THE MASTER SAMPLE OF AGRICULTURE

I. Development and Use

A. J. King

II. Design

R. J. Jessen

1945

THE MASTER SAMPLE OF AGRICULTURE*

I. DEVELOPMENT AND USE

By A. J. King

Bureau of Agricultural Economics, and
Statistical Laboratory, Iowa State College

The demand for accurate and timely information during the war has hastened the development of a new and efficient method of sampling. This method which has become known as "area sampling," is adaptable to many kinds of surveys where the sampling units are visited personally. At the Statistical Laboratory, Iowa State College, a Master Sample that will facilitate the designing and use of future area samples has been developed by the Bureau of Agricultural Economics and the Bureau of the Census. In developing the Master Sample, materials which are difficult to obtain have been collected and assembled, and much of the time-consuming clerical work required in designing area samples has been done; thus with the Master Sample it is now possible to design other area samples quickly and efficiently. The size of the Master Sample makes it possible to plan surveys so that no particular group of farms will be overburdened with interviews from different government agencies. Furthermore, the accumulated information obtained from the surveys taken will provide a clearing house of information for the various agencies using the sample. Such information in some cases will make it unnecessary to ask the farmers more than once for the same facts. The information will also be used to reduce the sampling errors of subsequent surveys.

Methods of sampling used in the past.--Among the most popular methods of extensive sampling used in the past, three are of special interest for comparison with area sampling; these are the mailed inquiry, the random point, and the quota method. All three have two basic weaknesses. In the first place, when applied, they usually give biased samples of the population under study. Secondly, they usually depend upon historical information as a basis for expansion. For example, many farm sample inquiries are designed merely to obtain an estimate of some average, such as a mean of an item per farm, which for an estimate of a total, is multiplied by a guessed, an assumed or an estimated total number of farms. These guesses, assumptions, or estimates are usually based on the last-published census figure. The census data are from one to two years old at time of publication and seven years may pass before new information becomes available. If the number of farms has changed in the meantime, any estimates based on this quantity is in error. Furthermore, if the census number of farms is used as a basis for expansion, it is essential that the definition of the farms in the sample be consistent with that used in the Census. Otherwise an additional bias may occur. This restricts the usefulness of the three methods because the farm definition used in the Census is not satisfactory for all kinds of inquiries.

---

*Papers presented at the 104th Annual Meeting of the American Statistical Association, Washington, D. C., December 27, 1944.

Journal Paper No. 1257 of the Iowa Agricultural Experiment Station, Ames, Iowa, Project 886, in cooperation with the Bureau of Agricultural Economics, United States Department of Agriculture and the Bureau of the Census, United States Department of Commerce.

The mailed inquiry is generally accomplished by sending schedules to a predetermined list of persons or by having them distributed by mail carriers as they travel their routes. This method of sampling has been popular among federal agencies because it is easy to administer and does not require a field force of enumerators. A large number of returns can be obtained at a low cost. In spite of its susceptibility to serious bias the method has been widely used where a large number of reports are required to estimate agricultural inventories which vary greatly from farm to farm, and also where geographic detail is needed. In these cases the smaller sampling error due to the large number of reports offsets somewhat the effect of the biases which are not taken care of in the estimation procedure.

Mailed inquiries have very definite limitations. The short questions that must be used are not feasible in many surveys which require indirect questioning, such as attitudinal surveys, or where questions are technically difficult, as in many farm management inquiries. Results from mailed inquiries are usually biased because it is difficult to obtain returns from those who are indifferent or who have writing or language difficulties.

The quota method of sampling is familiar to almost everyone because of the greater amount of publicity that it has received by its use in election forecasts, public opinion polls and marketing surveys. In this method the population is divided into a number of strata or cells, and the sample is drawn proportionately from each cell. The number of sampling units required in each cell is called the quota for the cell. Ordinarily the method is relatively simple to prepare, but it has some weaknesses which may be serious for many inquiries. The quotas which are set up to be filled by the enumerators are usually based upon census and other information. Since the census information is historical, the quotas may not be in agreement with actual population proportions at the time of inquiry. In periods of rapid change, such as in depressions and wartime, these quotas may result in biased samples. Then too, the selection of the individual within a quota group is left to the descretion of enumerators and this may lead to biased selection of individuals. Furthermore, this method is limited because it depends upon the Census as a basis for expansion.

The random-point method of sampling became popular for a few years but it is now probably losing favor for locating farms. In this method, points are located at random on maps and a certain number of farms nearest the points are included in the sample. A sample using this method can be quickly designed. Furthermore, the cost per schedule can be reduced by enumerating more than one farm at each point in order to reduce travel among farms. The returns are biased, however, because the random points are more likely to fall upon the large farms than on the small ones. Therefore, more complicated estimation procedures must be adopted and some outside information must be available both for removing bias and as a factor in estimating totals.

It is realized that in a particular inquiry a biased method of sampling may still give more accurate information per dollar spent, provided that a substantial number of returns can be obtained cheaply. However, since the Master Sample was designed to provide the basis for a larger number of surveys on many different topics, it was considered desirable that the sample should be free from bias.

In the area sampling with which we are concerned the sampling units are samll areas and every unit of observation is uniquely associated with one and only one such area. Furthermore, the total number of such area sampling units in the population is presumably known so that the proportion samples (i.e. the sampling rate) is kn To obtain an unbiased selection of farms we may select those whose headquarters are located within the boundaries of the sample area regardless of where the boundaries of the farm may extend (7). Similarly, an unbiased selection of persons can be obtained by taking all who reside within the sample area. As a further example, an unbiased selection of land area may be obtained by selecting for measurement all land confined within the boundaries of the sample areas (9), (10). Finally, an unbiased selection of fields may be obtained by taking those fields which have their northwest corners within the boundaries of the sample area.

Area sampling may be extended to provide the primary selection of units of observations which are then subsampled. The subsampling is usually accomplished either by a further sampling of smaller areas or by making a prelist of individuals; from this a small and possibly stratified sample may be drawn for a more lengthy interview. This method of sampling is commonly used by the Bureau of the Census and some others in sampling cities, counties, or other areas where the blocks are the areas which are prelisted and subsampled. The subsampling scheme is also contemplated for the Master Sample for its basic data can provide the necessary prelist information without the field work usually required.

Use of areas in sampling was widespread in Europe during the 1920's and early 1930's by Bulgaria (1), Italy (3), Poland (11), Sweden (5), etc. In all these samplings the areas were relatively large and often purposively selected. Jerzy Neyman (11) in his paper published in 1934 recommended: (i) that in general the stratified random selection of sampling units is preferable to purposive selection and (ii) that the sampling units should be small and widely scattered for maximum sampling efficience. In the later 1930's some experimentation with area sampling of farms was undertaken by members of the Bureau of Agricultural Economics and the Bureau of the Census. In the Bureau of Agricultural Economics an investigation initiated by Sarle (6) in 1936 was designed in part to study the sampling efficiency of the survey section as an areal unit compared with that unit comprising single farms and strings of from two to ten farms. J. B. Shepard in 1937 suggested that an agricultural sample census be taken (12) using the township as the sampling unit. Another investigation, initiated by Sarle (13), was directed to study the efficiency of the township as a sampling unit. Hansen has indicated that the Bureau of the Census also used the area method of sampling in 1936-37.

In most of these sampling investigations, it was generally hoped that the large area, such as the township, would win out over the smaller unit but nearly always the latter was found to be the more efficient. In 1937, a new use of area sampling was introduced for sampling crop acreages. This scheme, which appears to have originated independently in India (10) and in this country (9), requires the measurement of crop acreages contained within the sample areas, which in both of these samplings were relatively small. The study in this country was made possible by the extensive use of aerial photographs. The areas of fields in crops or parts of fields lying within the boundaries of the area units were measured on aerial photographs. In 1938 and 1939, Iowa State College cooperating with the Bureau of Agricultural Economics made two statewide farm surveys (7) in which the quarter-section was used as the sampling unit. Each farm with its farmstead falling within the unit was enumerated. Later in 1940 and 1941, statewide surveys (8) were made in Iowa and Arkansas by a large field force of WPA workers, using the survey section as sampling unit. In 1942, the Bureau of Agricultural Economics made surveys in California, Florida and New Jersey to obtain farm labor information (8). The Bureau of the Census in 1942 used the area method plus a subsampling scheme (2) in their new sample for estimating the United States labor force. About the same time, members of the Bureau of the Census (4) extended and further developed the theory for subsampling. Along with studies on the statistical efficiencies of areal units of different sizes, it has been found by Mahalanobis (10) and Jessen (7) that consideration should be given to cost because it too is a function of size.

Advantages and disadvantages of the area method.--It has been shown by Jessen and others that the area method gives an unbiased sample for agriculture. It has the further advantage that the units of observation are clustered in such a way that the sampling error can be minimized on the basis of cost. The reduction in cost is accomplished by reducing the travel between farms. The method is also adaptable to such modern techniques as stratification, double sampling and subsampling for reducing the sampling errors. An expansion factor available to the area method is known without error because the total number of sampling areas in the population may be determined by simply counting them on the maps. Hence, it is not necessary to bring in outside information which may be inaccurate (e.g., census data which may be out of date) in order to make total estimates. For agriculture it has the further advantage of giving an estimate of the number of farms which is seldom known with satisfactory accuracy. If repeated samples are taken, the correlation that usually exists on sample areas from year to year can be utilized in the area method to reduce the sampling error by appropriate sample designs and estimation procedure. The gains available by such a technique have been shown by Jessen (7) to be of considerable value. Designs of this sort are not free of bias if the sample units consist of farms, because farms are not permanent through time as are areas, and farms coming into or going out of existence may differ greatly from those of more permanent character.

The area method has certain disadvantages which are more operational than theoretical. It requires maps of detail sufficient to allow the delineation of the sample areas which may be located easily and accurately by the enumerator in the field. At the present time such maps are not always available. It is, therefore, advisable in some parts of the country to supplement the best available maps with aerial photographs, especially when the smaller, more efficient unit is used in the non-surveyed regions of the country. These requirements lead to several specific difficulties. In the first place, there is no single source from which the best available maps may be obtained. This causes great expenditure of time and money to bring them together. In the second place, although a large part of the United States has been photographed by the Agricultural Adjustment Agency and other agencies, there is the additional cost and time required to obtain reproductions of the photographs.

Another disadvantage of the area method is that, given maps of sufficient detail, it takes time and money to delineate on maps areas that are efficient in size and that are bounded by features which can be clearly identified in the field. Consequently, a sample using the area method cannot be designed quickly and at the same time efficiently. These operational difficulties have in part led to the development of the Master Sample. The Master Sample at the present stage of its development is made up of 67,000 areas which identify about 300,000 farms (the areas for the United States averaging about five farms) located within nearly every county in the United States.

History of the Master Sample.--Early in 1943, the idea of a Master Sample occurred to Rensis Likert of the Bureau of Agricultural Economics. The BAE at that time was working on a group of bureau-wide projects. It was evident that there was need for a procedure which would provide effective samples for various studies and more particularly which would enable the accumulation of data relating to a representative group of farms. By drawing subsamples for different studies from a large Master Sample and systematically accumulating the data from these farms and farm families, many important interrelationships affecting farm production, income and living which cannot now be studied economically could be analyzed. Thus a Master Sample could become a device for integrating and improving the effectiveness of the research of the Bureau. Likert sought the advice and help of the writer in preparing a proposal for developing a Master Sample. Because of its demonstrated advantages the area method of sampling was recommended.

It was first decided in April 1943 that the Statistical Laboratory at Iowa State College would design a set of areas which would provide a national sample of about 5,000 farms. At this point the name "Master Sample" was applied and is still being used although the scope of the sample has been enlarged materially. The work was assigned to Iowa because the group there was experienced in such sampling as a result of the research work on sampling that has been carried on for several years cooperatively by Iowa State College and the Bureau of Agricultural Economics. As the greatly increased demand for this type of sample became apparent, the planned size was increased at first to 25,000 farms and later to 300,000 farms, as the Division of Agricultural Statistics became interested in the proposed Master Sample as a basis for taking large surveys of farm inventories.

At this time plans were being drawn in the Bureau of the Census for the 1945 Agricultural Census. The Bureau of Agricultural Economics was interested in having the Bureau of the Census identify the schedules of the farms in the Master Sample so the census information could be used for subsequent sampling purposes. The Bureau of the Census became interested in using the Master Sample in connection with the 1945 Agricultural Census because it would provide a suitable group of farms from which information for a supplementary schedule could be conveniently obtained. The identification of the Master Sample farms would make it possible to separate these schedules for making estimates prior to the total tabulations. Furthermore, the maps and the costly materials for reproducing thh maps necessary for the Master Sample as well as the aerial photographs could be used in the Agricultural Census. When it became evident that the Bureau of the Census could also derive benefit from the Master Sample, an agreement was drawn up whereby the Bureau of Agricultural Economics and the Bureau of the Census could combine their technical and administrative resources to develop the most efficient sample feasible. Under this agreement the Master Sample was completed in time to be used in the Agricultural Census as planned.

However, one more development has arisen which further enlarges the scope of the Master Sample. Shortly before its completion, the Bureau of the Census wanted to extend the sample so that it would represent the entire population. The areas set up for agriculture were not efficient for population samples in the cities and villages where the size of the sampling areas was determined by the number of farms rather than the number of persons. In the village and town stratum the Master Sample is being redrawn with smaller areas, in the open-country stratum, however, the present sample is satisfactory for either farms or population inquiries. This work is going forward at Iowa State College where the agricultural sample was drawn. The Bureau of the Census is at present drawing what may be called a "Master Sample of City Areas" to serve in city areas as the Master Sample does in the rural areas. The two projects are being designed as integrated parts of an over-all plan. Work on the extension of the Master Sample and the city sample is just getting started and the entire job will probably not be completed for another year. Priorities, however, are being given to a sample of counties that can be used shortly.

When the Master Sample is finally completed in its extended form, it will be available for the proposed 1946 sample census of population. In an attempt to strengthen its statistical program the Bureau of Agricultural Economics plans to use the Master Sample for its more difficult and important inquiries. Requests from state agencies to use the Master Sample within the states have been received.

Steps are being taken to put the administration of the Master Sample on a sound basis. The maps, photographs and other materials will be put under centralized control. A joint committee has been set up by the Bureau of Agricultural Economics and the Bureau of the Census to lay

broader policies for administering the Master Sample. A traffic problem has already developed and the committee is now working on its solution. The Bureau of the Census and the Bureau of Agricultural Economics are being staffed to advise and help in the designing of samples that use the Master Sample. Plans are being drawn jointly by the two agencies and Iowa State College for extensive research to determine the sampling errors of the Master Sample with different methods of estimation when used as a whole or when subsampled. Facilities will be provided to redraw the sample in the counties where the sample has been overworked.

In summary it may be said that the area method of sampling is an improvement over other designs because (i) it does not depend upon advance knowledge of the characteristics of the population, (ii) it eliminates freedom of choice on the part of enumerators, and (iii) it is usually efficient from the standpoint of maximizing precision on the basis of cost. Although the designing of area samples is time-consuming the initial cost high, usually it has been found to be justified. The Master Sample of agriculture will greatly facilitate area sampling by providing maps and other materials, tabulations, and other information helpful in the efficient design of area samples. It has been indicated that the Master Sample will be widely used in many different kinds of surveys for national as well as state inquiries.

II. DESIGN

By R. J. Jessen

Bureau of Agricultural Economics, and
Statistical Laboratory, Iowa State College

As the idea of the Master Sample developed, its scope and objectives suggested a sample with the following specifications:

(i) That it be large enough to provide an adequate group of farms on which information can be obtained that will be useful for selecting efficient smaller samples for special inquiries and coordinated studies.

(ii) That it be general or flexible enough in basic design so that it may be used directly, or with simple modification, for sampling not only farms but rural population, land use, crop diseases, crop posts, etc.

(iii) That it be distributed geographically to such an extent that all of the primary sampling units, of the kind now commonly used, comprising single counties or groups of counties, will likely have an adequate sample of areas within each convenient for subsampling.

(iv) That it be specific enough as to the population sampled so that it may be conveniently and precisely extended to inquiries which deal mainly with elements other than farms.

(v) That it be efficient enough so that it may justify its costs.

(vi) That it be practical enough so that estimates from it can be made without unnecessarily complicated procedures and that the field work can be simply and accurately carried out.

(vii) That it be completed in time so that it may be used as an integral part of the 1945 agricultural census.

The plan or design of a sample proposed to meet those specifications was developed during the last half of 1943, when the above general specifications and others were taking final form and when many of the more technical features were still being tested. The purpose of the present paper is to present as simply as possible a description of the design of the Master Sample.

General.--The Master Sample is a sample of small areas which average about 2.5 square miles in size but which vary according to location and other circumstances. Nevada has the largest average size: 108 square miles; Indiana the smallest: 0.71 square miles. The sample areas have been selected from every one of the 3,070 counties 1/ of the United States

1/ Actually some counties contained such few sampling units that occasionally a county would fail to have a selected area.

and contain within their boundaries approximately 1/18 of the land area of the United States, 1/18 of the farms (about 300,000), 1/18 of the rural population, etc.

By design, these small areas were formed to contain, within reasonable limits, a given desired number of farms or persons within reasonably identifiable boundaries. A big procedural job was the setting up of criteria for use in determining the size, shape, boundaries, etc., of these areas.

The primary strata.--Three primary strata were defined on the basis of incorporation and density of population. The total area of the United States was classified into: (i) incorporated areas, (ii) unincorporated areas relatively densely populated, (iii) unincorporated areas relatively sparsely populated. For convenience those area groups will be called the (i) incorporated, (ii) unincorporated, and (iii) open country strata.

The incorporated stratum, as the name implies, includes all incorporated cities and towns. In addition, unincorporated places regarded as "urban" by the Bureau of the Census were also included. Partly for desired flexibility of grouping and partly for sampling efficiency, all places in this stratum were further stratified into "rural" and "urban" groups according to Census rule, that is, those places having a population in 1940 of less than 2,500 persons are classed "rural" while those of 2,500 or more population are classed "urban".

The unincorporated stratum consists of all name places, outside of incorporated area, which have an estimated population of 100 or more and all other areas which appear on the map to have a population density of 100 or more persons per square mile. Estimates of population for these places were obtained from the Rand McNally Commercial Atlas, Bureau of the Census report on unincorporated places and the county highway maps. Places within the stratum were further stratified into the three size groups: (i) population 100-499, (ii) population 500-999, and (iii) population 1,000 and over.

The area remaining after the designation of the incorporated and unincorporated areas may be labeled "open country". This "open country" stratum contains 96 per cent of the land area (see table) and about 91 per cent of the farms of census definition. In this stratum an attempt was made to obtain a design satisfactory for sampling either or both farms and persons, whether farm or non-farm.

PERCENTAGE OF LAND AREA, SAMPLE AREAS, 1940 CENSUS FARMS, 1940 TOTAL POPULATION IN EACH OF THE THREE PRIMARY STRATA

Stratum	Land area (Per cent)	Sample areas (Per cent)	Census farms (Per cent)	1940 Population (Per cent)
Incorporated	1	3.4	3.4	63.6
Unincorporated	3	6.4	5.9	10.5
Open country	96	90.2	90.7	25.9
Total	100	100.0	100.0	100.0

With the strata so defined the next step was to designate them properly on the work map (which was a general highway and transportation map of each of the 3,070 counties). Corporate limits are definite boundaries and their location can be obtained. Populated areas or places that are not incorporated do not usually have definite boundaries. We therefore, set up some rules for establishing boundaries for such places.

(i) The place or area should be kept as compact in area as possible and yet should include everything that is distinctly "non-open-country" in appearance.

(ii) Boundaries should be chosen that can be easily identified in the field. Therefore, wherever possible, boundaries were made to follow roads, railroads, streams and to a less extent survey lines.

(iii) Where satisfactory "natural" boundaries did not exist but where aerial photography was available, arbitrary lines were used. Where aerial photography was not available, rule (i) was relaxed.

The selected boundaries were drawn on the work map in green so that the unincorporated stratum could be readily distinguished on the map from the other strata. Similarly, the corporate limits were outlined in yellow.

The primary strata defined in this manner have served and will serve several purposes:

(i) Farms in the more densely populated areas, which we have reason to believe differ importantly from those in the more open areas, are assured their proper proportion in the sample;

(ii) Those areas which contain large non-farm populations relative to their farm populations are accurately designated, permitting a convenient means for selecting sampling units in these areas suitable for population and related types of samplings; and

(iii) A reasonably efficient use was made of available information, such as that of the census for incorporated places and the map culture for the number of and location of the farms in the open country stratum--the unincorporated stratum being the one in which information on either farms or population was least available.

Sampling within the open country stratum.--The general highway and transportation maps which were obtained for all counties in the United States show, with variable degrees of accuracy, the location of farmsteads and other dwellings in the open-country areas and to some extent in the smaller unincorporated places. These farm and non-farm dwelling indications, which together with other structure indications are generally called by the mappers "culture", were used as a measure of size for determining numbers of sampling units to be assigned a given area and to control the size in terms of indicated farms and dwellings of each. To utilize this information efficiently for these purposes it was decided that the farm and non-farm culture be counted and tabulated in such a way that it could be conveniently used. A unit of counting was

therefore established. This is referred to here as the "count unit", which is a "natural" bound area, except where minor civil division boundaries were used, and which includes a minimum of six farms of eight dwellings and a maximum of about thirty farms. These limits were established for three reasons: (i) A larger number of natural bound area segments, when inspected on maps, appeared to contain from six to thirty farms; (ii) The amount of work required was held at a minimum because, as will be shown later, much time was saved in the delineation of sampling units; and (iii) The resulting unit is of convenient size for some kinds of sampling where prelisting and subsampling schemes are desired. The average size count unit for the United States contained 10.9 farms. Within each minor civil division as a stratum, natural bound area segments were combined (by appropriate markings) whenever necessary to bring the total count to the minimum six farms or nine dwellings. Each minor civil division having open-country area was assigned at least one count unit even if fewer than six farms were indicated. The operating ceiling of thirty farms was relaxed if adequate boundaries for subdividing an area segment were not available.

For each count unit, the number of farms and the total number of dwellings (including farms) were marked on the map within the count unit area. These counts were made for all open-country areas. Every square mile of area in this stratum is covered by some count unit or units regardless of whether or not farms were indicated on the map within that square mile. Each count unit was numbered serially beginning with number one in the upper right-hand corner of each minor civil division and in a serpentine manner (such as sections are numbered within a surveyed township) throughout each minor civil division. The minor civil divisions were numbered in a similar manner within the county.

With the counts made and designated on the maps the next step was to determine the number of sampling units to be assigned to each count unit on the basis of this count information. Giving consideration to the problem of finding on the maps sufficient identifiable boundaries for small areas, it was decided that three general regions be set up in the United States within each of which a different average size of sampling unit would be sought. In region number one, which was largely in the Corn Belt, it seemed feasible to seek an average of four farms with individual sampling units permitted to vary from two to six farms in size, that is, a tolerated variation of $\pm$ 2 farms. In region two, the areas outside the Corn Belt but still within the Public Land Survey System, an average of five farms $\pm$ 3 farms per sampling unit was thought feasible. In region three, which included all other areas of the country (those not covered by the Public Land Survey System), the goal was set at six farms $\pm$ 4 farms per sampling unit. In areas not covered by aerial photography an attempt was made to obtain an eight-farm average and the tolerance relaxed to $\pm$ 6 farms, although some cases were left to individual circumstances.

These sizes, together with the tolerances, were the results of an attempt to strike a suitable balance between the smaller area which is usually statistically more efficient and the larger area which is less costly to enumerate (per farm) and prepare for, and which generally has fewer problems of boundary identification. Although some study has been made of these problems (7), (8), a considerable number of new situations required the quick and ready answers of judgment, the correctness of which can and will be evaluated by studies now under way and by subsequent experience.

Because the culture indicated on the county highway maps was of variable accuracy (when checked with 1940 census data), it was thought advisable to check the extent of the inaccuracy so that the expected number of farms per sampling unit would be based on the 1940 census rather than on inaccurate culture. As an example, suppose the culture on a map under-represented census farms by 50 per cent. To avoid the possibility of getting sampling units that had twice the average number of farms desired, the number of sampling units assigned to each counting unit was therefore based on half rather than the full number of indicated farms. Another criterion for determining the size of the sampling unit was the extent and quality of identifiable boundaries considering, of course, the availability of aerial photos. In order also to provide for a satisfactory sample of population in this stratum, the expected number of dwellings per sampling unit was kept, wherever possible, to ten or less.

Tables were set up so that the clerk, knowing the regional location of the county, the culture count and its degree of inaccuracy, could select the suggested number of sampling units (and the permissible sizes of each) to be assigned to each count unit. This result was modified if inspection of the count unit on the map showed that an improvement in boundaries, accessibility or photo use resulted from such a change. When the number of sampling units for a counting unit was determined, it was recorded on the map alongside the counts. Map data on counts and number of sampling units were then transcribed to listing sheets on which data for each counting unit were listed individually. These data were then punched on machine cards and a tabulation made for each county showing the following: the state code, county code, the minor civil division code, the counting unit serial number, its farm count, its dwelling count, its assigned number of sampling units, the cumulative totals for the two counts and sampling unit numbers. Count units on the tabulation were ordered serially within minor civil division which in turn were serially ordered within the county.

With the cumulative totals of sampling units indicated on the tabulation sheet, it was a relatively simple operation to designate the count units which would contain the required sampling units. A random number between one and eighteen was chosen as the starting point and the count unit containing every eighteenth sampling unit was designated on the cumulative total column on the tabulation sheet. The scheme of numbering count units within minor civil divisions and minor civil divisions within the county made possible the ordering of count units on the tabulation in such a way that geographic distribution of the sampling units would be maximized by this kind of systematic selection. The count unit selected in this way was then located on the map, and its subdivisions

into the appropriate number of sampling units were so indicated that each sampling unit could be identified. Only those count units selected in the sample required actual, subdivision into sampling units. This saved considerable clerical time. If the number of sampling units within a count unit was large, each sampling unit was numbered; if small, a systematic scheme of counting was followed. The sampling unit for the sample was chosen by selecting a random number from one to the number of sampling units in the count unit. The selected sampling unit was appropriately marked on the map and coded for future reference.

Sampling within the unincorporated stratum.--In the unincorporated stratum we were concerned with selecting a sample of farms without regard to its adequacy as a sample of population. Information as to the number of farms expected to be found within each of the unincorporated places was not shown by the maps as adequately as it was for the open country. It was, therefore, necessary to set up some rather arbitrary criteria for assigning the number of sampling units to each of the unincorporated places. These rules were: (i) a number of sampling units was so assigned that the expected number of farms per sampling unit was six; (ii) A minimum of two sampling units was required for each unincorporated place regardless of how few farms were indicated. The number above this minimum was determined by the culture count of farms, with adjustment when it appeared advisable; (iii) The number of sampling units assigned by rules (i) and (ii) was modified when necessary to utilize most effectively identifiable boundaries shown by the map. In some cases it was necessary to make arbitrary decisions as to the number of sampling units a place should have because no culture at all was shown on the map. It was found that the culture shown around densely populated areas did not sharply discriminate the farm from the non-farm dwelling. In some cases where the unincorporated stratum included a large area, it was found convenient to sample it in a manner somewhat similar to that of the open country. In this case count units were established and counts (of farms only) made to help determine the distribution and number of sample units.

Lists were prepared which grouped the unincorporated places into three size groups: (i) from 100 to 499 population, (ii) from 500 to 999 population, and (iii) 1,000 population and over. In addition to the name of the place, this listing showed its population, its expected number of farms, the number of sampling units into which it would be divided, and the cumulative total of sampling units through the three size groupings. Taking a number between one and eighteen for a random start, a sample was selected systematically from the cumulative column of sampling units in much the same manner as was followed in sampling in the open-country stratum.

Sampling within the incorporated stratum.--As in the case of the unincorporated stratum, a listing was made of all incorporated places within each county, grouped into the two size groups--those under 2,500 population (rural according to the 1940 census) and those 2,500 and over (urban). From special tabulations provided by the Bureau of the Census the number of 1940 census farms for each incorporated place was

obtained. Consequently, there was much better information on the expected number of farms for the incorporated places than for the unincorporated places. Each incorporated place was assigned a number of sampling units so that the average number of farms per sampling unit would be approximately six. With the sampling units so determined, sampling followed the same procedure as stated for the unincorporated stratum. Where incorporated places included an especially large number of farms, we obtained (from the Bureau of the Census) special enumeration district information for the location of those farms within the city. It was therefore possible to divide the city by enumeration districts in such a manner as to get a fairly equitable number of farms for each sampling unit.

Use of aerial photography.--The relatively small sample areas used in the Master Sample were made possible by aerial photographs, both the individual prints and the reproduced indexes. These served as detailed maps helpful in providing identifiable boundaries for designating sample areas and for use by the enumerator in finding those boundaries in the field. The scale of the ordinary aerial photograph is three inches to the mile and can be increased by the enlargement process to provide the necessary detail for showing sample area boundaries to the enumerator. Proper boundary identification is crucial for the prevention of possible bias in the improper inclusion or exclusion of farms for sample enumeration. The aerial photograph provides many useful landmarks, such as buildings, fence lines, clusters of trees, creeks, etc., which even the best detailed maps do not show. Although in areas covered by the Public Land Survey the sampling areas are generally more easily identified from ordinary maps, it was felt that even here aerial photographs are worth their cost. In areas not covered by the Public Land Survey the photographs are almost indispensable if the smaller, more efficient sample area is to be used.

For all sample areas in the open country the standard contact print of three inches to a mile was used to designate the sample areas. In areas around villages and towns or within villages and towns, enlargements ranging in scale from 4 1/4 inches to 9 inches to the mile were used to provide the greater detail desired. The usual contact print measures either 7 by 9 inches or 9 by 9 inches. It covers, therefore, from 7 to 9 square miles of area. Although the majority of sampling areas required but one aerial photograph for complete coverage, on the average it was necessary to purchase about 1½ prints per sampling unit. Aerial photos were obtained for over 91 per cent of the sampling units in time for the 1945 Census of Agriculture.

Techniques of using the Master Sample.--Sampling areas provide a convenient flexible base for sampling different elements. For instance, if a selection of farms is desired, all that is necessary is that criteria be set up so that each farm is associated with but one and only one sample area. This can be easily accomplished in the majority of cases by regarding a farm as associated with a given sampling area if its headquarters falls within that area. Where farms do not have distinct headquarters, rules can be set up which will establish a sort of pseudo-headquarters suitable for those sampling problems. Other types of surveys may require some different criterion of enumeration. As an example, for certain crop estimates it may be desirable to include in the sample area only those fields or areas actually contained within the boundaries of the sample

area, regardless of where the boundaries of the farms within which they are located happen to be. In this case the aerial photograph is particularly useful because boundaries can be established exactly and the desired areas may be easily planimetered.

What the Master Sample is at present.--The Master Sample at present consists mainly of materials. These are of two kinds: (i) 3,070 count maps (one for every county in the United States) on which sample areas have been designated, providing, when enumerated, a sample of about three hundred thousand farms, and (ii) materials useful for drawing samples so that the present kind of sample can be extended or supplemented for any samples either of agriculture or of persons. The Master Sample does not yet have any data on the farms selected. The bulk of these data is expected to be obtained by the 1945 agricultural census and will be made available for sampling purposes soon after the field work has been completed.

The Master Sample materials at present consist of large scale (usually one inch to the mile) highway maps of all counties of the United States with information on them useful for sampling; listings of all towns, villages, cities and other densely populated places and areas having an estimated population of one hundred persons or more; tabulations of the expected numbers of farms and dwellings by small geographic areas (count units) and their cumulative totals; materials by which an unlimited number of county maps showing sample areas for a three hundred thousand farm sample can be reproduced (by means of the ozalid process); and serial photographs for about 91 per cent of the 67,000 sample areas of the Master Sample.

Additional valuable materials for subsequent sampling will be made available when the 1945 agricultural census is completed. The Bureau of the Census will take a standard farm schedule from all farms and an additional supplemental schedule for more detailed information on farms in the Master Sample areas and specially designated large farms. Both schedules taken from the Master Sample farms will be coded accordingly during the regular census tabulation procedure. When punched on cards, these data will be useful for drawing both the general and the specialized types of farm samples for subsequent inquiries. With this information it will be possible to direct more efficiently the sample for specific inquiries to those farms most relevant to the inquiry. For instance, an inquiry of dairy farmers can be designed such that the enumerator is directed primarily to those sampling units having dairy farms because we will know, from the census information which sample areas contained dairy cows at the time of the agricultural census. Changes taking place on the sample areas through time do not impair the validity of sampling if appropriate designs are used. In fact, using the sample to measure change directly is usually an efficient way to estimate changes in the whole population.

Present plans for extending the Master Sample.--At present the Master Sample is an adequate sample of both farms and population in the open country stratum only. Work is now under way to extend the sample so

that it will provide, by means of a supplemental selection of smaller areas with an average of 8 to 10 dwelling units per area, an adequate sample of persons in the unincorporated stratum. This work is being carried out jointly by the Bureau of the Census, the Bureau of Agricultural Economics and Iowa State College. The extension will follow the same general method of sampling, that is, the use of small areas clearly designated on maps, but the size of the unit will be measured in terms of total dwelling units rather than farms. In some areas where adequate maps and aerial photography are not available, it will be necessary to adopt a prelist-subsample scheme to reduce the number of dwelling units enumerated in any small sample area and therefore to increase sampling efficiency. The use of this device, however, will be held to a minimum because of its relative costliness. This work is designed to prepare a sample in unincorporated strata suitable for use in sample surveys of population, households and related subjects and for the proposed sample of population in 1946.

Other plants for the improvement and extension of the Master Sample, which can be briefly summarized here, call for:

(i) The completion of the charting of the areas covered by aerial photography and the completion of a set of indexes showing for any given area the appropriate print or prints covering that area. When this is done, the Bureau of Agricultural Economics will have the most complete and centralized source in this country of such information useful for obtaining aerial photography.

(ii) The use of 1945 agricultural census information for the classification of the farms on the Master Sample are as useful for future sampling inquiries. Perhaps a classification of areas will be made for similar purposes.

(iii) With the Bureau of Agricultural Economics, Bureau of the Census, and Iowa State College cooperating, a program of research is underway on sampling and estimation procedures which is designed to make the Master Sample more efficient and adaptable. This program will include a study of such problems as the efficient selection of counties or other primary sampling units, the study of schemes for subsampling the present sampling units, count units or combinations of both, etc., study of various estimation procedures, the study of designs involving matching sampling units on repeated visitations and various other sampling schemes.

(iv) Modification of the Master Sample to utilize new information, when it becomes available, to improve the present sample.

LITERATURE CITED

(1) Anderson, Oskar, "Description of the Bulgarian agriculture sample census, "Bulletin de Statistique", Publ. Direction Gen. de Statistique de Bulgaria. No 8, 1934. (Summary in French.)

(2) Frankel, Lester R., and Stock, J. Stevens. "On the sample survey of unemployment", this Journal 37: 77-80. 1942

(3) Gini, Corrado. "Une application de la methode representative aux materiaux des dernier reconsement de la population italienne," Bulletin de l'Institut International de Statistique 23(2): 198-215. 1938.

(4) Hansen, Morris H., and Hurwitz, William N. "On the theory of sampling from finite populations," Ann. Math. Statis. 14: 333-362. 1943

(5) Hoijer, Ernst. "Organization of the 1935-36 extraordinary census of Sweden; "Iowa Agri. Exp. St. Mimeo. Series" No. 3702. 1937.

(6) Holmes, Irvin. Research in sample farm census methodology. Part 1: "Comparative statistical efficiency of sampling units smaller than the Minor Civil Division for estimating year-to-year change." U. S. Department of Agriculture, Agricultural Marketing Service. 1939.

(7) Jessen, R. J. "Statistical investigation of a sample survey for obb- obtaining farm facts," Iowa Agri. Exp. Sta. Res. Bull. 304: 1-104

(8) Jessen, R. J., and Houseman, Earl E. "Statistical investigation of farm sample surveys taken in Iowa, Florida, and California, "Iowa Agri. Exp. Sta. Res. Bull. 329: 263-338, 1944

(9) King, A. J. and Simpson, G. D. "Now developments in agricultural sampling, "Jour. Farm Econ." 22: 341-349. 1940

(10) Mahalanobia, P. C. "On large-scale sample surveys", Philosophical transactions of the Royal Society of London. Series B. No. 584 231: 329-451. 1944

(11) Neyman, Jerzy. "On the two different aspects of the representative method, the method of stratified sampling and the method of purposive selection, "Jour. Roy. Statis. Soc." 97: 558-606. Discussion, pp. 607-625. 1934.

(12) Shepard, J. B. "Selection of areas for sampling agricultural enumeration, "Jour. Farm Econ." 19: 454-464. 1937

(13) Strand, N. V., and Jessen, R. J. "Some investigations on the suitability of sampling Iowa Agriculture," Iowa Agr. Exp. Sta. Res. Bull. 315: 616-650. 1943.