on the grid would be equal to one square kilometre with a number of the squares being divided into quarters or even tenths. The most convenient grid size for frame construction depends on the average size of PSUs being formed. In addition each enumerator should have a transparent grid fitted to the scale of the photo enlargement used during data collection.

* A *table*, one meter by two meters, should be provided for each worker on the frame. Two or three extra tables should be on hand to form larger areas for laying out stratum areas when numbering CU's.

* *Chairs* - each person working on frame construction should have a comfortable chair. They will be seated for hours doing tedious work. Providing quality chairs may be very important in maintaining team motivation.

* *Map storage cases* are needed for maps and satellite images.

* *Special flat storage* for mosaics will be needed if the decision is made to construct mosaics.

* *Standard filing cabinets* are needed for the frame records and extra contact prints. Cabinets with a lock could be used to store planimeters, lettering sets, pencils, and other supplies.

5.6 Office Space

It is best if the entire frame construction crew work together in one large, well lighted room. Windows and lots of daylight are important for visual interpretation of photos and images, measuring and distinguishing colours and shades of gray. Storage cabinets will be located along the inside walls of the room. The team leader, his assistant, and the construction supervisor can have desks in the same room. Make sure that each frame construction worker still has 15 square meters of space after accommodating these staff and additional items. This is just the space needed for frame construction. Desk space will be needed for those who do the office review of questionnaires, error printouts, preliminary tabulations and final output.

Allow for space and desks or tables for the computers and chairs for the computer operators. The computer group should be in separate office space from that of the frame construction crew.

Many of these items may be unnecessary if plotters, Geographic Information System software, digitizers, Computer Graphic Systems, Zoom Transfer Scopes, digital maps and other advanced equipment is available. In any event, if the basic concept and the equipment needed to carry it out are understood, new and more convenient procedures are readily adopted.

CHAPTER 6

CONSTRUCTION OF THE AREA SAMPLING FRAME: LAND-USE STRATA AND PSUs

In this chapter is assumed that the general planning of the *area sample survey* have been undertaken, as described in the previous chapter. Therefore, the basic characteristics and parameters of the area sample survey have been determined including: the geographic area of interest, the sample size, the criteria for defining the strata, the target size of PSUs and segments, as well as the survey variables, their required level of precision, the estimation methods, the data collection and the data processing procedures. Also it is assumed that the survey requirements in personnel, equipment, materials and budget have been established.

This chapter covers the techniques used for the subdivision of the total survey area into *strata* and each *stratum* into an ordered group of measured *PSUs*. The accumulated size (target number of segments) of the PSUs provides the size of each stratum and of the total frame.

6.1 General Procedure for Area Frame Construction with Full Cartographic Materials

In this section, fundamentals of the area frame construction are presented assuming an ideal situation exists when all cartographic materials, including maps, satellite images and aerial photography, can be employed. Maps form the accurate scale base for measurements, aerial photography and satellite images identify the physical boundaries and zones of different land-use strata. Current satellite images will show current land use. If there have been appreciable changes since the photography, it may be worth the expense to order the acquisition of new images covering the areas of change. Stratification and formation of PSUs can be conveniently done on the 1:100,000 scale TM/LANDSAT scenes. If the photography is fairly recent and readily available, it may be less expensive to construct photo-mosaics than to order the 1:100,000 images. If only large scale (1:250,000 or 1:500,000) images are available, stratification and formation of PSUs should be done on mosaics made from the contact prints. Strata and PSUs will be transferred to maps for measurement. The photography will provide more detail as needed for stratification and will be required for the subdivision of PSUs into segments. The photography will also furnish enlargements of the sample segments for use during data collection. The variations in area frame construction procedures needed for different available materials and combinations of cartographic materials will be covered in the following sections. The emphasis must always be on obtaining the most recent suitable materials possible for the frame.

Steps in Area Frame Construction

Step One	• Prepare the Frame Materials, Delineate Frame Limits, and Measure Total Frame Area
Step Two	• Delineate and Measure Areas Covered with Water (Lakes, Large Rivers, etc.), Heavy Forests, High Mountains, National Parks, Military Reserves, and other Non agricultural Land, except Urban Areas
Step Three	• Outline and Measure the Urban and Agro-Urban Strata
Step Four	• Delineate Strata
Step Five	• Review of Stratification
Step Six	• Transfer Strata Boundaries to Maps. Measure and Field Verify the Strata
Step Seven	• Construct PSUs
Step Eight	• Transfer PSUs to Maps and Order PSUs
Step Nine	• Measure Area of PSUs
Step Ten	• Assign Measures of Size to PSUs, Strata and to total Frame

Step One • Prepare the Frame Materials, Delineate Frame Limits, and Measure Total Frame Area

Prepare the frame material that will be used to provide the "view" of large areas for stratification purposes. At this point, the decision may have been made to construct simple photo-mosaics, use available orthophotomaps, use satellite images, use only maps, or to use a combination of any or all of these. If 1:50,000 scale maps are available, it is a good idea to organize the material in units coinciding with individual maps. These units are referred to as *map pages*, using the name and number of the map. Where maps are not available, the map pages can be the size of a 1:100,000 scale enlargement of a LANDSAT image or, in a few instances, the map page can *be* a 1:50,000 or larger scale orthophotomap. It is convenient to assign work on the basis of map pages. Stratification is undertaken one map page at a time. Once the material is ready, draw in the *regional and provincial or departmental boundaries.* The external boundaries that are also the frontier of the country must be drawn in exactly as they are shown on the official maps of the country. Many of these will be arbitrary straight lines that may follow map coordinates but can't be located on the ground without sophisticated instruments. Internal boundaries must be drawn in along physical terrain features using a "give

Figure 2. A TM/LANDSAT 5 scene with PSUs

stratum block. It is convenient to delineate the PSUs so that their boundaries do not cross those of the smallest administrative divisions of the country (e.g. districts) or large fields; cf. section 3.2.3. To achieve this, the district boundaries have to be transferred to the satellite images or photo-mosaics. The stratum color is used for delineating the PSUs in a particular stratum. PSUs must be complete within each level of estimation and not overlap into another region or into another province or department even if the latter are not currently the level for which estimates are required. This will facilitate the elaboration of special frames and the adaptation of the overall frame in the future, if it is decided to change the level of estimate from the region to the province. If the urban areas need to be sampled, begin PSU construction with the urban strata; go next to the most intensively cultivated stratum and proceed in order of intensity of use to the least intensive and finally the non-agricultural stratum. PSUs are outlined in the color of the stratum within which they are formed. As explained earlier, the target size of a PSU depends on the chosen target segment size for the stratum. Use grids to assist in the size approximation. Avoid forming overlarge or excessively small PSUs. The objective is that each PSU should reflect on a small scale what are seen to be the stratum characteristics. If the stratum is defined as being 50% to 75% cultivated, each PSU ideally should contain 50% to 75% cultivated land. All the PSUs formed in a stratum should be very similar in makeup. Do not create PSUs that are mostly pasture within an intensely cultivated stratum and do not create PSUs that are mostly cultivated land within a pasture stratum. If the inclination to do this arises, a review of the overall stratification is indicated.

As with previous steps, the newly formed PSUs are passed on to another team member for review. The check will be for physical boundaries, size, and makeup of the PSU in relation to others in the stratum. Checks are made, not to place blame on any one person for making mistakes, but to improve the final product of the team's labours; this attitude must be instilled in each member of the team. Teamwork is essential. Again, as in the case of the stratification, it is important to organize field checks of at least part of the PSUs, using Global Position Systems, if available.

Step Eight • Transfer PSUs to Maps and Order PSUs

In this step of the process, the PSUs are transferred to the maps. Again, the initial transfer is made with an ordinary soft (HB) pencil. After review of the transfer is complete, they are outlined on the map in the appropriate stratum color. (Working with maps alone, PSUs would first be formed with pencil and put on in color after checking). Each PSU is given an identification code consisting of three consecutive numbers. An example is 1-2-11. The first number refers to the stratum, the second number is the number of the PSU, and the final number is the area in square kilometres. Thus, the PSU in the example is the second PSU in the ordering sequence; it belongs to stratum 1 and has an area of 11 square kilometres. Numbering is done prior to measuring, so only the first two numbers of the code can be entered at that time. Numbers for the PSUs are sequenced in a serpentine manner beginning in the northeast corner and ending either in the southeast or southwest corner *without regard to strata*. The ordering could be carried out from east to west or west to east if that is seen as the best way to distribute the sample. The following example is set up to show serpentine ordering of PSUs and the way that PSUs often overlap various map pages. The dark squares are the map pages laid out as described above.

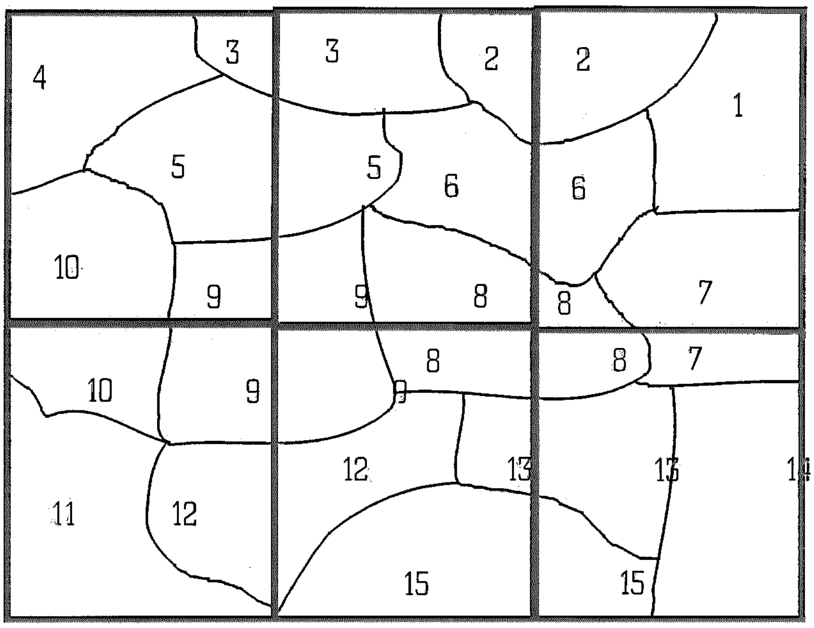

Figure 3. Order of PSUs

The order number of the PSU is the only number used in this example. The stratum number is missing but it has no effect on the ordering.

This is best done by laying out all the maps for a province or a region on a flat surface (or on the floor), joining them as closely as possible with the north of each map always at the top. The serpentine numbering system of the PSUs is applied for the entire province or estimating level, regardless of strata. The serpentine numbering system provides a north south distribution within each stratum that is important if a systematic sample is selected.

Step Nine • Measure Area of PSUs

Measurement of PSUs can be conveniently done with a computer graphic system or a digitizing table, especially for large areas involving a large number of PSUs when the use of planimeters becomes a burden. The results should be verified by measuring a sample of PSUs with a planimeter. Each PSU is measured in square kilometres to two decimal points. A listing or control sheet is prepared to properly record measurements and assist in determining the area to be accepted for the PSU, if several measures are taken.

If computerized measurement instruments are not available, and measurement of all PSUs have to be done with planimeters, it will be necessary in most cases to make overlays of images and orthophotomaps using a heavy grade quality tracing paper because planimeters are subject to error when used on slick surfaces of photo maps and images. Measurement will be done on the overlay. Measuring with planimeters on the maps is no problem except that the surface underneath must be absolutely smooth. Planimeters must be checked each day before beginning work by measuring a known area. Measurements should be within 5% of the control area. A first planimeter reading is obtained for each PSU on each map page. To make the "first reading", the technician takes two measurements of the PSU; if these two measurements are within 5% of each other, they are averaged to provide what is called the "first reading". If the first two measurements are not within 5% of each other, a third measurement will be made. If the third measurement is within 5% of *one* of the first two, the two closest measurements are averaged to form the "first reading". If the third measurement differs by more than 5% from the first two, repeat the entire sequence. If two measurements within 5% of each other are not achieved after repeating the sequence, the PSU is passed to another person using a different planimeter for measurement.

After "first readings" are obtained, the map page is passed on to another technician who measures the PSUs with another planimeter and record the measurements. The supervisor collects the first planimeter readings and second planimeter measurements and records them on the control form. For this to be a valid independent checking process, technicians making the second set of measurements must not have access to the first result. If the difference between the two first planimeter readings is 5% or less, the two are averaged to establish the area of the PSU. PSUs with a difference of more than 5% are sent to a third team member for planimeter reading. The third reading is conducted in the same manner as for the "first reading". If the third reading is within 5% of the first reading, their average is the PSU final area. If the third reading differs by more than 5% from the first but is within 5% of the second measurement, the average of the third reading and the second measurement is the final PSU area. With a difference of more than 5% between the third reading and the other two readings, the cycle is repeated from the beginning with completely new measurements until readings are achieved that conform to the guidelines. Differences should not exceed 5%. The area will be entered as the third number in the three number PSU identification code on the frame materials and the listing sheet. Note that the PSUs are grouped by stratum as they are measured. The segments in each stratum form a sampling population.

An example of a functional listing sheet for this purpose follows:

PRIMARY SAMPLING UNIT SIZE - CONTROL SHEET							
STRATUM: ______			PROVINCE: ______		DISTRICT: ______		
PSU #	Topographic Map Name	Map Number	Planimeter First Reading (Km^2)	Planimeter Second Reading (Km^2)	Diff. (%)	Planimeter Third Reading (Km^2)	FINAL Area (Km^2)

Table 1. PRIMARY SAMPLING UNIT SIZE - CONTROL SHEET

Areas of the PSUs can now be summed at the district and province levels and the totals checked against official area estimates. Any significant differences must be examined and justified or corrected before assigning the final area to PSUs.

Digitizing tables for measuring areas are now quite common and accessible. A frame map is fixed on a digitizing table and a reference point for X-Y coordinates established on the map. A technician goes around the perimeter of each PSU with an electronic device, registering coordinates as he goes; the more points registered, the better the approximation of the PSU area. The X-Y coordinates, tagged with the proper identification, uniquely describe the borders of a PSU so that a polygon can be created for each PSU. The digitizing software records the coordinates in a file which is used in conjunction with the map scale to calculate the area of each PSU in the desired unit of measurement. Fewer people will be involved for measurement as those required with planimeters, and verification and checking is done by error-check programs in the computer and by overlaying original frame maps with plotted frame maps on a light table. The succeeding steps in sample preparation are carried out by the computer program when strata and PSU areas are digitized.

Step Ten • Assign Measures of Size to PSUs, Strata and to total Frame

As PSU areas are determined and approved, they are entered on another listing sheet by stratum and district that shows map identification, PSU number, PSU area in square kilometres, segments assigned, accumulated number of sample segments, and up to ten columns on the right hand side for the selection of replicates, as shown in Table 3 in Section 7.1 -Sample Selection Control Sheet. The PSU area is rounded to the nearest whole square kilometre using standard rounding rules when it is transferred to this final work sheet.

Next, segments are assigned to the PSUs as determined by the area of the PSU and the *target size* of a segment in that stratum. The number of assigned segments (the measure of size of the PSU) is equal to the area of the PSU divided by the target size of the segment in the stratum, and the result approximated to the nearest integer. The measure of size of the PSU should be an integer. For example, if segments of one half square kilometre are desired, a PSU with an area of 5 square kilometres would be assigned 10 segments. Use standard rounding rules to assign an integer measure of size to each PSU within a strata. When measures of size have been assigned to all PSUs, the cumulative column in the listing sheet shown in Table 3 can be completed. The frame is now ready for sample selection.

6.2 Area Frame Construction with Maps and Aerial Photography. Construction of a Photo-Mosaic

In the past maps an aerial photography were the common set of materials used for area frame construction. Satellite images did not exist or were not readily available. Even today aerial photos on a scale 1:50,000 or larger generally provide better recognizable *boundaries* for area frame construction of strata and PSUs than all types of existing satellite images. And have generally less percentage of cloud cover.

Maps are used as the measurable base of the frame. The photography makes it possible to see the terrain features to establish good boundaries and thereby to classify most of the land into strata and construct PSUs with a minimum of field work. The assumption in this section is that the photography is available for large areas only in contact prints on a scale of about 1:50,000. If photography is available as orthophotomaps, the following discussion on the formation and use of mosaics is not applicable since stratification and formation of PSUs would be done directly on the orthophotomaps.

Being able to see large areas with the photography facilitates delimitation of strata and PSUs. The contact prints can be laid out in a loose arrangement in order to visualize an area the size of a 1:50,000 scale topographic map. However, prints will become damaged and/or lost in the daily laying out and gathering up that will take place in the stratification, transferring, forming PSUs, and transferring again. Experience has shown that it is practical to construct a permanent photo-mosaic of the contact prints.

Photo-Mosaic Construction. A good grade of cardboard about 2 to 3 millimetres in thickness is required to be used as backing. Sheets about one and one half meters by two meters are a convenient size. Unfortunately, this is not a common size so it may be necessary to join two sheets with masking tape. The joint will be sufficiently strong after the contacts are glued over it. Reinforce the edges of the cardboard with masking tape. Next, lay out the contact prints on the backing so as to provide visual coverage for the ground area of a 1:50,000 scale map. Only simple coverage is needed. Allow an additional row of prints along each edge extending beyond the boundary of the map. Arrange the photos by flight line with only a minimum of overlap - like tiles on a roof. The second row extends slightly over the edge of the first row, etc. Orient the mosaic with the top of the photos toward the north. After the photographs are laid out, the flight number, photograph numbers, scale, and date are recorded on the margin of the mosaic. Each photo must have its individual data recorded on the back. Trim the east, west, and, south margins off the photos. The top or north margin can be left on. Work with the contact prints until you are satisfied that there is complete coverage of the map plus one extra row of prints on each side and that no spaces are left uncovered. The prints can now be glued in place. Rubber cement should be used. It dries rapidly but allows some time for adjustment. It will not wrinkle the prints. If the cement is applied correctly, prints can be taken up if necessary without excessive damage. Glue is applied in about a two centimeters band along the top margin of the back of each photo. Prints are glued only at the top so they can be lifted (the scene on the bottom may be better than the one on the upper print). Begin by gluing down the base print. The base print is the lower right hand print. Proceed from right to left and from bottom to top. After gluing the base print firmly in place, lay the next print to the left (without gluing) in the best possible match. Distortion along the edges and changes in altitude between flight lines will make it impossible to exactly match the features between prints. Choose an approximate match that gives the best continuation of the scene from side-to-side and top-to-bottom. When the best match is made, mark the location of the second print along the edges and corners with a pencil on the cardboard backing. This will make it easier to position after applying the glue. Continue in this manner until the first line is glued from right to left. Then go to the second line, working from the right to the left and so on until the mosaic is complete.

A standard procedure for recording needed information on a mosaic could be as follows:

Left margin:	an arrow indicating north.
Upper left corner:	the state, department or province where mosaic is located.
Central upper margin:	name and number of the topographic map corresponding to the mosaic.
Right margin:	scale and date of the photography, with number of the flight line and each photo at the end of its corresponding line.
Lower right corner:	initials of person constructing the mosaic.

In most cases it will be possible to use prints of the same scale within each mosaic. Again, prints of 1:30,000 scale are most convenient. With 1:20,000 scale, a mosaic covering the area of a 1:50,000 map is larger than would be convenient but still can be used. Scales of 1:40,000 to 1:60,000 require more use of magnifying glasses to identify boundaries, but can be used and will provide more visible detail than most satellite images.

6.3 Area Frame Construction with Maps Alone

Area frames using physical boundaries for strata, PSUs and sample segments have been constructed where maps were practically the only material available. This is not advisable and is seldom necessary today. It is hard to find sufficient physical boundaries on maps, this resulting in the formation of PSUs and segments larger than desired and a significantly larger amount of field work required for stratification and division of selected PSUs, since many of the boundaries will have to be defined by field work. Accurate scale maps of segments will be required for data collection if the closed segment or weighted segment estimators are used. Non sampling errors will be high due to boundaries that are imprecisely identified.

An advantage is that no transfer from one material to another are required. Maps have been employed with some success where only the open segment estimator is used. If there is no alternative, be sure the most recent maps are obtained. Use maps of 1:25,000 scale whenever possible, with total coverage in 1:50,000 and use 1:5,000 or 1:10,000 scale maps for cities and villages.

To make the stratification, field experts familiar with the areas must be brought in so as to locate or confirm stratum boundaries. After approximate stratification is achieved, the maps are taken to the field for verification.

In areas with scarce physical boundaries where maps are the only available material, two solutions have been used to define appropriate PSUs: the use of point sampling methods, as described in Section 7.4, or the use of square segments using grids, examples of which are included in Volume II.

6.4 Area Frame Construction with Aerial Photography Alone

The availability of aerial photography without maps is an unlikely occurrence at a country level but it may well be true for certain portions of the area for which a frame is to be constructed. Where such is the case, an acceptable map can be prepared from a carefully constructed photo-mosaic. The mosaic must be assembled with particular care matching terrain features with no overlap. After all mosaics for the area in question are ready, a firm boundary for the area covered by each mosaic must be established so that the area in question is revealed without omission or overlap. The mosaics must also fit precisely with adjoining map-based mosaics. Stratification and formation of PSUs is undertaken after the outside boundaries are placed on the mosaics. An overlay is then made of the mosaic using with tracing paper showing the stratification, PSUs, outside boundaries, and perhaps a few prominent features such as cities, major rivers, and major roads. Measurements are then made on the overlay. Such maps are not exactly to scale but they can be used with satisfactory results where necessary.

If the photography is in the form of orthophotomaps, there is no problem. They can be used both for boundary identification and measuring since they are corrected to scale.

6.5 Use of Satellite Images and Satellite Photos

Satellite Images in a 1:100,000 scale (LANDSAT or SPOT) can be most conveniently used to delineate strata and PSUs. The use of satellite images constitutes a very important improvement and simplification for area frame construction. It avoids the laborious and meticulous work involved in the construction of photo-mosaics, and allows for more precise area measurements. Also, the acquisition of current satellite images of a given large area is much cheaper than obtaining current aerial photos. Furthermore, satellite images that used are generally more recent than available aerial photography providing a more updated cartographic base for the frame. Satellite images, however, cannot be utilized in areas with frequent cloud cover, which is less of a problem for aerial photography.

Naturally, neither aerial photos nor images can be utilize to recognize crop areas under trees.

Satellite images are not usually suitable for subdividing PSUs into segments since they provide imprecise boundaries. The best solution is to commission new aerial photography for the selected PSUs. The photography provides the detail needed to subdivide the PSU into segments and also the enlargement of the selected segment for survey data collection. This photography is best carried out by an aerial photography company but it can be taken from a light plane and a relatively unsophisticated camera, taking care about focal length of the camera lens, camera stability and altitude of the plane from the ground surface.

Satellite images cannot efficiently substitute the aerial photographic enlargement of sample segments used for data collection, requiring a scale of 1:5,000; at that scale, 2 centimeters on the photo equals 100 meters on the ground making it feasible to mark off half hectare fields. Currently satellite images available cannot be effectively enlarged to a scale larger than 1:50,000 far from the 1:5,000 scale required for such purpose.

Future sensors aboard satellites, having higher resolution on the ground than the MSS (Multispectral Scanner) aboard the LANDSAT V (80 meters), the TM (Thematic Mapper) sensor of the LANDSAT V (30 meters) or the XS multispectral sensor (20 meters) or panchromatic sensor (10 meters) aboard the SPOT 3, may provide images at such level of resolution that they will be used to subdivide PSUs and even provide segment enlargements for enumeration. The SPOT 4 will be launched in 1997; and the LANDSAT VII is to be launched in 1998 with the 15 m panchromatic band as an important new feature. The IRS-1c satellite (Indian Remote Sensing) will be launched in 1995 equipped with a high resolution (better than 10 meters) panchromatic sensor.

If the only available satellite images are MSS/LANDSAT in 1:250,000 scale - which is not very likely to be the case-, then only the strata can be delineated on imagery. It would not be practical to try to form PSUs. The strata would have to be measured on the imagery (actually on a tracing of the image). Random points within each stratum could be used to locate sample segments. Then, it may be possible to photograph the area around each selected point and form a PSU from which a sample segment is chosen at random. It is, of course, possible to form only a sample segment around the random point so requiring less photography but risking bias into the formation of the segment. Stringent rules are necessary when forming single segments around random points. Development of hand held Global Positioning System instruments and devices for measuring distances have made the use of random points (point sampling) more precise.

There are different types of satellite images obtained from different sensors aboard satellites. Satellite data can be obtained on paper prints (like maps) in false color or in digital form for display in a computer screen. The selection of the appropriate imagery for the stratification of an area frame must take into account the appropriate combination of bands, geometric corrections and acquisition dates.

Satellite Photos in scales 1:50,000 or larger can in principle be used for construction of strata and PSUs. The problem is that large scale space photos are not readily available for large areas. For instance, the space photos taken from the series of satellites SOYUZ, that meet such requirement and have been evaluated, did not deliver with any reliability which is an important factor when constructing an area frame.

Also, a number of large scale space photos that could be used to subdivide at least some PSUs and as substitute of the corresponding aerial segment photo enlargements, are either classified for military reasons or too expensive to acquire for a large area.

CHAPTER 7

SELECTION OF A REPLICATED SAMPLE OF SEGMENTS

Following the procedures outlined in earlier chapters, the total frame, the strata and PSUs have been delineated and assigned a measure of size, and the PSUs will have been ordered following geographic criteria of contiguity or ordered of similarity of agricultural characteristics. Assume that for a given province the sample size (n) has been allocated to the land-use strata (size n_h in stratum h), and that the number of replicates (r_h in stratum h) has been determined.

This chapter deals with the allocation of the sample to the strata, the definition of *substrata* or *zones* within strata, and the selection of a sample of segments within each stratum using a random or systematic, replicated selection procedure. In addition, a point sampling selection method is outlined to be used in strata with inadequate physical boundaries.

The allocated sample size in the strata, n_h , should be adjusted so that n_h is proportional to r_h, the number of replicates. L_h will denote the number of substrata in land-use stratum h.

Before beginning sample selection in a *stratum*, the first six columns of a listing sheet such as an example shown in Table 2 (Accumulated Sample Unit and Sample Selection Control Sheet) must be completed for each stratum from data on the PSU Size Control Sheet (see Table 1). In other words, for each stratum a list should be prepared with the accumulated measures of size of the PSUs ordered following the procedures already indicated.

Each stratum should be completely subdivided into non-overlapping substrata of equal sizes called *zones*. The number of replicates is dictated by the desired sample rotation procedure. For a given land-use stratum, the *number of zones* formed should equal *the sample size for each replicate*. For example, if 5 sample replicates are selected, $r_h = 5$, then $L_h = n_h/5$.

Within each stratum, substrata of equal number of segments (except the last one of each stratum that could have a slightly different number of segments) are formed by accumulating PSUs following their established sequential order of similarity of agricultural characteristics.

For each zone, the determination of the selected PSUs and the number of sample segments per PSU is accomplished by using the accumulated measures of size of the PSUs (following the established order). Then, each selected PSU is subdivided into a number of segments equal to its measure of size. In each zone, segments are selected with equal probability and without replacement (EPSEM) contributing one segment to each of the r_h replicas of the stratum.

Systematic and random replicated selection procedures are described. For each stratum, the practical procedures for sample selection involve two steps:

* Selection of PSUs with Probability Proportional to Size Measures (PPS)

* Selection of a Segment within selected PSUs, with Equal Probability (EPSEM)

7.1 First Step: Selection of PSUs with Probability Proportional to Size Measures (PPS)

Systematic Replicated Selection Procedure

For systematic sampling, the zones (with equal number of segments) are determined by the subdivisions created by the application of the sampling interval in the stratum following the adopted order of PSUs as shown in the above-mentioned listing sheet (Table 2) with the accumulated measures of size of all PSUs in the stratum. The definition of the given number of equal size *zones* (substrata) within a strata depends therefore on the established order of PSUs.

Example 1. Assume for stratum 1 that N=990, n=60, that the desired sample rotation is 25% each year, and that r=4 replicates in the active sample will be selected. Each replicate will consist of samples of n=15 segments (60/4) selected independently. The sampling interval becomes I=990/15=66. If I is not an integer the procedure is still valid.

Select a random number between 1 and 66 from a random number table for each replicate to be selected. Each replicate must have a different random start. If a random number that duplicates one previously chosen it is rejected and another selected until there is a unique random start for each replicate. Since the active sample is to made up of 4 replicates, it is advisable to select 8 (or even 10) replicates in the first selection process. This provides enough extra replicates for rotation for 4 or 5 years without new selection plus a pool for an increase of the sample size if that need is indicated by the result of the first survey. The eight random starts will be entered in columns 1 through 8 on the control sheet under the heading "Replicates". They are entered on the line occupied by the PSU that contains the indicated segment. A random start of 21 for replicate 1 would be entered on the line with the PSU that contains segment 21. The remaining segments in each replicate are selected by adding the I successively such that the next segment in replicate 1 is segment 87 (21+66), which is entered on the line occupied by the PSU containing segment 87. Next segment 153 (87+66) is selected and so on. This procedure is carried out for each replicate. The result is that the PSUs that contain sample segments are identified on the control sheet. In the above example, the first substratum in a given *stratum* will contain segments 1 through 66, the second substratum contains segments 67 through 133 and so on, forming 15 substrata. Each substratum includes one sample segment of each replicate. The serpentine ordering of PSUs imposes a geographical distribution on the systematic sample.

Example 2. Table 2 shows the "Accumulated Sample Unit and Sample Selection Control Sheet" for an imaginary population. In the example shown, N = 216 and n=18 for the stratum illustrated, and r=3 replicates are selected systematically. The total sample size of n=18 means that each replicate will have 6 sample units (18/3=6). The sampling interval for each replicate is I=216/6=36. Different random number between 1 and 36 are selected for each replicate, namely: 7, 17, and 34 respectively. The number 7 is entered in the column for replicate 1 on the line of the and take" process to keep overall areas approximately correct. Scaling rulers are an important asset for locating boundaries. Points and features that are common to both materials such as where a road crosses a river, crossroads, railroad crossings, a road or a river, a school, a cemetery must be found. Then, with measurements from known points, a boundary can be transferred from one to the other. Occasionally, notable differences will be found between the photography or images and the maps. In this case, the photograph or image must be accepted as correct. Drawing on photography or images is to be done with grease pencil or special soft lead pencils. Red is a good color to use for the boundaries of the area that defines the level of estimate; usually a region. Grease pencils and soft lead pencils are *not* appropriate for use on maps.

Step Two • **Delineate and Measure Areas Covered with Water (Lakes, Large Rivers, etc.), Heavy Forests, High Mountains, National Parks, Military Reserves, and other Non agricultural Land, except Urban Areas**

This step follows on from Step 1 using the same procedures as described above.

Step Three • **Outline and Measure the Urban and Agro-Urban Strata**

Classify the central (commercial) portion of the large cities (areas with high population density) as non-agricultural. If a city is known to be growing, move the urban stratum line out to form an allowance for future expansion. The width of this band can be estimated by comparing older with newer images or photos and by taking into account growth projections. It may be that the most rapid, or perhaps all, growth will take place in specific zones around the city. Incorporate this knowledge into the formation of the expansion allowances. These expansion areas can cause an increase in variance now but will extend the life of the frame. The agro-urban stratum which is an area with high population density that also includes patches of agriculture, is also delineated at this time.

Step Four • **Delineate Strata**

The *strata*, in agricultural areas, are defined by proportion of cultivated land, predominance of certain crops, average size of cultivated fields and special sites of agricultural activities.

Frame construction personnel must memorize the strata definitions and must spend time working together with the images and photos to ensure that they all agree on the appearance of each stratum on the material. Decide on the color to be used for each stratum and review the rules for drawing lines on the material. The most intensively cultivated strata will be considered the higher level strata so that the most intensive stratum is stratum 1, the next most intensive is stratum 2 and so on. The line between two strata will be drawn in the color of the higher level (most intensive) stratum. The number of each stratum will be written in their respective stratum blocks in the appropriate color. Make use of all sources of information in identifying strata and strata boundaries. Prepare a wall board with the definitions of the strata and the minimum block size for easy of reference and coordination of the team work.

Begin, *not by drawing lines*, but by carefully studying the map page or map page equivalent and identifying blocks of the various strata; writing in the proper number in the proper color in the identified blocks. If adjoining map pages have been stratified, mark the entry point of all strata that extend into the map page where stratification is just beginning. After as many blocks as possible have been located, identify the physical terrain features that will serve as boundaries and draw in the lines enclosing each block. Draw the boundaries for stratum 1 (most intensive) first. Proceed to strata with less proportion of cultivated land then pasture and, finally, outline the non-agricultural stratum.

Step Five • Review of Stratification

The stratified map page will be passed on to another team member for review. Differences in interpretation or ambiguities will be settled between the reviewer and the person in charge of stratification in conference with the supervisor. The proper continuation of strata boundaries from one map page to another is the first check the reviewer should make. Land classification and the remaining strata boundaries are checked next. The stratification on a map page is not considered complete until adjoining map pages have also been stratified and checked.

Step Six • Transfer Strata Boundaries to Maps. Measure and Field Verify the Strata

The reviewed strata boundaries are now transferred from satellite images or photo mosaics to maps. The initial transfer will be made using an ordinary soft lead pencil. After this initial transfer has been reviewed and the boundaries confirmed, coloured pencils (erasable because it may be necessary to modify the strata boundaries using information from the map or to better conform with PSU make-up and boundaries) are used to place the final lines upon the map. The base lines will be the color of the higher level stratum but, the limits of each block should be lightly shaded on the inside with the stratum color. In this way, blocks can be readily identified. If maps or orthophotomaps are the only material being used, no transfer is necessary. Transfer of boundaries and area measurement of strata can be effectively completed by using a computer graphic system that allows measurement of the stratum blocks before the transfer to maps.

The stratification should also be confirmed or corrected in the office by field experts with detailed knowledge of the areas.

At this point the strata should be measured and the total area compared with available data. Then, it is important to verify the stratification in the field and implement the appropriate corrections. For this purpose, in addition to the maps, images and photos available, portable Global Position Systems will be used, if available. Then, following any corrections as a result of field verification, strata are measured again.

Step Seven • Construct PSUs

Returning now to the images or photography, PSUs are constructed within each

ACCUMULATED SAMPLE UNIT AND SAMPLE SELECTION CONTROL SHEET															
STRATUM: .... DISTRICT: .... PROVINCE:															
Target Size of Segment:															
Topographic Map Name and Number	Map#	PSU Order Number	Area Km^2	Measure of size	Accumulated MOS	REPLICATES									
						1	2	3	4	5	6	7	8	9	10
Manyu-5937	4	1	5	10	10	7									
"	"	2	7	14	24		17								
"	"	3	4	8	32										
"	"	4	9	18	36 50	43		34							
"	"	5	6	12	62		53								
"	"	6	8	16	72 78			70							
Ndian-5836	3	7	5	10	88	79									
"	"	8	5	10	98		89								
"	"	9	7	14	108 112			106							
"	"	10	4	8	120	115									
Meme-5837	3	11	6	12	132		125								
"	"	12	9	18	144 150			142							
"	"	13	3	6	156	151									
"	"	14	8	16	172		161								
"	"	15	7	14	180 186			178							
Fako-5836	2	16	5	10	196	187									
"	"	17	4	8	204		197								
"	"	18	6	12	216 216			214							

Table 2. Accumulated Sample Unit and Sample Selection Control Sheet

PSU containing segment number 7; the number 17 is entered in the column for replicate 2 on the line of the PSU containing segment number 17, and so on. Succeeding segments in each replicate are selected by adding the sampling interval (36).

Random Replicated Selection Procedure

Substrata are formed as described above, adjusting PSUs so that there is the same number of segments in each zone. Then, segments are selected at random within each substratum using a table of random numbers *without* replacement.

Example 3. Assume that, as in the Example 1 above, for stratum 1, N=990, n=60, and r=4. Each replicate will consist of samples of n = 15 segments selected independently within each substratum. Show the limit of each *substratum* on the control sheet with a horizontal line under the PSU that contains the last segment of the substratum, and show the number of the last segment (number #66 in the example, since 66=990/15) above the line in the upper right hand corner of the cell of the accumulative column of the number that contains the limit of the substratum. Eight to ten replicates can be obtained on one trip through the random number table. If 9 is the first number encountered within the range of 1 to N = 990 (total number of segments in the stratum), 9 is entered on the line of the PSU that contains segment 9 and in the column for *replicate 1*. If the next random number falls also in the first substratum; say 51, enter 51 on the line of the PSU that contains segment 51 in the column for *replicate 2*. In other words, the first random choice *within* a substratum will be for replicate 1, the second random choice within the substratum is for replicate 2, the third for replicate 3, etc. If the next random number after 51 in this example is 81, it would be entered on the line for the PSU containing segment number 81 in the column for the first replicate, since it is the first segment selected from the second substratum. Once the limits of the substrata are established, each random number within the population range can be assigned to a substratum and a replicate. *Numbers previously chosen and numbers outside the population limits are skipped over.* Selection continues until the required sample size is acquired for each replicate in each substratum.

Example 4. Assume that, as in the Example 2 above, for stratum 1, N=216, n=18, and r=3. Each replicate will consist of samples of n=6 segments selected independently within each substratum. For a random selection within substratum, the boundaries of the substratum can be marked off on the control sheet. There are 6 substrata containing 36 segments each. The first substratum is made up of segments 1 through 36. The lower limit of the first substrata would be indicated by writing in the number 36 in the upper right hand corner of the cell in the column of accumulated sample units containing the number 50. PSU number 4 contains segments 33 through 50. The lower limit of the next substratum would be at segment 72 (36 + 36) and would be indicated by writing the number 72 in the upper right hand corner of the cell in the accumulative column containing the number 78 followed by 108 written in the corner of the cell containing 112; 144 in the cell containing 150; 180 in the cell containing 186; and 216 in the cell containing 216. Random sample selection could be done using three columns of a random number table, accepting only numbers from 1 to 216. The first number encountered within the range of a given substratum is assigned to replicate 1, the second number to be found in that same range is assigned to replicate 2 and so on until 3 segments are selected for each substratum, making the sample of n = 18 complete.

7.2 Second Step: Selection of a Segment within a Selected PSU, with Equal Probability (EPSEM)

The previous examples have shown how the PSU containing a sample segment is identified. The next step is to subdivide the indicated PSUs into their allotted number of segments and select one of those segments at random. To do this, one team member subdivides the selected PSU into the assigned number of segments with another team member checking the work. When the two team members agree that the subdivision of the PSU is satisfactory, each segment is assigned a number. Then a random number is selected by the supervisor which designates the sample segment. The supervisor records the number used for the PSU. The random numbers should be selected by the supervisor to avoid intentional or unintentional bias in the selection process by those who subdivided the PSU into segments. The subdivision of a PSU into segments is usually done in yellow: first on the mosaic or on other photography, and then transferred to a topographic map. Selected sample segments are given a three-number code in red that reflects the substrata, replicate and segment order number. For example, the third segment selected in the second replicate of stratum 1 might be numbered 1-2-3.

The subdivision of PSUs influences the efficiency of data collection and, thereby, the quality of survey results. Each potential sample segment should be a small representation of the PSU; containing the same proportion of cultivated land, pasture, occupied dwellings etc., as the PSU itself. Variation between segments formed within the PSU must be minimized. A subdivision where one or two segments contain most of the occupied dwellings or where some segments are 100% cultivated in a stratum that is classified as only 50% cultivated should be avoided. It is also important to use clear visible boundaries when forming the segments. Ambiguous boundaries require excessive time to identify the segment in the field and will cause errors to be made in segment identification, thereby increasing non sampling error.

If there are not enough clear visible boundaries to allow the PSU to be subdivided into the assigned number of segments, outline as many individual segments as possible of the proper size. The remaining number of segments is assigned to the undivided portion (block) or portions. If the random number selects one of the individual segments of the proper size, it becomes the sample segment for the PSU. If the random number indicates a segment in an undivided portion of the PSU, it will be necessary to go to the field and subdivide the block into the assigned number of segments and then select one of these at random to be the sample segment. If it is impossible or impractical to subdivide the block in the field, use the entire block as the sample segment and divide the survey result by the number of segments assigned to the block. This reduction factor can be a permanent portion of that segment's record and applied automatically by the computer at summarization. Or, it could occur that one or two individual segments can be defined in the field, still leaving a block equivalent to two or more segments that cannot be divided. Again, a random selection is made in the hope that an individual segment may be chosen. If it is not, the block is used as the segment and treated as described above. Always work toward the definition of segments as many segments of the desired target size. The undivided blocks require more time for enumeration and may present other problems of definition.

An example follows showing a PSU with an area equivalent to three segments (4-5-6) that could not be divided on the photo for lack of proper boundaries.

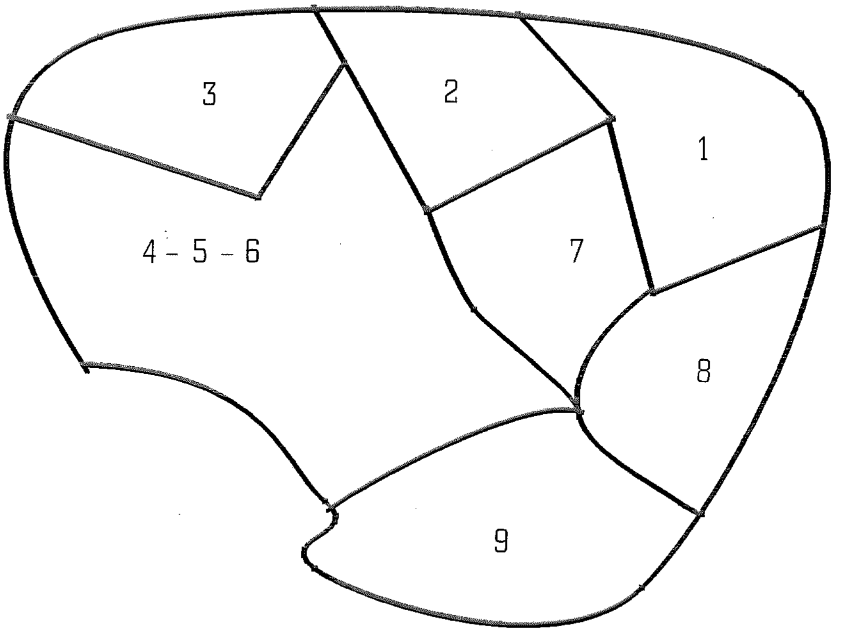

Figure 4. PSU subdivided into six segments and an area corresponding to three segments

A random number from 1 to 9 would be chosen to select the sample segment. If 7 is selected, segment number 7 is used as the sample segment. If 3, 4, or 5 are selected, the solutions stated above on dealing with the selection of an undivided block of segments are applied.

accurate way possible to determine scale of segment enlargements. Measure one or more centrally located large fields that can be identified on the ground and on the enlargement. Measure with a planimeter the same areas on the enlargement in square centimeters. The scale is determined by the formula in item 2. above.

7.3.2 Review of Segment Enlargements

A person experienced both in the field and in working with photography should review all the segment enlargements before they are approved for field use. The purpose of this review is to pick out segments with boundaries that will be difficult to locate, and to identify segments that probably have too many tracts. Subdivide as many as possible of the segments identified as having too many tracts, working on the enlargements in the office. As a general rule, any segment with 20 tracts or more should be subdivided. The objective is to have about 10 tracts in each portion. A segment can be divided into 2, 3, 4 or even more parts if that is necessary to attain the objective. The rules for dividing a segment are the same as those for dividing a PSU, i.e. endeavour to make each portion a smaller representation of the original segment. The problem is to subdivide a normal sized segment that evidently has a large number of tracts and would take an excessive amount of time to enumerate. This occurs in localities containing large numbers of tiny individual plots (minifundia). Data collected is weighted by the factor of subdivision before being entered into the summary system. Data from a sample unit that is one third of the original segment is multiplied by three and so on. It is advisable to resolve the problems of poor boundaries and an excessive number of tracts prior to the survey data collection, by means of field investigation for the cases that cannot be resolved in the office. When this cannot be done, the problem segments are marked for the attention of the supervisor at the outset of the first survey round.

7.4 Point Sampling Selection for Strata with Scarce Physical Boundaries

This type of sampling consists of selecting random points on a map or image within a specified area. It is most often done by imposing a grid with X-Y coordinates over the area of interest and selecting the desired number of sample points on the grid using a random number table. The points are marked on the map or image. Point sampling is most often used in strata where land use is not intensive (range or pasture land) and acceptable boundaries for the formation of PSUs are not readily available. The sample unit is defined by finding the nearest set of boundaries that encompasses the point. In an extensive land-use stratum "nearest" can mean within a number of kilometres; forming segments containing 30 or 40 and up to 100 square kilometres or more. Each of these large and highly variable segments will have a different probability of selection (proportional to size).

Point sampling can be used even in intensive land-use strata with PSUs (preferably) or individual segments being formed around the selected point. This may be done in any case when the material is adequate for land-use stratification but not good enough to form PSUs. The area around the point could be photographed. A PSU or a single segment can be formed on the photography. Forming a PSU and selecting a sample segment at random as described in sections 7.1 and 7.2 ensures the most unbiased construction and selection of sample units. If photography is impractical, the point must be exactly located on the ground. Hand held Global Positioning System instruments offer a practical means of locating points on topographic maps or satellite images. Otherwise, one must work with distances and direction from identifiable landmarks which usually requires a considerable amount of time and walking.

In Pakistan, segments have been formed around a point in intensively cultivated land-use strata without the use of photography. After locating the point, one way to define the segment is to determine the radius of a circle required to encompass a segment of the desired size. A strong rope of this length is then used to circumscribe a circle around the point. A scale drawing of the circle can be made on grid paper showing (with considerable internal measuring) field boundaries and prominent terrain features. A previously stated rule is that segment boundaries must follow physical ground features and must not divide fields. To satisfy this rule, fields with more than half their area *inside* the circle are included in the segment; fields with more than half their area *outside* the circle are excluded. The circle is *not* the segment boundary but is used to indicate the physical features that define the segment. A permanent drawing of the segment from the field sketch must be made for survey use. This procedure takes about four times as long to form a segment compared with the use of photography.

Figure 5. A TM/LANDSAT scene with a PSU broken up into segments.

7.3 Segment Photo Enlargements or Scale Drawings to Assist Data Collection

7.3.1 Determining the Scale and Securing Segment Photo Enlargements or Scale Drawings

Enlargements or drawings with a north arrow and segment identification number must be available for all segments in the sample prior to the survey field data collection commencing. The scale should be established and shown in one corner of the photo or drawing. Field review of problem segments, subdivision of segments with too many tracts, and scale determination must be completed well ahead of the survey. The area of each segment must be measured and noted in office records. It is usually not advisable to give the segment areas to enumerators, although they must be aware of the approximate scale of the enlargement for locating reference points and new field boundaries.

Photo Enlargements. Where photography is available, segment enlargements can be ordered for the sample segments. Enlargement scales ranging from 1:3,000 to 1:10,000 have been used to control segment enumeration; the most common, and possibly the most practical, being the scale of 1:5,000. Most practical because many of the photography scales are a multiple of 5,000. There is no absolute rule about the scale to be used; the basic need is for the enumerator to be able to delineate the fields. If fields are small (one half hectare), a 1:3,000 or 1:5,000 scale will be best, but if fields are consistently large, 1:10,000 scale enlargements can be used. In any event, it is advisable that all segment enlargements should be the same size, so that storage cabinets and carrying cases can be procured. The size of the photo enlargement must be convenient for the enumerator to carry and manipulate. Another consideration is the size of individual sheets of enlargement paper which usually comes in rolls about 1 meter in width and can be cut to many different sizes. The point is to have a standard size. This can be achieved by using a frame of plastic or cardboard to outline the area on the negative (preferably) or contact print that will be enlarged to, say, 60 cm x 60 cm and provide a scale of 1:5,000. A frame 15 x 15 cm would serve for contacts at 1:20,000. Contacts of 1:30,000 would require a frame of 10 x 10 cm.

The segment enlargement should be ordered from the contact print on which the segment is nearest the centre of the photo. The centre of the segment should be placed near the centre of the frame in order to include as many landmarks as possible around the segment that will help to locate and define it in the field. The area encompassed by the frame is then drawn off on the negative or contact with a red grease pencil. It is probably best to order enlargements in small lots, particularly at first until the quality and timeliness of the work is assured. Technicians will outline the sample segments on the enlargements with a soft (HB) pencil. After the transfer is approved by the supervisor, the segment is outlined in red ink with the ink line just outside but touching the physical feature that serves as the sample unit boundary. Ink is used because it is more durable and to prevent changes by field personnel. An arrow indicating north should be placed in the upper right hand corner of the enlargement, as mentioned. A segment identification number showing region, stratum, replicate and segment order number can be placed in the upper left hand corner. An eight digit code allowing two digits for each classification provides a unique identification in a system that will last for 99 years if one replicate is rotated out each year. The segment boundary should also be indicated on a transparent overlay of the segment photo enlargement. The enumerator should indicate tract and field boundaries and codes only on the transparency. This will prevent damage from repeated use of the photo and will allow comparisons between different survey rounds.

Scale of Segment Enlargements. A reasonably accurate measurement of the segment enlargement is needed in order to be most useful in the control of data gathering. Measurements are dependent on knowing the scale of the enlargement. Unfortunately, the scale of the contact prints is only approximate and will vary between prints as the altitude of the land and the plane vary. Also, the enlargement process may not have been be too carefully controlled, resulting in varying degrees of magnification between batches of enlargements. Rectified (true to scale) photography solves these problems, but it is not usually available. An alternative is to establish scale by means of ground measurement using one of the following procedures:

1. Find two points that can be identified on the enlargement and on the ground. The distance between the two points on the ground must be at least 200 meters, assuming an approximate enlargement scale of 1:5,000. A distance of 100 meters may be satisfactory where the approximate scale of the enlargement is 1:2,000 or 1:2,500. The points should be within the segment and near the centre. The two points and the line between should be at the average elevation of the segment (definitely not at the lowest or highest point). Measure the distance on the ground in meters and on the enlargement in millimetres. The scale can be calculated by the formula shown below.

$$Scale = 1 : \frac{Ground\ distance\ -\ meters}{Photo\ distance\ -\ millimeters} X\,1000$$

For example, if the distance between the two points on the ground is 213 meters and the distance on the photo is 35 millimetres, the scale would be:

$$1 : \frac{213}{35} X\,1000 = 1 : 6086$$

Planimeters can be adjusted to this scale to give the proper area readings for the segment. This is the fastest and simplest way to establish a scale for an enlargement. If applied with care, results should be acceptable.

2. A more accurate method is to use a triangle the area of which is calculated and related to the same area measured on the photo. The triangle is defined by points that can be identified on the ground and on the photo. The triangle should be located near the centre of the segment and at about the average elevation of the land within the segment. Construct the triangle as nearly equilateral as possible with sides of at least 200 meters if the approximate enlargement scale is 1:5,000.

a) Identify the points that define the triangle in the field on the photograph. Circle each object with red grease pencil and label the points A, B, and C. It is important that the field measurement between the points be done on a straight line. Where a natural straight line such as a road, fence line or field boundary does not exist, stretch a strong string or small rope tightly between the two points and measure along the line it indicates to the nearest centimetre.

b) Draw straight lines on the photo between points A and B, B and C and C and A. Measure the resulting triangle with a planimeter, obtaining the result in square centimeters.

c) The area of the triangle on the ground is calculated using the following formula wherein distance AB=a, BC=b, and CA=c.

$$Computed\ area\,(CA) = \frac{1}{2}\sqrt{a^2c^2 - \left(\frac{a^2 + c^2 - b^2}{2}\right)^2}$$

The scale is then calculated by

$$Scale = \sqrt{\frac{ground\ area\,(CA) \in M^2}{planimetered\ area\,(PA) \in CM^2}}\ X100$$

As an example, assume ground measurements are a=217 m, b=198 m, and c=204 m. The area of this triangle is measured on the photo is an area of 7.39 sq. cm. The scale of the enlargement in this example is:

$$CA = \frac{1}{2}\sqrt{217^2\,X\,204^2 - \left(\frac{217^2 + 204^2 - 198^2}{2}\right)^2} = 18351\ M^2$$

$$Scale = 1 : \sqrt{\frac{18,351}{7.39}}\,X100 = 1:4983$$

A factor can be calculated for converting planimeter measurements in square centimeters to their ground equivalent in square meters. In this example the calculation is: 18,351/7.39=2483.22. This factor is simply the number of square meters on the ground equivalent to each square centimetre measured on the enlargement. The scale is more useful than a conversion factor because a digital planimeter will make the conversion automatically once the scale is known.

3. If there are personnel with experience in measuring fields by taking angles and distances around the plot to be measured, and using a computer or programmable calculator to derive the area, this would probably be the most

CHAPTER 8

THE FIELD DATA COLLECTION

8.1 Preparing for Data Collection

8.1.1 Questionnaires and Forms

The design of a new questionnaire should not begin until the end product of the survey is defined in detail by drawing up examples of the output tables and listing all the desired comparisons and cross tabulations to be made. Keep in mind that data users are the only justification for data gathering, so include them in all the planning. The person in charge of data processing must be included from the beginning to advise on the layout for data entry and the practicality of the desired tabulations as well as the type of output needed for preliminary and final analysis. All data gathered must have a demonstrable utility. If there is no use for it, don't ask for it. The following are guidelines for good questionnaire construction:

1. Use good quality paper that can withstand rainstorms, erasures, and a lot of handling and make it of a physical size that is manageable by the interviewer (fits on a clipboard).

2. A short questionnaire is the best questionnaire. For a multiple-purpose national agricultural survey, it is best to limit average interview time to about 30 minutes. A properly designed questionnaire should enable to collect a wealth of information in that time.

3. In some cases, tables can be used effectively with experienced enumerators but as a general rule, each question should be written so that the enumerator will obtain results by reading the question to the respondent.

4. Use simple language that can be understood by the enumerator and the respondent. Simplicity is even more important when the enumerator has to translate the questions into a local language. Always have the questionnaire reviewed by someone with extensive field experience. Test the translation of a questionnaire into a local dialect by having an outsider translate it back into the original language.

5. Arrange questions so that the interview flows naturally with no skipping backward and forward between questions except as indicated by skip techniques. Begin with questions that are easy to answer and of interest to the respondent. Hold questions that might be considered difficult or controversial until the latter part of the questionnaire.

6. Use "skip" techniques and short enumerator instructions as needed throughout the questionnaire. Use a different type style or color for the enumerator instructions. For example:

Do you have any goats?___YES- *continue* ____NO- *go to section 2*

(Enumerator Instruction) add the classifications and see if respondent agrees with total. If there is agreement, continue; if not check each entry and make corrections.

7. Where possible, use the same terminology as holders are accustomed to use and ask questions in a sequence that will make sense to the respondent in terms of their relative importance.

8. Reserve a space for enumerator comments and observations on each page. There will be many unusual situations that require explanation.

9. If data is to be coded for summarization, pre-code as many data items as possible on the questionnaire. Later manual coding should be kept to a minimum.

Have the preliminary version of the questionnaire reviewed by data users, the statisticians, the data processing personnel and the field supervisors. When all are in agreement, prepare the enumerator's manual. The enumerator's manual should cover each question on the questionnaire, explaining exactly what is expected and giving specific instructions for that question. It must define *all* terms used in the survey such as *cultivated land, annual crops, fallow land, resident holder, interplanted crops, cow, heifer, steer,* etc. The manual should also cover interviewing techniques, the overall survey procedure and timetable, enumerator responsibilities, and administrative details. Prepare the manual with the idea in mind that it not only serves as a guide for the enumerator in this survey, but also as a reference for future surveys when the purpose is to collect comparable data.

Next, test the preliminary questionnaire in all types of areas where the survey is to be carried out. Use enumerators for the test who have been trained with the new questionnaire and manual. Do not use people who worked on developing the questionnaire efficiently. The purpose of the test is two-fold, to find out how the respondents react and to see if enumerators can handle the questionnaire efficiently. The questionnaire test can also be the pilot survey that tests enumeration procedures. The test will often show that some of the questions are not understood as asked. It may also show where more emphasis is needed in training and that some field procedures need modification. It is appropriate to complete enough questionnaires during the test to also try the data entry and summary system. This may indicate the need for further changes. The final questionnaire and manual can be printed after all changes needed as indicated by the test have been incorporated. Detailed editing instructions for each question can be prepared when the questions are finalized. These editing instructions would include checks such as expected range of crop yields; relation of certain questions with preceding or succeeding questions; procedure to be used in correcting livestock age and sex classifications; how to code answers for questions that were not precoded; etc.

Data collection for area frame surveys require a *control sheet* in addition to a questionnaire. The control sheet is designed to register *all* the tracts in the segment. Some tracts may not merit a questionnaire but their existence and area are recorded on the control sheet to make sure that everything in the segment is accounted for. The control sheet includes screening questions to identify those tracts for which a completed questionnaire is required. For a first survey in an area, print at least 20 questionnaires per segment plus enough control sheets to handle 25 tracts in each segment. In addition to this add 10 questionnaires and control sheets for each person involved in enumerator training then add 15% for extras and keep the masters just in case. Provide a durable waterproof envelope of an appropriate size to keep the questionnaires for each segment.

In addition, the area frame survey requires a third form for each segment in the sample indicating how to access the segment from a nearby village or location included in the maps. This form, the *segment location control sheet*, can be prepared in advance or during the first round of the survey, and is intended to facilitate the location of the sample segments for supervision and future rounds of the survey.

8.1.2 Timing

A survey is a very expensive undertaking that involves an important logistic effort; it is vital that the enumerators are in the field at the proper time for gathering the desired information. Plan well ahead so that the survey is conducted at the proper time. For instance, if a priority interest is planted area of spring crops, the beginning of the survey should coincide with the end of the spring planting season. If the main interest is harvested area and production, the survey should begin immediately following the harvest. Being there at the time of an event or immediately thereafter avoids memory bias which can be significant in countries where the farmers do not keep records. Avoid, if possible, periods of heavy rains, in order to facilitate the logistics and survey data collection.

8.1.3 Hiring and Training Supervisors and Enumerators

The enumerators must be able to read and understand the enumerator's manual and learn to work with aerial photography and maps. A short quiz can be devised on reading comprehension, ability with arithmetic, and aptitude for learning to use maps and photography. Generally, qualified people from each locality of the survey area who have worked in other types of agricultural surveys will make the best enumerators.

Supervisors should be hired and trained *before* the enumerators. The training for supervisors should be intense and demanding. Topics covered would be mostly the same as for enumerators. The supervisors must also attend and assist with enumerator training.

Enumerator training should be carried out in a pleasant area without distractions with proper seating and teaching facilities. Show enumerators that they are expected to learn by checking and criticizing their work. Commence training with more enumerators than will actually be needed so that unsuitable students can be dismissed from training. Also, reserve enumerators will be needed due to losses during the survey through illness, etc. The following list of subjects to be covered in enumerator training is not exhaustive, but outlines the type of coverage in the course. The training and the actual work will be similar to that in any other agricultural survey except for learning how to work with the photography and using the segment as a guide for the application of the questionnaire. Make sure that the enumerators understand the importance of obtaining a response for each question on the questionnaire from every reporting unit (tract or holding) in the sample.

1. Purpose of the Survey - In order to explain this to the respondents the enumerator must understand and believe in the purpose of the survey, and how the survey will be beneficial to the respondent and the country.

2. Responsibilities as Enumerators

3. Interviewing Techniques

4. Definitions and Sampling Procedure- One of the first questions a respondent will ask is " why did you pick me for this?" and his neighbour will ask "why didn't you pick me?" An understanding of the sampling frame and sample selection process allows the interviewer a means of answering these questions.

5. Enumeration Procedures and Use of Maps and Photography

6. Questionnaire in Detail (line by line)

7. Practice Interviews (role playing)- This should include both acting as interviewer and recording answers on the questionnaire from a simulated interview. The filled out questionnaires must be checked and errors pointed out.

8. Field Practice- This is usually carried out near the location of the training school at cooperative households, chosen and notified in advance. The trainees should work in teams and each person needs to have the opportunity to do an interview.

9. Critique of Field Practice- The teams can do a critique of their own members. Supervisors should have an opportunity to make comments. An overall evaluation could be presented by the survey supervisor.

10. Administrative Procedures- The enumerators need to know how, when and where they will be paid and the records that must be turned in to receive per diem etc. Enumerators should know how to contact supervisors in case of difficulty and extraneous events affecting the enumeration.

8.1.4 Maps Showing the Location of Sample Segments

Maps showing the location of sample segments must be prepared for each region. A file copy that stays in the office and a copy for the field are needed. Topographic maps in the scale of 1:50,000 are ideal for this as they also show roads and other means of access and can be used by the supervisor or team leader for planning each day's work. Segments are outlined in red permanent coloured pencil. The segment number is shown inside or alongside the segment also in red.

Master segment location maps for regions or larger areas can be prepared on larger scale maps (1:250,000 or 1:500,000) with copies for field offices and a copy for the central office. Up to date road maps will be needed by each enumeration team.

8.1.5 Equipment for Field Staff

Each enumerator will need a clipboard, a scratch pad, a box of common pencils, 3 or 4 soft white block erasers, a dozen red and a dozen blue all media marking pencils (erasable grease pencils), a small plastic ruler, a small first aid kit, a compass, a 100 meter tape, a carrying case for enlargements and some sort of briefcase or backpack for questionnaires and things besides enlargements. An inexpensive calculator should be provided. A transparent grid to assist in estimating areas can also be provided if segment enlargement scale is uniform. The grid may confuse enumerators if enlargement scales are highly varied. Finally, they will receive enough control sheets, questionnaires and segment envelopes to cover the number of assigned segments plus extras.

Supervisors should have a clipboard, a scratch pad, a box of common pencils, black ball points, a box of coloured pencils (green will do), 3 or 4 soft white block erasers and a larger first aid kit, a compass, a transparent grid to assist in estimating areas, and a calculator. The supervisor will also maintain a supply of extra pencils, grease pencils and erasers. Each supervisor should also have a 100 meter tape for making scale determinations and in case field or segment measurements are necessary. The supervisor will also maintain a supply of questionnaires, control sheets and segment envelopes.

8.2 Organization of Field Work

Data collection should begin simultaneously in all regions having a similar crop calendar so that the information is comparable. Enough enumerators and vehicles should be used so that the survey can be completed in two to three weeks. If a survey takes 6 to 8 weeks, it is likely that much of the data gathered at the beginning is not comparable to that gathered near the end.

If harvesting/planting seasons vary widely in a country, it may be possible to finish a survey in one region and then move the team on to the next region in time to begin at the proper moment - and so on. For this to work, each region must be distinctly different, the survey time in each region must be short and well coordinated and the survey must be well focused on the priority information.

The size of a segment is supposed to be such that an (experienced) enumerator can finish data collection in one day. This will not be the case with the first pilot survey where it is advisable to allow three person days to complete each segment.

The basic building block of the survey organization is the survey team consisting of 3 to 5 enumerators, a supervisor and a vehicle. In some situations it is effective to hire supervisors who can drive the vehicle rather than hiring a chauffeur. The number of survey teams assigned to each region or survey area will depend on the number of segments to be enumerated and the time allowed to complete the work.

The teams are organized under a regional supervisor who reports to the central office. Survey planning requires good sets of national and regional maps showing all roads, cities, and towns plus the location of sample segments. Groups of segments can be assigned to teams, routes for arriving at the segments can be selected and vehicle mileage estimated. Allowances must be made for gasoline, oil, tires, batteries, greasing and washing, minor repairs, and probably for one major accident. If coupons are used for gasoline, make sure they are accepted in all survey areas.

8.3 Enumeration Procedure

The enumerator uses the topographic map or a road map showing segment locations to identify routes of access and arrive at the segment. Or, the supervisor, using the maps, will drop the enumerator off at the segment. The enumerator should go around each segment on foot or in a vehicle, making certain that the boundaries are completely identifiable. Segments with ambiguous boundaries are reported to the supervisor as soon as possible. However, if the enumerator cannot contact the supervisor until the end of the day and has no other segment to work on in the area, the enumerator should try to enumerate the segment using the nearest real boundary to the one on the enlargement that cannot be located. The enumerator marks the limits of the enumeration with the red grease pencil. Hopefully, only a small portion of the original boundary will be unidentifiable. This segment is reported to the supervisor who must either locate the boundary drawn on the photo or adjust the boundary and explain the source of the problem.

The enumerator should complete the form that describes the access route to the segment.

After the enumerator is satisfied that the boundaries are well defined, enumeration commences at the nearest occupied dwelling from where he is, or by approaching the nearest visible worker in a field inside the segment. The next step is to identify the *holder* of a *tract* within the segment. The holder helps the enumerator delineate the tract on the transparent overlay of the photo in blue grease pencil. The tract is given an identifying *letter* in blue grease pencil within the tract boundaries and the screening questions on the control sheet are asked to determine if a questionnaire should be completed for the tract. Where no questionnaire is administered, only the holder's name, area of the tract, and land use are entered only on the control sheet. As soon as a tract is identified, the enumerator will ask the names and whereabouts of the neighbouring holders for future reference, meanwhile continuing with the interview of the current holder. The enumerator should be required to ask *all* questions and record an answer for all questions except those passed over by skip techniques.

Figure 6. A segment delineated on aerial photography.

8.3.1 Data Collection for the tract as reporting unit (closed segment estimators)

Following the entry of the holder's name and other identifying data on the questionnaire, the enumerator proceeds to fill in the data for variables defined in the tracts corresponding to closed segment estimators. In this process, each *field* is outlined in red grease pencil and given a sequential number within each tract. Numbering starts over in the next tract. The fields may have been outlined at the same time the tract itself was being delineated. The data for each field, which probably consists of area and crop grown, is then recorded on the questionnaire. The area of each field is provided by the holder. If the holder has no idea about field areas, the enumerator should make sure the fields are carefully drawn on the photo so they can be measured with a planimeter or a digitizing table in the office. Only fields and tracts that are *within* the segment are included.

8.3.2 Data Collection for the holding as reporting unit (weighted and open segment estimators)

For a given variable defined in the holdings corresponding to a weighted segment estimator, total holding data is collected from each holding that has land in the segment. This total holding data is then given a "weight" according to the proportion of the holding that is in the segment (tract). If 15% of the holding is in the segment, 15% of the all holding data from the questionnaire is assigned to that segment to be used in the summary process.

The weighted segment estimator requires to define the holding and find an informant that can provide data for the totality of the holding.

If data is collected on livestock herds for animals that are *controlled* in pastures and/or *corrals*, first the data on animals that are kept within the tract and cannot leave the segment should be taken. Then, find out if there any animals that can move freely (gates are open) into and out of the segment. Determine the area outside and inside the segment over which they are free to roam. A proportion of the total of these free-moving livestock will be assigned to the segment based on the proportion of the total area over which they can roam that is inside the segment. The animals must be able to move in and out on their own free will. If a gate is closed, they are either in one place or the other. This does not work for animals that are turned out of a corral to run on free range during the day, returning to the corral at night.

Upon completing an interview, the enumerator must review the questionnaire to make sure all the spaces are filled with reasonable answers, that recorded field areas add to the tract area, and that other classifications, such as livestock, add to the indicated total. This will be done before leaving the area of the interview so that it is possible to return to the respondent if a doubtful entry is found. The enumerator is the *only* person handling questionnaires that may erase entries to make changes.

8.4 Supervision

The supervisors are key people in the survey process. If working in a survey team with a vehicle, they plan each day's work and leave enumerators at the segments to be enumerated that day. They pick up the enumerators and review completed questionnaires with the enumerator before leaving the segment. The supervisor looks at each field in all tracts, checking the recorded size against the apparent size. Field areas are summed again to be sure the tract total is correct. Tract totals are then summed as a check on segment area. The supervisor will check this against the planimeter measure of the segment. If the segment is relatively level, the enumerated area can be allowed to vary from the area measured with the planimeter by plus or minus 10%. If the segment is mountainous with steep slopes, the enumerated area should never be less than the area measured with the planimeter and could exceed it by as much as 30%. The rest of the questionnaire is checked to be sure that answers are present where required and that all answers are reasonable. The only truly valid corrections are those that are made with the respondents in the segment. Once the supervisor and enumerator leave the segment, it is unlikely that they will be able to return for the purpose of making corrections. All changes made on the questionnaire after this will be based on averages or will be opinions or guesses. The supervisor makes changes by lightly striking through the table to be changed and writing in the correction above or to the side of the original entry. This is done in green pencil. Each editing level uses a different color so that it can be determined where changes were made.

The supervisor must observe an interview by each of the enumerators in his team as soon as possible after the start of the survey to detect and correct any bad habits or misunderstandings before too much damage is done. Arrangements with the local authorities for the survey in their area is another responsibility of the supervisor. Maintaining cordial relations with the local populace is also part of the supervisors job.

The supervisor must also make a random selection of completed segments from each enumerator (perhaps 2, at least 1) to be re-interviewed for the purpose of quality control. The main value of this is in having the enumerators know it will be done. The regional supervisor should arrange to spot check one or two segments in each team under his control. The regional supervisor reviews all completed questionnaires as they are passed to him by the team supervisors. The regional supervisor makes corrections in blue pencil.

The supervisors form the chain that moves completed questionnaires from the field to the central office. The flow is from team supervisor to regional supervisor to the central office.

CHAPTER 9

DATA PROCESSING AND ANALYSIS OF SURVEY RESULTS

9.1 Questionnaire Flow and Review (Editing)

As described earlier, the enumerator makes the first review of the completed questionnaire before leaving the scene of the interview, making sure that all questions are answered except for valid skips and that answers are reasonable. The enumerator should look at fields again to check enumerated area against apparent area and a check should be made that field areas sum to the listed tract area. The enumerator may erase to make corrections.

The team supervisor reviews all the questionnaires with the enumerator, together with the enlargement and the control sheet of completed segments with the enumerator before leaving the segment. The team supervisor uses a green pencil and strikes through entries to make changes. The team supervisor passes the completed segments on to the regional supervisor who reviews each questionnaire and passes those with errors back to the appropriate team for correction. The regional supervisor makes changes in blue pencil and sends the questionnaires on to the central office. Completed questionnaires need to move through the review system quickly so as to detect cases where enumerators and even supervisors have misunderstood instructions or are failing to perform their jobs properly. With early detection, instructions can be clarified and changes made in personnel in time to salvage the survey result.

Each level of review must have a control sheet to show which segments were completed on what date and the date the questionnaires were passed on to the next higher level. Clerical staff at the central office will check in completed segments received from the field and verify all additions on the questionnaires. Segment questionnaires are then edited by statisticians or experienced survey technicians. One will edit and initial the segment envelope and return it to the edit box to be edited by another who will proceed as if no edit had been done. Any disagreement is settled by discussion between the two editors. The second editor initials the envelope and takes the segment to the box that is ready for data entry. Another control sheet is used here that shows segments and date sent to data entry. It will also show when these same segments were returned by data entry.

Editing must be held to an absolute minimum. It is necessary to prepare in advance detailed editing instructions for each question that show, among other things, the expected ranges for variables such as crop yields, pigs per litter, eggs per day, etc. Data outside these ranges would be checked and accepted if they are possible and supported by enumerator comments and/or other data on the questionnaire. For example, a crop that was grown with a high level of fertilizer and irrigated.

Some respondents will refuse to answer any questions in which case the enumerator should be instructed to record all the observations possible about the tract. The editor then uses the observations and information from adjoining tracts to fill out the questionnaire.

Changes made during the final edit are made in red pencil by striking out the figure to be changed and writing in the new figure above it or to the side. All original figures must remain in a legible condition. The reason for this is that the original figure is sometimes found to be the correct figure after all or it is useful in some way for making decisions about other sections of the questionnaire.

The editing process requires the careful checking of all items including area of fields, area of tracts and total segment area. Disagreement between enumerated area and area measured with planimeter is often caused by an underestimation of pasture land, forest or wasteland. If cultivated land is correct, review usually shows that other land was underestimated.

9.2 Data Entry

Plans for data entry should be made when the questionnaire is being designed, because the questionnaire must conform with the data entry system and vice versa. As much as possible of the questionnaire should be pre-coded with only a limited amount of coding to be done by the editors. Another system is to have a program (software) that displays the questionnaire page by page on the monitor. Data is entered in the blanks and stored properly by a data base program. In either case, edit checks can be programmed into the system that will alert the data entry operator if certain errors are made. The question always comes up about whether or not to verify (repeat) the data entry. A well tested system to check on the accuracy of data entry personnel uses what are known as "hash totals". Clerical staff add up all the data items on each one of a group of test questionnaires. The hash total for each questionnaire is entered in a data cell on the last page. The data entry personnel enter the data from the questionnaires as normal, including the hash total. For this to work, the data entry people must not be aware of the "test". The computer programmer then makes a program that adds all the data items for each questionnaire and checks that sum against the hash total. An error printout will show the questionnaires where the computer sum doesn't match the entered hash total. If the hash totals are done carefully and checked, what appears on the error print are due to data entry errors. This is occasionally very illuminating.

It is often sound practice to summarize results by hand or with the help of a calculator in order to verify the computer results; quality control being of primary importance in survey work. It is possible to summarize even relatively large surveys (3,000 to 4,000 questionnaires) by hand using listing sheets and adding machines. Computers should obviously be used where available but with proper controls. Many surveys can be processed without them.

Data entry should begin as soon as a number of segments have been edited. This allows early detection of any problems with entry or summary programs and the accumulation of data in a "clean" file so that final results can be obtained quickly when the final segments are received, edited, and entered.

9.3 Summary of Results. Data Analysis

9.3.1 Preliminary Tabulations

Review of preliminary tabulations is a part of survey work that is often not done as well as it should be. If time allows, it is advisable to generate a data listing. This is simply a printout of each data item for each questionnaire in the survey. This can be visually checked for reasonableness and compared with each questionnaire as a verification of data entry.

With computer summarization, the next product generated is an error listing. The error listing is the result of a computer program that identifies those data items that don't conform to predetermined values or ranges, are inconsistent with other answers, or are incorrectly summed. The questionnaires are compared with the error listing and a decision is made to accept or reject the questionable item. Corrections are made on the error listing, entered into the data file and subjected again to the edit program. Any additional errors are corrected and resubmitted. Data items that are not singled out by the edit program automatically go to a "clean" data file.

When all the data passes through the edit program and is entered in the "clean" data file, other preliminary tabulations such as frequency distributions, totals and averages by region and by replicate and ranges can be produced prior to expansion and final summary. If a range extends beyond what is expected, the computer can show which item or items are extreme. If an average or total for one replicate is very different from the other replicates, investigation of the odd replicate is in order. There may be no error but the source of the difference should be known.

9.3.2 Estimates and their Coefficient of Variation

The next step after all the preliminary reviews is to carefully study the expanded data for each land-use stratum, observing the coefficient of variation (CV). Any land-use stratum that appears to contribute excessively to the overall variance is examined further for possible errors. If no errors are found, consideration should be given to increasing the sample size in that land-use stratum if it causes the overall CV to exceed desired limits. The final regional and national expansions are checked against all estimates from other sources and against check data such as exports, market records, cotton ginnings, purchases by marketing authorities, etc. New survey data will often disagree with these established estimates making it necessary to prepare a sound justification prior to publishing survey results.

9.3.3 Treatment of Outliers in the Summary Calculations

An outlier is a data value which differs considerably from the average of the other data values in the set (perhaps beyond 3 standard errors above or below the mean) and often contributes a disproportionate amount to the direct expansion estimate and its variance. Lists of large holdings and specialized and localized producers must be developed prior to any large scale area frame survey for the precise purpose of properly accounting for units that will be outliers if they should be encountered in the area sample. Holders' names from the area sample are checked against these lists and any overlap with the lists are removed.

Even small lists can be incomplete and list checking is a tedious task so some area sample questionnaires will contain outlier data even after checking against the lists. Proper checking and editing prior to data entry should identify many of the possible outliers before they get into the summary process. Once into the summary, outliers may be best detected by frequency distributions or by examining stratum totals and variances, replicate totals and geographic totals and variances. A higher variance or expansion will indicate the data grouping that should be examined further either by the computer or manually to find the outliers.

One way of dealing with outliers is to remove them from the direct expansion; making a direct expansion without them. Then add them back in to the final expansion at the level of representation they would have had if they had been on a special list. This would be done where the outlier is an extremely large report and it has been possible to get most of the potential outliers on a list.

Where the condition of being an outlier is caused by large expansion factors, the outliers are removed from the direct expansion and then put back in with a weight that is the value the unit should have received if it had been classified correctly.

In all cases, the estimate and the sampling error with outliers present should be computed. After adjustments for the outliers have been made, new estimates and sampling errors should be computed. These two estimates- unadjusted and the adjusted - need to be evaluated together when determining the final estimate. If the ranges of the sampling errors of the adjusted and unadjusted estimates overlap, the area of overlap represents a compromise between the two. Remember that even though the outlier will probably cause the initial estimate to be too large, the adjustment procedure may cause an under estimate; cf. Vogel (1986). This possibility of an underestimate is particularly important in cases where knowledge about the number of large holdings is incomplete.

In areas of heterogeneous agricultural activity, the detection of outliers is extremely important to the accuracy and precision of the estimates. Care should be taken by the statisticians to analyze the data for potential outliers.

9.3.4 Non response and Missing Data

Non response can be the result of a refusal to provide data on the part of the holder or of inability to contact the holder. In these cases, the enumerator must record as much as possible information based on observation and interviews with neighbours. The local supervisor must contact refusals in an attempt to obtain a response. If that fails, the regional supervisor should then try to contact refuses. The supervisors should also be making notes. If they fail to initiate a response, the notes from all levels are used to fill in as many blanks as possible. Remaining blanks are filled in by looking at responses from other similar sized holdings in the surrounding area and observing the segment enlargement. Also, the report of the respondent for the previous year is checked. The respondent may not have refused the survey then. Tract data is usually not a problem since land use is easily observed.

Segment inaccessibility can be caused by floods, snow, political unrest, etc. and may result in the loss of data for entire segments. When a segment is lost, it must be checked whether data is available from a previous survey. If so, the historical data may fill the need if there is no indication of radical change in the area. Reducing the expansion factor in the stratum where the segment is missing is a solution that is frequently applied but may not be appropriate if the segment differs from the average. A flight in a light plane over an inaccessible zone to observe missing segments can often provide some information to be used in estimation. Treat each case of missing data (segment or portion thereof) as an individual problem.

Computer programs that impute missing data have been developed by a number of agencies after (and even before) they have gained experience with surveys. These programs fill in blanks with numbers based on averages and ratios for nearby holdings and/or regional parameters. This is often referred to as "hot deck" procedures and can easily become addictive since it relieves the statisticians of editing and trying to find missing data. This leads to the "hopper syndrome" - pour everything into the hopper and let the machine worry about it. In a short time, it is often found that an uncomfortably high percentage of the data items is provided by the hot deck.

PART III: THE MULTIPLE FRAME SURVEY DESIGN

The multiple frame sampling methods described, as indicated in the Introduction, combine an area sample design in which segments have identifiable physical boundaries (discussed in Part II), with a complementary list frame of agricultural holdings or holders addresses, to be completely enumerated during the annual survey field data collection.

This Part III describes multiple frame estimation, the construction of a list of special agricultural holdings, and refers also to data collection for the special holdings.

CHAPTER 10

MULTIPLE FRAME ESTIMATORS AND THE LIST OF SPECIAL HOLDINGS

10.1 Multiple Frame Estimators

A multiple frame estimator is the sum of an area sample estimator (using data collected in the sample segments) plus the total data collected from the special list of holdings. If the list is sampled, the multiple frame estimator is the sum of the estimators from both frames, the area sample frame and the list of special holdings. As already mentioned, for variables with a distribution such that a large percentage of the total is concentrated in a relatively small number of holdings, it is often convenient to use multiple frame estimators in order to improve the precision of the direct area sample estimators. Standard techniques are used to make adjustments for missing data.

The area sample estimator should *not* consider the tracts corresponding to holdings of the list of special holdings, in order to avoid duplications. Therefore, all holdings with tracts included in the sample segments should be compared with those of the special list of holdings, in order to eliminate duplications from the area sample.

For a survey variable, if the list of special holdings is completely enumerated, then the variance of the multiple frame estimator is equal to the variance of the area sample estimator considered. If the list is sampled, the variance of the list sample is added to the area sample variance to arrive at the total multiple frame variance.

If the list of special holdings is completely enumerated, the variance of the multiple frame estimator is equal to the variance of the area sample estimator, but the CV of the multiple frame estimator will always be less than or equal to the CV of the area sample estimator. Multiple frame surveys obtain therefore more precise estimates than area sample surveys for those variables partially accounted with the list of special holdings.

10.2 Construction of the Complementary List of Special Holdings

This section sets out the criteria for choosing an adequate *list of special holdings*, complementary to the area sample survey.

<u>Large Scale producers</u>. Holdings which account for a large share of the total production of some item to be estimated always present a special problem for sampling, and specially for area sampling.

For an area sample, livestock and poultry estimates, using both the closed segment and open segment estimators, are usually the most affected by the occurrence of one or more of these "extreme operators" in the sample. This is readily understood by assuming a situation where a holding with 7,000 cattle is picked up in a segment, although there are only very few such holdings in the stratum. Segment expansion factors often are in the range of 75-100, so that each sample segment is, in theory, representative of up to 100 segments in the population. The inclusion of this holding will completely distort the estimates and increase the sampling variance well beyond what was desired or expected. Also, it is very unlikely that there are 75 or more holdings of that same size in the stratum. The effect on crop area estimates can be equally destructive when using the open segment estimator. A closed segment estimator of crop area may not be as seriously affected but estimates could still be seriously in error in the case of certain crops where there is large scale production in a very localized area, such as occurs with rice in some countries.

The development of a multiple frame procedure that employs a list of the large producers is a procedure most often used to improve the area sample estimates for important variables. A list is needed of such large operations for each variable to be estimated. The list can be compiled from a combination of sources such as; the latest agricultural census, tax records, banks, producers' associations, extension agents, and other local agricultural technicians. The list should contain the location of the holding and the name and address of a person (holder, manager or administrator) who can consistently supply information about the holding. Information about crop areas, total land area, and size of herd or flock are needed in order to be able to place the unit at its proper level of importance on the list. Enough information is needed for a proper identification of the holding if it (or a portion thereof) should be encountered in a sample segment. The list should be updated frequently; once a year or before any major survey if surveys are sporadic. The original sources can be provided with a list of those holdings in their area and asked to delete those who are out of business, indicate new holders of old holdings, and add new holdings.

The size of the list of extreme operations is frequently determined by the capacity to enumerate extra holdings in addition to the area segments. A good place to begin is with the classifications by size of holding from the most recent census to determine the number of holdings in the largest classification for each variable to be estimated. If the sum of the holdings in the largest category is manageable, that can set the size. However, if calculations indicate that encountering a holding or holdings of the next largest size group will increase the variance of the estimate beyond the desired limits, that size group must also be included. After the first survey, the length of the list can be determined more accurately.

The largest holdings on the extreme operator list should all be enumerated if possible. Smaller size classifications can be sampled using simple random sampling. Any duplication between the list and the area sample must be eliminated from the area segment data. One way of eliminating this duplication is by checking the names of holders encountered in the area segments against the list. A better way is to outline on the frame material the area operated by each holding of the list. In this way, it can be determined immediately which segments will contain overlap. The area and location of land operated would have to be updated before each survey because the area operated by this type of large holdings can change significantly from year to year. Estimates from the two sources (area sample and special holding list) are added to make the overall estimate for a province or region.

Localized or Specialized Holdings. A list of holders can also be used in conjunction with the area frame when it is desirable to estimate rare or localized production even when holding sizes are not abnormally large; examples are: cut flowers, vegetables, fish farms, fur farms (mink), high technology poultry, horses, etc. A recent census of agriculture may provide a basic list for these specialty items and indicate the area where they are concentrated. Other sources of lists as well as the management and application of a specialized list are the same as those described above for large holdings.

In view of the considerations and examples above, important survey variables, with a skewed distribution in relation to holdings should be selected. For each of these variables, from the census or other sources, as already indicated, a short list of holdings that account for a relatively large percentage of the total known estimate should be selected to form at least part of the list of special holdings. The complementary list of special holdings should ensure the inclusion of a significant percentage of the total estimate for important survey variables, thus improving the precision of the area estimates.

The list of special holdings should be fairly easy to update annually prior to the survey field data collection, since the holdings involved are usually easy to recognize and well known in the region where they are located. Information can be obtained from extension agents, producers' associations, banks, tax records, agricultural censuses and from government agencies that control and purchase production of certain crops. The preparation of such lists should include the accumulation of data on each operation such as holding size, crops grown, type of livestock and inventory, etc. for stratification purposes if the list is to be sampled. In addition, in order to avoid duplications, a short list will be much easier to combine with the area sample than a large list.

Finally, it should be noted that the additional work involved in constructing a short list of special holdings and in its combination to the area sample is generally minor but gains can be very significant for the estimation of some important variables.

10.3 Data collection for the List of Special Holdings

Data collection requires two types of questionnaire, one for the data collection from the area sample and the other for the complementary list of special holdings.

The questionnaire for the special holdings is usually deviced to cover more specialized information than is included in the area sample questionnaire and does not generally involve objective measurement of areas or the use of the aerial photo enlargements, which are essential features of the area sample data collection procedure. Use a different coloured paper or some readily visible mark to distinguish between area frame and list frame questionnaries.

For large holdings, that overlap several administrative areas it will be necessary to break down the information in order to conform with the desired level of estimates (province, region, etc.).

The annual or seasonal data collection for the list of special holdings can be done by the supervisors by direct interview with holders or managers, or by delivering the questionnaire to the holder with a later collection. In some countries the mail can be used for delivery and return of the completed questionnaire.

10.4 Examples of Multiple Frame Estimators and their Variances

The following are examples of multiple frame estimates derived from some real data presented in the Brazilian case study (cf. Volume II). The proportion of the estimate from each frame is approximated and may not be exact but the totals are what are shown in the case study.

In the State of Paraná, the area frame sample consists of n = 525 segments distributed over 7 strata and selected from a total population of N = 99,540 segments. The average expansion factor for a segment was 189.6.

A survey of the area sample using the weighted segment estimator yielded a direct expansion of 9,091,000 head of cattle with a standard error of 399,040 and a CV of 4.4% (399,040 divided by 9,091,000). In this same survey, a list of 259 cattle producers with over 2,000 head each, was completely enumerated. All producers interviewed in the area frame survey were checked against the list and the cattle data was removed from the area frame expansion for those who were also on the list. The enumeration of the list produced a total of 885,000 head to be added to the area sample direct expansion with no contribution to the sampling variance since it was a complete census of the list. The sum of the two independent data sources is the estimate for the state, 9,091,00 + 885,000 = 9,976,000. Only the area sample contributes to the variance so the CV of the estimated total is 399,040 divided by 9,976,000 which equals 4%. Therefore the CV have been reduced. The list contributes to the accuracy of the estimate by (a) adding in a significant number of cattle with no sampling variance attached and (b) removing these large holdings from the sampling scheme where, if encountered, they would have caused a considerable increase in the sampling variance.

In the State of Sao Paolo the area frame sample is 658 segments chosen from a total of 127,170 segments distributed over 15 strata. The average segment expansion factor is 193.3.

The area frame was used in conjunction with a list of 181 cattle producers that have over 3,000 head on their holding. In this case, a random systematic sample of 35 producers was selected to be interviewed from the list of 181. The number of cattle estimated by the list frame was 941,000 head. The standard error of this estimated total was 84,945 head. The CV of the estimated total was 9% (84,945 divided by 941,000). All holders' names from the area sample segments were checked against the entire list (181 names). Cattle data for any of those that appeared on the list was removed from the area sample. The area sample direct expansion (estimated total) using the weighted estimator was 12,830,000 head with a standard error of 641,500 head expressed as a CV of 5% (641,500 divided by 12,830,000).

The combined total estimate (area frame + list frame) was 13,771,000 head (941,000 + 12,830,000). The standard error of this combined total is calculated by adding the squared standard error of each of the two subtotals and then extracting the square root of the sum which is equal to 647,100. The CV of the estimated total is calculated by dividing 647,100 by 13,771,000. This is equal to 4.7%.

CHAPTER 11

SPECIAL TOPICS

11.1 Use of the Agricultural Census

The latest agricultural census data and maps are important for the planning and implementation of a multiple frame annual agricultural survey. It may often be the basic reference and also the best source to determine the list of special holdings and to provide general information on the structure of agriculture.

The agricultural census list of Enumeration Areas (EAs) may be used for the construction of the PSUs of an area sampling frame, as explained in the sequel.

In order to conduct an agricultural census, most countries need to estimate in advance the approximate location of holders' housing units or holdings, and to assign to census enumerators well defined areas of work. The total area of the country (or the area to be covered by the census) is often unambiguously divided into identifiable areas corresponding to approximately equal enumerator workloads. These non-overlapping subdivisions of the country are called enumeration areas (EAs) or enumeration districts. Accurate mapping of the EAs, or eventually sketches if small scale maps are not available, ensures a proper coverage, avoiding omissions and duplications.

More precisely, the census enumeration areas (EAs) are geographic areas such that:

a) They constitute a complete subdivision of the land, covering all holders' housing units, with no overlap.

b) The boundaries of an EA should not cross urban, rural or political subdivisions of the country; and preferably the EA should have recognizable permanent physical boundaries;

c) An EA should be a compact piece of land, so that an enumerator could walk or travel between any two points of the EA without crossing its boundaries. In particular, a large river should never cross an EA.

d) Their area should correspond to approximately equivalent workloads weighting:

   - the approximate number of holders' households;
   - the distances and difficulties of access to the holders' dwellings;
   - the average time needed for each interview and the established length of the census enumeration period.

An agricultural census frame is a list of enumeration areas, with the estimated number of holdings or holders' households in each EA.

The EAs often are the smallest subdivisions of the country for which agricultural census data is available. In some countries, the census EAs are defined as subdivisions of relatively small administrative or political divisions (districts, for instance) if good maps are available. In other countries, villages are adopted as enumeration areas. If the latest agricultural census Enumeration Areas (EAs) were defined with recognizable physical boundaries, they may be useful in the construction of the area frame of PSUs.

However, the latest agricultural census data and maps of EAs may not be available or, if available, they may be either entirely or partially useless because they are too old or incomplete to define the units of an area frame.

The use of the agricultural census EAs to construct area frames is illustrated in the Brazilian annual agricultural and livestock surveys; cf. Volume II.

11.2 Use of the Population and Housing Census

For a population and housing census the enumeration units are the households (or housing units for the Housing census). Population census enumeration areas (EAs) are defined as those of the agricultural census, but considering households or housing units instead of holdings as enumeration units.

The EAs often are the smallest subdivisions of the country for which population census data are available and generally determine the boundaries of the urban, suburban and rural zones of the country.

The population census EAs are usually defined as *subdivisions* of the smallest administrative or political divisions (districts, for instance). In many countries, however, there are no good maps of EAs and only sketches are used. In a number of countries, the agricultural census EAs are *groups* of contiguous population census EAs within each political subdivision and urban or rural zones.

Population and housing census EAs have been prepared by most countries. These EAs should be considered in the search for a sampling frame and may prove to be usable depending on accuracy required, budget available, and style of agriculture in the country. The following questions should be asked when checking a population census frame:

- Are all the maps and records for all the EAs still intact and readily available? In practice, one finds that maps badly used or destroyed during the census are not replaced and finding specific records of population and agricultural holdings for each enumeration area is not always possible;

- Do the boundaries of the enumeration areas follow physical terrain features? Many times the boundaries are arbitrary lines drawn on the map that cannot be accurately located on the ground;

- Is there a good estimate of agricultural holders or holdings for each enumeration area?

If one of these questions cannot be answered positively, the population census EAs cannot be used for constructing an area frame for an agricultural survey. In any case, before using a population census frame, it should be kept in mind that it's purpose was to record population, and agriculture was usually of secondary importance if recorded at all. Therefore, the use of population census EAs is not generally recommended for constructing an agricultural area sample frame. When they are used, it is usually necessary to subsample in the census EAs because they are often constructed to contain 100 to 250 households/housing units.

The EAs prepared for carrying out the latest population census may not have always been corrected and adjusted on the basis of the information collected during the census. This is also often the case for the agricultural census EAs. In these cases the EAs should be updated if it is considered worthwhile.

Population census questionnaires, in most cases, do not allow a direct link to be established between agricultural holdings and holders. However, the population census questionnaires generally identify the population whose main activity is agriculture within the context of a fairly short time reference period.

The PSUs of an area frame must also cover the agricultural holdings that are not directly associated with holders' households, such as large plantations or cooperatives. If a list of holders is obtained by screening a list of households or housing units from the population census EAs, it is quite possible that those holdings will be omitted in the list. A special effort is required, therefore, to compile a list of such large holdings from other sources in order to ensure their coverage. Although the population census may provide very useful data, it should not constitute the sole source of information for the preparation of an agricultural census frame.

11.3 Classification Methods of Digital Satellite Data for Improving Area Sample Estimates of Selected Planted Crop Areas

The variance of direct area sample estimates of certain planted crop areas can be reduced by using regression estimates applied to digital classification methods of satellite data (LANDSAT and SPOT).

Such methods, have been successfully applied in a number of countries under different conditions. However, they have serious limitations and, as a consequence, they have not been integrated in the periodic national agricultural surveys in any country.

The main limitations of these methods, for such a purpose, are: a) they require the use of *current* satellite data, which is in many situations difficult, expensive or impossible to obtain; b) they refer only to area of *planted* crops with a certain maturity, not to all planted area, nor crop harvested, nor to the area of the crop to be planted; and c) The classification procedures for digital satellite data, which rely on interpretation techniques for identifying the above mentioned planted areas of crops are applicable only to few crops in a given planting season, for a given region.

11.4 Objective Yield Surveys

The area sample provides a simple way of selecting a self-weighting sample of fields for crop cutting and crop forecasting. Holders in many countries are not able to accurately estimate the production obtained from a specific crop, or perhaps they can only provide an answer in some units of measure that are not easily converted to a standard unit. Crop cutting surveys are widely relied upon to provide acceptable estimates of crop production. Crop area surveys are usually timed for early in the growing season when most or all of the crops are planted. For an area sample survey, the area, location, and crop grown is registered for every field in all the sample segments. To select a self-weighting sample for crop cutting, the first step is to complete a table such as the one illustrated below.

Segment Number	Wheat -area in hectares	Segment Expansion Factor	Expanded Area of Wheat	Accumulated area of Wheat
1	13	81	1,053	1,053
2	25	78	1,950	3,003
3	18	105	1,890	4,893
4	31	78	2,418	7,311
5	23	105	2,415	9,726
6	11	105	1,155	10,881
7	16	81	1,296	12,177
8	22	81	1,782	13,959
9	41	78	3,198	17,157
10	28	68	1,904	**19,061**

A self-weighting sample of 4 fields for the placement of crop cutting sample units can be selected using the imaginary data shown above. First, a sampling interval (I) is established by dividing the total accumulated area of wheat by 4, I = 19,601 / 4 = 4,765.25. A random start between 1 and 4,756.25 is selected from a random number table - say 2,346.17 - indicating that segment number 2 is the first segment in the sample. The next segment is selected by adding the sampling interval; 2,346.17 + 4,765.25 = 7,111.42; selecting segment number 4 as the second in the sample. Continuing to add the sampling interval selects segments 7 and 9 to complete the sample of 4 segments. Next, the individual field within the segment to contain the crop cutting plots must be chosen. To do this, a listing of fields within the segment is prepared for sample segment 4. The letter in the field number denotes the tract.

Select the field, in this case, by choosing a random number between 1 and 31. If the number selected is 7, field number 4 in tract B is chosen. The field can be located precisely on the segment enlargement. The sample plots can be placed in the field and harvested when the wheat is mature. A simple average produces the regional and national estimate of yield.

Field Number	Area in Hectares	Accumulated Area
A-2	4	4
B-4	11	15
B-6	6	21
D-2	10	31

These sample fields can also be used to provide sample plots for forecasting crop yield by measuring crop characteristics before harvest that can be correlated with final yield. Reliable forecasts for a number of crops can be made up to three months prior to harvest.

11.5 Some concluding views for the planning and implementation of multiple frame current agricultural surveys in developing countries

Technology is advancing rapidly in the field of remote sensing, global positioning, and measurement of areas and distances. Data from heretofore classified satellites is being released for public use. Satellite images are now available for many countries of the world. Competition in the field of satellite imaging is likely to reduce costs and improve quality. This means that the possibilities for building a precise area frame at a lower cost are improving. However, remote sensing satellite technology is still unable to provide a quality scale enlargement for measurement of segments and for measurement of the fields within the segments. Aerial photography is still required for such purposes.

National or large-scale agricultural surveys based on area sampling methods from an accurate area frame employing physical boundaries for strata and sample segments, defined by the use of aerial photography, satellite images, and topographic maps are currently used in a large number of countries in North Africa, Asia, North, Central and South America and Western Europe.

In a number of countries, although the construction of the area frames was successfully carried out, often with technical assistance, the country did not have the personnel, budget or the motivation to properly maintain the frame and support the expensive field staff and vehicles needed to carry out a program of periodic surveys. In most of those cases, effective support from data users was lacking. Launching and maintaining a survey programme has been an expense many countries were unable to bear. In other cases, political instability has completely disrupted installed sampling programmes.

In order to initiate or to rehabilitate an existing agricultural data collection programme using a multiple frame (area/list) survey method, the following points are considered critical to success:

* Enlist the support of data users. If they do not feel the need for an improved estimating survey system, all the effort and money invested in a multiple frame method can easily be lost. An informational, educational project on the

use of accurate agricultural estimates is necessary to promote the new survey system.

* The management of the survey should be in an agency where political influence and nepotism are not widespread. This may have more to do with a certain level in an agency than in the agency itself. Managers that seek the work without having the resources or staff should be avoided. An agricultural survey system does not have to be managed by the Ministry of Agriculture. In some countries, the National Statistical Office or another agency might be a better choice.

* Promote a regionalization of the country that is coordinated among all the agencies in the agricultural sector and accepted by the major data users.

* Begin the area frame at a province or at a regional level. Pick a province or region that is of particular importance to the government and to the country in terms of food production or export crops. Conduct a small pilot survey for testing and training. Then move to a sample of 250 to 300 area segments so that reliable estimates can be made. This is the best way to begin if a statistician with frame construction experience is available-usable results can be produced rapidly and a sizable staff can be trained for a continuing activity leading to a national survey. On the other hand, if local technicians are of good quality but inexperienced and if no technical assistance is at hand, a very practical way to get started in area frame construction is to select a small pilot area, allow the team leader statistician, his assistant and the cartographer/photo interpreter to do all the frame construction and selection of the sample. Records of time required, costs, etc. should be kept. Develop a simple questionnaire and perhaps 4 enumerators should be hired and trained. Information gained in this way will serve as an up-to-date planning guide for expansion of the program. At this point, it is highly advisable to have an experienced area frame statistician review the work done and the plans for the future.

* Budget for *current* satellite image coverage at a scale of 1:100,000 of the frame area and for some manner of obtaining current photography in a scale of 1:30,000 to 1:60,000 for subdividing the PSUs and ensuring that quality photographic enlargements of a known scale (1:5,000 or smaller) will be available for all sample segments. Some enlargements of satellite images to 1:10,000 and even to 1:5,000 have been attempted but they are usually not of sufficient quality for segment enumeration and measurement.

* Budget not only for the purchase of vehicles but also for their maintenance and replacement.

* Develop a list of good part time enumerators that are available for periodic work. Avoid using extension agents for enumerators. They may tend to feel that they know more about the holder's business than the holder himself so, many of the answers will be from the extension agents, not from the holder. Interviewing will be a job for them that is in addition to their normal work and there may not be a direct chain of command from the survey leader to the

extension agents. It is also possible that they are not very well liked by the local holders.

The multiple frame agricultural surveys described in this manual that combine an area sample of segments that have identifiable physical boundaries, with a short list of agricultural holdings that are completely enumerated may prove to be the best way for many countries to establish or improve an existing current agricultural survey programme in order to obtain the basic annual data for the agricultural sector.

BIBLIOGRAPHY

I. General Survey Sampling Methods

Cochran, W.G. (1977). *Sampling Techniques.* Third edition, John Wiley and Sons.

Deming, W. Edwards (1960) *Sample Design in Business Research.* John Wiley and Sons

Hansen, M. H.; Hurwitz, W.N. and Madow, W.G. (1953) *Sample Survey Methods and Theory.* Vols. I-II, New York: John Wiley and Sons.

Hansen, M. H.; Hurwitz, W.N. and Jabine, T.B. (1963) *The Use of Imperfect Lists for Probability Sampling at the US Bureau of the Census.* Invited Paper, 34th Session ISI , Ottawa.

Kish, L. (1965) *Survey Sampling.* John Wiley and Sons.

Murthy, M.N.(1967) *Sampling Theory and Methods.* Statistical Publishing Society, Calcutta

Raj, D. (1972) *The Design of Sample Surveys.* Mc Graw-Hill

II. Agricultural Survey Sampling Methods

Cotter, J.J. and Nealon, J. (1987) *Area Frame Design for Agricultural Surveys.* National Agricultural Statistics Service, USDA, Washington.

Cotter, J.J. and Tomczak, C. M. (1994) *An Image Analysis System to Develop Area Sampling Frames for Agricultural Surveys.* Photogrammetric Engineering & Remote Sensing, Vol. 60, N° 3, March 1994, pp. 299-306.

Fecso, R., Tortora, R. D. and Vogel, F. A. (1986) *Sampling Frames for Agriculture in the United States.* Journal of Official Statistics, vol. 2, N° 3, 1986.

Geuder, J. (1983) *An Evaluation of SRS Area Sampling Frame Designs* - Ordering Count Units and Creating Substrata. Statistical Reporting Service, USDA, Washington.

Geuder, J. (1984) *Paper Stratification in SRS Area Sampling Frames*; Statistical Research Division, Report 79, Statistical Reporting Service, USDA, Washington.

Houseman, E.E. (1975) *Area Frame Sampling in Agriculture* . Statistical Reporting Service, SRS N° 20, USDA, Washington.

Huddleston, H.F. (1980) *A Training Course in Sampling Concepts for Agricultural Surveys.* Statistical Reporting Service, SRS No.21, USDA, Washington.

King, A.J. and Jessen, R.J. (1945) *Master Sample of Agriculture.* JASA, vol. 40: 30-46.

Kish, L. (1989) *Sampling Methods for Agricultural Surveys* . FAO Statistical Development Series 3, Rome.

Kott, P. S. (1990) *Mathematical Formulae for the 1989 Survey Processing System (SPS) Summary.* USDA/NASS, Washington.

Nealon, J. P. (1984) *Review of the Multiple Frame Estimators.* Statistical Reporting Service, SRS N° 80, USDA, Washington.

Vogel, F.A. (1986) Sample Design and Estimation for Agricultural Sample Surveys. Statistical Reporting Service, USDA, Washington.

III. Other Related References

FAO (1995) *Programme for the World Census of Agriculture 2000.* FAO Statistical Development Series N° 5, Rome.

USDA (1983) *Scope and Methods of the Statistical Reporting Service.* Miscellaneous Publication N° 1308, Washington, 1983

USDA (1995) *Area Frame Design Information. Implementation Year of Current NASS Area Frames.* National Agricultural Statistics Service, Washington, 1995

BACK COVER

Much of the required information for the agricultural sector, such as crop production, livestock inventories, and basic social and economic data, is obtained through periodic national, multi-purpose agricultural data collection programmes called "Current Agricultural Surveys".

This manual describes the sample design, organization and implementation procedures for these current agricultural surveys using multiple frame probability sampling methods, which is a new methodology currently being adopted by many countries around the world. Specific case studies will be presented in Volume II.

This handbook introduces the subject and describes the practical aspects of list and area frame construction and sample selection methods while recognizing the resource and specialized staff constraints faced by developing countries.