1978 CENSUS OF AGRICULTURE - AREA SAMPLE DESIGN AND METHODOLOGY

David D. Chapman and Charles E. Rogers, U.S. Bureau of the Census

BACKGROUND

The 1978 Census of Agriculture is the twenty-first nationwide Census of Agriculture. The first agricultural census data were collected in 1840 as part of the sixth Decennial Census of Population. Between 1850 and 1920 an agriculture census was taken every 10 years in connection with the Census of Population. Beginning with 1920, an agriculture census has been taken every five years.

Data collection for each Census of Agriculture prior to 1969 was based on a canvass of rural areas by enumerators. These enumerators either located and interviewed farm operators (prior to 1950) or collected and reviewed completed forms previously mailed to farm operators (1950, 1954, and 1959). In 1969, and continuing in 1974, the basic method of enumeration was changed to a mail-out, self-enumeration, mail-back procedure. The names and addresses for the mailing list were obtained from administrative records of the Internal Revenue Service, U.S. Department of Agriculture, other government agencies, and agriculture-related associations and organizations. The compilation and combination of multiple source lists into one final mailing list results in duplication and in information approximately two years old before use. The mailing list often omits current farm operators.

Coverage evaluation studies of the 1969 and 1974 censuses indicate the present method of compilation and combination of source lists is not adequate to assure complete coverage of farm operators. While the aggregate value of agriculture products contributed to the economy by missed farms is small, the total number of missed farms and the total number of missed farms reporting specific types of products is significant. The net undercoverage estimated in the last five censuses is given in Table 1.

Table 1.--Estimated net census farm undercoverage in percent in Census of Agriculture

Characteristic	Year				
	1974[1]	1969	1964	1959	1954
Number of Farms	11.5	15.0	11.3	8.4	8.1
Land in Farms	7.4	9.1	6.1	6.0	5.4
Value of Product	2.9[2]	3.3[2]	2.9	(NA)	(NA)

[1] Farm definition was changed in 1974. The estimated net census farm undercoverage with a farm definition comparable with the 1969 census is 14.3-percent.

[2] Estimated value of product for missed farms only.

As a result of the coverage problems of the 1969 and 1974 censuses, a major goal in planning the 1978 census was the reduction in the number and effect of missed farms.

The 1978 Census of Agriculture Area Sample is proposed as a supplement to the mail list portion of the census. The primary objective is to estimate the number and characteristics of census farms not included on the 1978 mailing list for all states except Alaska and Hawaii. In the absence of information on missed forms, the survey is designed to estimate the number of farms and their characteristics with a fixed reliability.

1978 Mailing List

Construction of the final 1978 mailing list for the Census of Agriculture is divided into the three basic operations:

(1) Construction of a preliminary list of names using source lists available in March 1978.

(2) A Farm and Ranch Identification survey of selected names and addresses from the preliminary list to delete nonfarm operators and add tenents and successor farm operators.

(3) Construction of the final mailing list in October 1978 using the Farm and Ranch Identification survey, and additional lists not available previously.

Since both preliminary and final mail lists are the result of combining multiple lists of names and addresses, substantial duplication is expected. Unduplication of these lists is conducted through a computer comparison and/or clerical review of records using 1) Employer Identification Numbers, 2) Social Security Numbers, truncated surnames, and 3) coded names and addresses. This unduplication process is used to create both the preliminary and final mailing list. Initial multiple source lists contributed over 13 million names and addresses prior to unduplication. The unduplication process reduced this to a preliminary mailing list of approximately 6 million names. The portion of the preliminary list identified as having a high probability of containing nonfarm operators was selected for inclusion in the Farm and Ranch Identification survey. In the Farm and Ranch Identification survey each selected name and address was mailed a short questionnaire to determine if it represented a farm operation. Individuals identified as not being farm operators are to be excluded from the final mailing list. Farm operators from the Farm and Ranch Identification survey, individuals from the preliminary list not included in the Farm and Ranch Identification survey, and new source lists will be combined to construct the final mail list

of less than four million names. The construc-
tion of a list of farm operators is an inexact
process. Names and addresses may be deleted in
error, duplicates remain, and some names and
addresses of farm operators never occur on any
source list.

The final 1978 Census of Agriculture mail-
ing list will be used to construct labels to
mail questionnaires to farm operators about
January 1, 1979. This initial mailing will be
followed by five additional mailings of forms
and/or letters to nonrespondents at approxi-
mately three-week intervals. Large operators
not responding to mail questionnaires will be
enumerated by telephone. A sample of the remain-
ing nonrespondents will be selected and enumer-
ated by telephone. Information on the sample of
nonrespondents will be used to adjust the mail
list estimates for nonresponse.

Farm operators on the mailing list may have
their residence and mailing address anywhere.
Table 2. gives the reported location of farm
operator's residences from the 1974 Census of
Agriculture.

Table 2.--Residence of census farm operators in
1974.

Residence	Percent
On Farm Operated	80
On Another Farm	5
Rural Area Not on Farm	4
City, Town or Urban Area	11

Evaluation studies indicate farms missed in the
Census have approximately the same pattern of
residence as farms included in the Census.

SAMPLE DESIGN

The 1978 Census of Agriculture Area Sample
is a stratified one-stage cluster sample. Clus-
ters are selected independently from each stratum
with equal probability using a systematic
sampling procedure. Estimates will be made of
stratum totals using either unbiased or ratio
estimators and accumulated to population totals.
A ratio estimator is being considered where the
auxiliary variable is the estimated number of
either housing units or farms in the segment.
After data collection a detailed investigation of
the relative efficiency of the ratio estimator
and an unbiased estimator will be conducted. The
precision of the estimator will be estimated
assuming all clusters are selected without
replacement using simple random sampling within
strata.

Target Population

The target population for this survey is all
census farms with operators residing in places of
less than 2,500 population in 1970. The elemen-
tary population unit is the census farm. A
census farm is defined as a place on which agri-
cultural operations were conducted at some time
in the census year, under the day-to-day control
or supervision of one person or partnership, and
from which $1,000 or more of agricultural products
were sold or would normally be sold during the
census year. An operator, who exercises control
through ownership, management, a lease, or a
rental cropping arrangment is defined for each
farm. The reporting unit for the farm is this
farm operator.

The goal of the mail list is to identify
farm operators, using administrative lists; while
the goal of the Census of Agriculture Area Sample
is to select a probability sample of all land in
areas of less than 2,500 population in 1970 and
all farm operators living on that land.

The area sample will be selected only from
areas classified as places of less than 2,500
population in the 1970 Census of Population and
Housing. This will include a high proportion of
farm operators in the survey while contacting a
relatively small proportion of households. While
not all Census of Agriculture farm operators were
classified as farmers or farm managers in the
1970 Census, the distribution of farmers and farm
managers relative to households in the US (shown
in Table 3) illustrates the potential cost
savings of excluding farm operators in urban areas
from the population to be sampled.

It is estimated that approximately 13-percent
of all farm operators will not be included on the
mailing list. Since the mailing list is construc-
ted from administrative lists, whether or not a
farm is missed should be relatively independent
of its physical location (this is supported by
earlier coverage evaluation studies). An estima-
ted 7-percent of farms are located in places of
2,500 or more population.

The expected relationship between location
and presence on the mail list is given in Table 4.

Table 4.--Estimated proportion of farms by
location, by included or not included
on mail list and by location and mail
list.

Location	Mail List		
	Included	Not-Included	Total
Rural	.8091	.1209	.93
Urban	.0609	.0091	.07
US	.87	.13	1.00

Excluding places of 2,500 or more population from
the survey is expected to exclude approximately
one percent of the farms from the combined mail
list and area sample. However, by excluding these
places, major urban areas are avoided and 75
percent of the housing units in the country can
be eliminated from the survey resulting in a major
savings in costs.

Table 3.--Distribution of Occupied Housing Units and Farmers and Farm Managers in 1970 Census of Population and Housing

	Number			Percent	
Location	Occupied Housing Units	Farmers & Farm Managers	Occupied Housing Units Per F&FM	Occupied Housing Units	Farmers & Farm Managers
Urban[1]	47,672,276	101,102	417.5	74.9	7.1
Rural[2]	15,965,445	1,317,644	12.1	25.1	92.9
TOTAL	63,637,721	1,418,746	44.9	100.0	100.0

[1] Places with 2,500 or more population 1970

[2] Places with less than 2,500 population in 1970

An independent household survey will be used to select a sample of farm operators in urban areas, places of 2,500 or more, missing from the mail list.

SAMPLE UNIT

The sample unit in the 1978 Census of Agriculture Area Sample is a defined geographic area of land. The size of the sample unit in geographic area, number of housing units or number of farms varies with the stratum to which the segment is assigned. Within a stratum sample units are constructed to have an equal number of farms.

The open segment concept is used to associate farm operators with a sample segment of land. A farm will be included if the operator lives inside the segment. Each farm operator and the farm operated will be uniquely associated with a single segment. The number of farms in a segment was selected based on the estimated farm density of the area. The desired number of farms per segment is given in Table 5. by estimated farm density. The estimated farm density is the ratio of the estimated number of farms and the number of occupied housing units.

Table 5.--Desired segment size in farms and maximum segment size in housing units by estimated farm density.

	Segment Size	
Estimated Farm Density	Desired Number of Farms	Maximum Number of Housing Units
.1 and above	12	120
.05 to .1	10	200
.02 to .05	5	250
.01 to .02	2	200
.005 to .01	1	200
Less than .005	0	150

The desired segment size based on estimated farm density keeps the size of segments to less than 250 housing units. In the area of highest farm density, the number of housing units was reduced to equalize the workload due to the large number of farms. For the State of Kentucky, the relationship between expected number of farms and the expected number of housing units in a segment is given in Table 6. by estimated farm density.

Table 6.--Desired segment size in farms, average segment size in farms and housing units in Kentucky by estimated farm density.

		Segment Size Average	
Estimated Farm Density	Desired Farms	Farms	Housing Units
.1 and above	12	11.9	32.5
.05 to .1	10	9.5	130.0
.02 to .05	5	5.1	157.3
.01 to .02	2	1.8	125.0
.005 to .01	1	1.0	124.3
Less than .005	0	.004	127.23

Stratification

The purpose of stratification is to divide the sample units into groups with similar characteristics. In the 1978 Census of Agriculture Area Sample, units are stratified by two criteria, geography and agricultural intensity. Each of the 48 conterminous States is a domain of study for which estimates are to be published and each was made a separate stratum. Within each State, enumeration districts are stratified by an estimated farm density. Since no estimate of the number of farms is available below the county level from previous Censuses of Agriculture, the number of farms in each enumeration district was estimated as the larger of either rural farm

housing units or farmers and farm managers in the 1970 Census of Population and Housing. Table 7. shows the distribution of enumeration districts by estimated farm density for the State of Kentucky.

Table 7.--Percent and cumulative percent of rural enumeration districts in Kentucky by estimated farm density.

Estimated Farm Density	Percent	Cumulative Percent
Less than .01	33.7	33.7
.01 to .05	11.8	45.5
.05 to .15	11.8	57.3
.15 to .25	7.6	64.9
.25 to .35	9.3	74.2
.35 to .45	9.5	83.7
.45 to .55	8.2	91.9
.55 to .65	5.0	96.9
.65 to .75	2.1	99.0
.75 to .85	.7	99.7
.85 to .95	.1	99.8
Greater than .95	.2	100.0

All enumeration districts and consequently all segments are divided into one of six strata based on estimated farm density. The six strata are defined below.

Strata	Estimated Farm Density
1	.1 and above
2	.05 to .1
3	.02 to .05
4	.01 to .02
5	.005 to .01
6	Less than .005

The same six strata are used within each State. The distribution of the population, housing units, and farms to strata will differ from State to State. The distribution of segments, housing units and farms by strata in the State of Kentucky is given in Table 8.

Sample Size and Allocation to Strata

The primary objective of the Area Sample survey is to estimate by State the characteristics of farms not on the mailing list. Considering this objective, information useful in determining sample size for each State is generally not available. Little is known concerning the geographic distribution (to the level of enumeration districts or segments) of farms missing from census mailing lists. Sampling errors were calculated from historic data for major data items on all farms using various assumptions concerning intracluster correlations. These calculations and an overall cost consideration were used to determine overall sample size (approximately 6,000 segments) and the allocation to States. The allocation to strata within States was approximately optimum for between enumeration district variance using number of farms estimated from the 1970 Census of Population and Housing. The allocation of segments for Kentucky is given in Table 9.

Sample Selection

The sampling frame for the 1978 Census of Agriculture Area Sample is a list of all enumeration districts from the 1970 Census of Population and Housing and their characteristics and maps. The list of the enumeration districts and their characteristics was constructed using the Fifth Count (File C) Summary Tapes from the 1970 Census of Population and Housing.

The Fifth Count (File C) Summary Tape contains a list of all enumeration districts with geographic and administrative codes from the 1970 census, counts of population, housing units, number of farmers and farm managers, and farm laborers and farm foremen. This information is used to assign enumeration districts to strata and to estimate the number of farms and housing units in the segment.

Prior to sample selection, the frame of enumeration districts was sorted by Census of Agriculture County Code, Census Tract, and Enumeration District. Agriculture Census codes are constructed to arrange the counties in geographic order. In general enumeration districts are numbered in a serpentine fashion beginning with the northeast corner. This sorting arranges the enumeration districts into an approximate geographic order. In theory the sample selection is a one-stage process. In practice, two steps are used. The number of segments located in an enumeration district is determined by dividing the expected number of farms in the enumeration district by the desired number of farms per segment for the stratum. Within each stratum an array is created which

Table 8.--The number of segments, the number and percent of occupied housing units and the number and percent of farms by strata in Kentucky.

Strata	Segments	Occupied Housing Units		Estimated Farms	
		Number	Percent	Number	Percent
1	9,372	304,879	52.6	111,860	95.9143
2	290	37,694	6.5	2,742	2.3511
3	276	43,411	7.5	1,407	1.2064
4	284	35,495	6.1	518	.4442
5	93	11,545	2.0	93	.0797
6	1,155	146,931	25.3	5	.0043
Total	11,470	579,955	100.0	116,625	100.0000

Table 9.--Number of segments, estimated strata standard deviation, relative strata (Neyman) allocation, expected sample size, and sample interval by strata in Kentucky.

Strata	Number of Segments	Strata Standard Deviation	Relative Strata (Neyman) Allocation	Expected Sample Size	Sample Interval
1	9,372	.55	.877	156.12	60.03
2	290	1.60	.079	14.06	20.62
3	276	.59	.027	4.8	57.50
4	284	.16	.007	2	142.00
5	93	.001[1/]	.00002	2	46.50
6	1,155	.04	.0078	2	577.50
Total	11,470		1.00000	180.97	

[1/] Zero -- for allocation set equal to .001

contains the cumulative number of segments by enumeration district. A random start is selected independently for each stratum. The stratum random start will be successively incremented by the sampling interval to identify the segment number of the sample segment selected. The combination of a systematic sampling plan and sorting the frame in an approximate geographic order produces a sample in each stratum which is distributed uniformly throughout the State. Sample selection will identify the segment number and the enumeration district of each sample segment. The actual physical boundaries of the segments are not, however, identified.

The actual identification of the area of land to be enumerated is accomplished using enumerator maps from the 1970 Census of Population and Housing. These enumerator maps spot the location of housing units. Using these spotted housing units, the enumeration district will be divided into segments of equal size with recognizable boundaries. These segments are numbered serially in a serpentine fashion beginning in the northeast corner. After numbering, the selected segment will be identified and a special enlarged enumerator map will be prepared.

DATA COLLECTION

The data collection phase of the 1978 Census of Agriculture Area Sample survey is separated into three distinct parts: (a) enumeration, (b) quality control, and (c) evaluation. Enumeration refers to the procedures used for locating and collecting information from the respondent. The quality control refers to the methods used prior and during enumeration to control the nonsampling error in enumeration. Evaluation is to measure the magnitude and effect of nonsampling error on survey estimates.

Enumeration

During enumeration the enumerator will locate the boundaries of the area segment using a map of the segment. The enumerator will locate all dwelling units in the segment. The head of each household or group quarters and vacant structures will be listed in a record book. A series of

preliminary or screening questions will be asked of each head of household for every member of a household or group quarters. All individuals for which a "yes" response is received to one or more screening questions, will be considered a "potential farm operator". An additional series of questions will be asked of all "potential farm operators" to uniquely identify an operator for each farm and to determine alternate names and individuals associated with the farm.

For households with no one at home the enumerator will attempt to obtain enough information from neighbors or others to contact the head of the household by telephone. When necessary the screening interview may be conducted by telephone. Complete farm information will be collected by personal interview.

Timing

The field data collection phase of the survey will begin in late September and continue through mid-November. Supervisory crew leaders and crew leaders personnel will be recruited and trained in September. Enumerators will be recruited in late September and enumerator training will begin in early October with enumeration starting in mid-October and continuing through mid-November. Enumeration will be completed prior to mail-out of the regular census in January 1979.

Field Coordination

Approximately 1,700 enumerators will be hired to survey approximately 6,400 segments. The field organization for the survey will be to have approximately 1,700 enumerators, 215 crew leaders, and 27 supervisory crew leaders. A crew leader will supervise an average of 8 enumerators; A supervisory crew leader will supervise an average of 8 crew leaders. Enumerators will be recruited from the general area in which the segment is located. Experienced enumerators will be used when possible; but most enumerators are expected to have no previous survey experience. Each enumerator is required to pass a general examination prior to selection and will receive one week of training prior to beginning enumeration. The crew leader is responsible for training enumerators, observing and monitoring their work, and reviewing the finished work. The average crew

leader is expected to be an individual possessing experience with the Bureau of Census's current surveys but with no previous supervisory experience. Each supervisory crew leader is expected to be an experienced enumerator with supervisory experience.

Questionnaire

The questionnaire used in the 1978 Census of Agriculture Area sample is similar in form and content to the sample questionnaire used in the mail-out/mail-back portion of the census. The order and general content of the questions is the same. Question wording has been adapted for collecting information by personal interview. Additional information is to be collected on individuals associated with the operation or to assist in matching the questionnaire to the mail file.

Quality Control

Quality control activities are designed as an integral part of enumeration. Quality control in data collection is to be conducted in two distinct parts: (a) prior to enumeration and (b) during enumeration.

Prior to Enumeration

Before enumeration begins a quality control sample of segments will be selected from each state. The enumerator's initial assignment will be a quality control segment. The crew leader will visit each quality control segment and pre-list approximately 15 households in the segment. Screening questions will be asked of each household to be prelisted. After the enumerator completes the segment, the crew leader will match the sample of prelisted households to the enumerator's complete list of households. The number of missed housing units and "potential" farm operators will be compared. Unacceptable segments will be corrected. Enumerators with high error rates will be retrained or released.

During Enumeration

During enumeration two types of quality control procedures will be used. One will be personal observation. A crew leader, supervisory crew leader, or regional office personnel will observe the enumerator to correct errors, discuss problems, and answer questions. The other is to review completed work in detail to see that segment maps, listing books, and questionnaires are being completed correctly. Problems discovered will be discussed with the enumerator and, when appropriate, the crew leader.

Evaluation

After enumeration is completed, a Post Enumeration Survey, PES, is proposed. A sub-sample of segments will be selected to evaluate the quality of the survey. Each PES sample segment will be re-listed using the best enumerators from the area sample and using more intensive

procedures. The results from a re-listed segment will be compared to the original enumeration. The difference between the enumeration and the post-enumeration will allow estimates of the number and characteristics of missed and duplicated farms.

The enumeration of the PES sample is proposed to begin in December after the regular enumeration is completed.

DATA PROCESSING

When enumeration is complete and enumeration materials, questionnaires, record books and segment maps are returned from the field, completed segments will be processed to estimate the desired characteristics. The processing of information in the Census of Agriculture Area Sample will be divided into six operations:

(1) Receipt, check-in and filing of enumeration materials

(2) Matching Area Sample questionnaires to the mail list

(3) Clerical review, geocoding and keying

(4) Adjustment for nonresponse

(5) Tabulation and Publication

(6) Variance Estimation

Each step contributes to processing an individual respondent's questionnaire into estimates of the number and characteristics of farms not included on the mail list.

Receipt, Check-in and Filing

As segments are completed by enumerators, each segment and questionnaire in the segment will be reviewed in the field by supervisory personnel and then forwarded for processing. Upon receipt of segment materials, they will be reviewed for completeness. When segments are considered complete, the questionnaires will be matched to the mail list.

Matching

Since the primary objective of the 1978 Census of Agriculture Area Sample is to estimate the number and characteristics of farms not on the mail list, each completed questionnaire and the name of each nonrespondent and not-at-home must be matched to the mail list and classified as "matched" or "not matched." The matching procedure is of major importance in estimating characteristics of farms missing from the mail list. Operators not matched to the mail list will be considered missed from the mail list. Matching will be organized into two parts: (1) Mail List Search and (2) Self-Matching. All area sample respondents will be matched in November and December to the complete Census of Agriculture mail list by comparing reported names and addresses

on a questionnaire to a microfilm copy of the mail list. Each area sample respondent will be classified a match or nonmatch. All respondents matched to the mail list will be identified and not contacted in the follow-up program.

It is expected that most respondents will match themselves to the mail list. When enumeration at a household is completed, the enumerator will leave an adhesive label and a letter of instruction. The label will contain the name, address, segment number, and listing number of the respondent. The letter will instruct the respondent to attach the label to the census questionnaire he or she will receive by mail in January 1979, and to return the questionnaire without completing it. Area sample respondents who return a mail questionnaire with label attached will be classified as matched to the mail list.

During the matching operation, identifying information -- such as alternate farm names and additional individuals associated with the farm-- collected during enumeration will be used to match individuals to the mail list. After the microfilm search and sticker match is completed, all area sample respondents will be classified as matched or not matched to the mail list. Questionnaires for those included on the mail list will be inserted into mail list processing and treated as mail list respondents. Questionnaires for those not matched to the mail list will be classified as not included on the mail list and used to make estimates for each State of the characteristics of farms not included on the mail list.

Clerical Review, Geocoding and Keying

After matching, each questionnaire enumerated in the area sample will be assigned to a state and county for tabulation. Farms not matched to the mail list will be assigned to a state and given "Psuedo" county code for tabulation. Since unmatched farms will be used in the estimates only at the state level, the "psuedo" county will estimate the number and characteristics of farms in the state not on the mail list.

After matching and geocoding, all questionnaires will be clerically reviewed and keyed.

Nonresponse

Nonresponse refers to the failure to obtain information on an item of interest. Complete nonresponse is the failure to obtain any information from the farm operator. Item nonresponse refers to obtaining information on some but not all the items for the farm operation. Once keyed, all questionnaires, both matched and not matched, will be edited to detect missing answers or errors by the respondent for major items. The values of missing or incorrect information may be imputed. After editing and imputation, the questionnaires will be adjusted for complete nonresponse. Adjustments for complete nonresponse may be made separately depending on whether the nonresponding

operator was classified as on the mail list or in the area sample but not on the mail list.

Tabulation and Publication

Information on all farms not on the mailing list will be expanded and tabulated by state. Data will be tabulated as a "psuedo" county. "Psuedo" county estimates and mail list county estimates will be combined to give state estimates. State estimates will be combined to produce division, region and US estimates.

In publications, estimates for the number and characteristics of the farms not on the mail list will be listed as "Not allocated to County."

Variance Estimation

After tabulation, the sampling error for information published for the "psuedo" county and the state will be estimated. Sampling errors for estimates of number and characteristics of the farms not included on the mail list will be published with the final State estimates.

SPECIAL STUDIES

A series of special studies will be conducted during the 1978 Census of Agriculture Area Sample to provide basic planning information for similar surveys and to evaluate alternative statistical methods for estimating farm characteristics. The data will be used to evaluate components of variance and to explore the effect of segment size on survey cost and sample variance. Information will be collected on the cost of enumeration for use in future surveys. A special study is anticipated on the statistical efficiency and practicality of the open, closed and weighted segment methods of estimation. Methodology for making substate estimates using a synthetic estimator or other procedures may also be studied.