ACKNOWLEDGEMENTS

This paper stems from the efforts of Ralph Gann, Chuck Hudson, Fred Vogel, Richard Allen, Bob Cole, and others who wrote letters and documents early in the integrated survey program to explain its purpose and progress. Parts of a paper by Mr. Hudson to producers entitled "Strengthening Crop and Livestock Estimates Through An Integrated Survey Approach" have been incorporated here. Thanks go to Bessie Johnson for turning batches of pages into a typed report.

CONTENTS **PAGE**

ILLUSTRATIONS

INTEGRATED AGRICULTURAL SURVEYS

by R. R. Bosecker

INTRODUCTION: This paper has been prepared to describe the integration of agricultural surveys by the National Agricultural Statistics Service (NASS).

The purposes of integrating independent surveys were to:

1) provide the agricultural community with statistically reliable crop and livestock statistics based on uniform probability survey sampling methodology, and

2) reduce the number of individual producer contacts by consolidating major commodity surveys.

The concept to accomplish these objectives was known as the Integrated Survey Program. To fully illustrate the importance of this survey and sampling program, this report presents a background of survey methodology used by NASS, discusses advantages and disadvantages of integration, and then describes the key features of the integrated surveys.

DEVELOPMENT OF NASS SURVEY PROCEDURES: Estimates of harvested acreage and production for most field crops and annual livestock inventories have been provided since 1866. Large-scale non-probability mail surveys were employed from that beginning during the Lincoln Administration as a satisfactory means of obtaining agricultural data. These large non-probability mail surveys worked well as a means of obtaining agricultural data when U.S. agriculture was less specialized than it is today. Since most neighboring farms grew the same crops and frequently had similar livestock and poultry inventories as well, characteristics of respondent and nonrespondent farms were not greatly different.

From these large mail surveys, each State developed a list of reporting farmers representing all areas across the State to provide month to month dairy and egg production. These were known as Monthly Farm Report Panels. The monthly panel surveys and large non-probability annual mail surveys were the backbone of agricultural estimating methods during the first half of the 20th century.

During the 1940's and 1950's, as farms became larger and more specialized, non-probability mail survey techniques became less

reliable, particularly for livestock, poultry, and other specialty commodities. Respondent and nonrespondent farms were likely to be greatly different. It became apparent that improved methodology involving scientific probability sampling procedures were needed. Iowa State College, now Iowa State University, began development in 1943 of a land area based sampling frame known as the "Master Sample of Agriculture" (see King and Jensen, 1945). This effort was funded by a cooperative agreement between Iowa State, the U.S. Department of Agriculture and the U.S. Bureau of the Census.

The Master Sample was used extensively in the late 1940's and early 1950's for probability surveys but not as the basis for a recurring statistical program due to lack of funds. In 1954, with the appropriation of funds for methods research and development, a mid-year area frame survey of planted acres and livestock inventory called the June Enumerative Survey was begun on a pilot basis in 10 states (703 sample segments in 100 counties). Research on an end-of-year December Enumerative Survey, primarily for livestock data, began in 1955 using a subsample of land parcels from the June survey.

Area Frame Sampling provided true probability estimates and the ability to calculate sampling errors for the first time. It was possible to select sample units with a known probability of selection. The concept of area-frame sampling is simple. The land area to be surveyed is divided into small units called segments, with unique and identifiable boundaries that can be delineated on aerial photographs or maps. These segments encompass the total land area of each State as well as the U.S. A random selection procedure is used so all areas of agriculture are represented and the sample properly estimates for the populations of interest.

In 1965, the June and December enumerative surveys had been extended to all 48 conterminous states. The area sample consisted of about 17,000 segments and included all or parts of 70,000 farms. Most of the country still relied upon the Master Sample frame for sample segments, but new frames had been constructed for western and northeastern states. These new area frames had utilized stratification based on land-use categories according to intensity of cultivation or urbanization. Visual interpretation of aerial photography, provided the means to stratify according to land-use. Reduction of 10 to 25 percent in sampling variance were afforded by the new frames. The program to update area frames was made continuous.

It did not take long during the research phase of area frame sampling to realize that a few large livestock operations could distort survey results by their presence or absence in the sample. Their presence contributed greatly to the sample variance as well as the survey expansion. Therefore, a list of these "extreme operators" was developed independently of the area work. A list sample could then be selected, and the livestock data for extreme

operators removed from area samples. This approach was the beginning of multiple (dual) frame sampling for agricultural surveys. Theoretical foundations for the NASS estimates from two overlapping sampling frames were provided by Hartley (1962) and Cochran (1964).

In 1969, four states began a continuous series of multiple frame surveys utilizing extensive lists stratified by size for hog and cattle estimates. By 1974, multiple frame surveys were conducted in 14 states to provide hog estimates and in 28 states for cattle estimates. Multiple frame surveys using these extensive list frames provided significantly lower sampling errors than surveys using only small lists of extreme operators.

Various lists of farm operators were compiled and maintained within State Statistical Offices during the 1970's. Multiple frame surveys for rice (1977) and farm grain stocks (1979) were added to those for hogs and cattle. In 1979, work began to develop a comprehensive national list frame system that would standardize the maintenance of names and associated size data for farms and ranches throughout the U.S. By 1981, a national list had been consolidated in one computer system. Maintaining the list frame remained the responsibility of individual State offices. Sampling of the list frame for national multiple frame surveys became centralized in headquarters.

Multiple frame sampling has some distinct advantages for NASS, particularly for livestock, specialty crops, and economic data such as agricultural labor. These items are poorly correlated with land area and are inefficiently estimated by the area frame sample. In multiple frame sampling, most of the data for the population of interest can be collected more efficiently through the list frame by mail or telephone. Also, it is usually possible to develop and incorporate historic information for units in the list frame for stratification and sampling purposes. The area frame portion of the survey provides an estimate for incompleteness in the list. In this way, the two frames can be used to complement each other and provide a valid estimate for all operations. An excellent presentation and discussion of the various estimators employed by NASS are given in "Review of the Multiple and Area Frame Estimators" by Nealon, 1984.

Since implementation of the multiple frame concept, revisions to cattle estimates at the national level have never been over 1 percent and revisions to semi-annual hog estimates at the U.S. level have been over 2 percent only 3 times since 1971.

INTEGRATION OF SURVEYS: In 1983, a long-range study entitled "Framework for the Future" was completed. This study strongly stated that the first building block for all future activities of NASS must be a defined set of statistical standards and that the

agency should develop and operate its statistical program under these guidelines. It further stated that once the standards were defined, the policy should be to organize all survey, estimation, and publication efforts to most efficiently meet these standards. Since multiple frame sampling provided standards for evaluating individual survey estimates, the study called for expanding this type of measurable statistical standard to all agricultural estimates in an integrated survey program.

The "Framework" report recommended -- "the integrated survey program will involve the development of a large 'omnibus' survey that will provide baseline estimates and a basis for subsampling for subsequent surveys. This is an extension of the developmental work that has been going on for the past twenty-five years with the use of the area frame and the development of the list frame system. The integration of area and list frames leads to further development of a statistical program based on multiple frame sampling."

Experimentation with the integrated survey program approach started in three States in 1984 (Arizona, Illinois and Tennessee). The "omnibus" crop and livestock survey approach was replaced with a replicated sample for two reasons. First, livestock studies had shown that bringing in new replicates in each survey period moderated possible survey biases from multiple contacts of the same sample. Also, the omnibus survey approach would require much larger sample sizes than needed for mid-year estimates which would increase cost.

In 1985, the Integrated Survey Program was expanded into 27 major hog, cattle, crop, and grain stocks States. The final expansion phase to an operational program in all States was undertaken in December 1986. Although the area frame had provided national level indications of crop acreage, the integrated survey was the first to provide national estimates of crop production.

Estimates for many commodities were now provided by a limited number of integrated multiple frame surveys. Commodities were grouped for sampling and data collection based on shared reference dates.

Common reference dates for crop and livestock estimates were essential to integrate the surveys. A special task force representing all areas of NASS activity was convened to develop an action plan for consolidating survey reference dates (July 1985). Data users need to have specific points of reference upon which they can base their analyses. In some cases it was necessary to modify the dates for a data series in order to integrate surveys. In other cases it was impossible to integrate because the reference dates could not be changed. A summary of changes made to integrate surveys follows:

1. Quarterly Hogs and Pigs - Reference dates remained March 1, June 1, September 1, and December 1 but data collection was changed from being centered around the first day of the reference month to a two week period following the first of the month. The multiple frame hog and pig survey became integrated with questions on crops and grain storage.

2. Quarterly Grain Stocks - Grain stocks surveys were changed from January 1, April 1, and October 1 to coincide with hog dates in order to integrate these separate surveys. This change was made possible by a change in the marketing year for corn from starting October 1 to starting September 1.

3. June Acreage Survey - This non-probability mail survey for planted acreage was replaced by including crops in the integrated probability survey with a June 1 reference date.

4. June Enumerative Survey - June 1 area frame survey to define the nonoverlap domain for the integrated surveys was continued. In addition, separate area frame estimates continue to be generated for acres planted to individual crops and for June 1 inventories of hogs and cattle. Data collection takes place during the first half of June. Data collection had begun about May 23 for completion by June 10.

5. December Enumerative Survey - Independent area frame survey in December was abolished with the implementation of the integrated quarterly survey.

6. Mid-year and End-of-year Cattle Inventory - Independent multiple frame surveys were replaced with integrated surveys including sheep and goats. Reference dates were kept at January 1 and July 1, based on user preferences.

7. Sheep and Goat Inventories - These surveys were integrated with the January cattle survey.

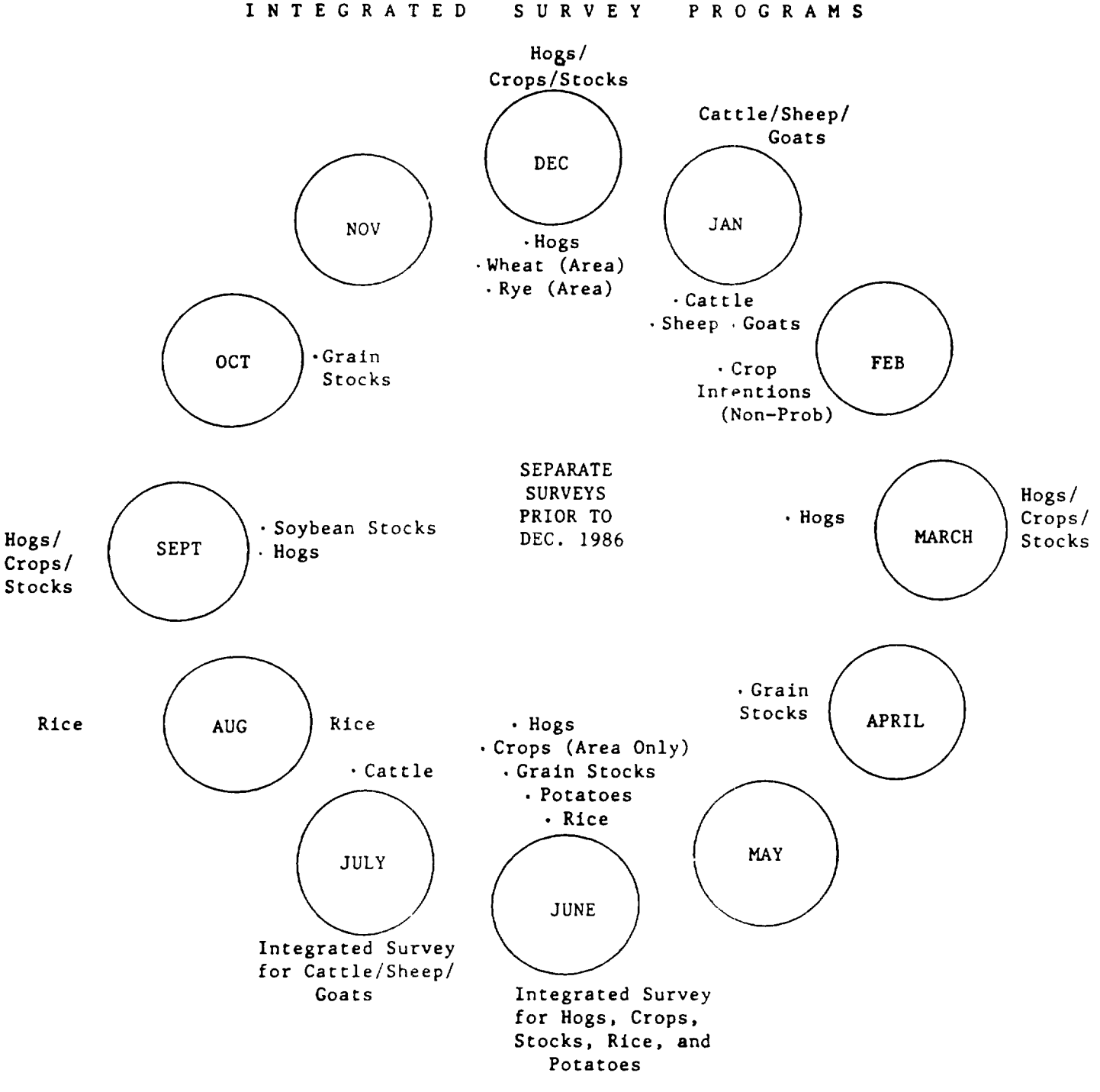
8. Monthly Crop Reports - These previously independent monthly surveys are currently being changed from non-probability panel surveys to subsamples of the quarterly integrated survey.

9. Acreage Intentions - Independent cropping intention survey was integrated into the March 1 quarterly survey.

A schematic representation of the NASS integration of surveys is shown in Illustration 1. This illustration also shows a one-year survey cycle beginning with a new list sample and area frame base survey in June. Survey activities are fully described in a continuously updated "Supervising and Editing Manual" [3]. This instruction manual is utilized by NASS statisticians to promote consistency in survey procedures across 45 State Statistical Offices throughout the United States.

ILLUSTRATION 1

NASS Multiple Frame Surveys Before and After Integration



United States
Department of
Agriculture

National
Agricultural
Statistics
Service

Research and
Application
Division

NASS Research Report
Number SSB-89-05

June 1989

INTEGRATED AGRICULTURAL SURVEYS

R. Bosecker

DISADVANTAGES OF INTEGRATED SURVEY METHODOLOGY: As new and improved techniques and procedural changes were implemented, some trade-offs and sacrifices of old ways were inevitable. Disadvantages were as follows:

- Stratification and sampling for multiple purposes were more complex than for separate surveys for each commodity.
- Questionnaires used for quarterly surveys were more lengthy than previous forms for individual commodities.
- Change in reference dates for grain stocks estimates may have affected historic data bases compiled by data users.
- Later release of hog inventories and pig crop estimates (no more than 10 days) was necessary.

ADVANTAGES OF INTEGRATED SURVEY METHODOLOGY: The Integrated Survey Program offered several advantages which were important improvements to survey methodology. Some of the more important advantages were as follows:

- Smaller total sample size (less cost) was necessary with integrated surveys, while maintaining comparable sampling errors, and extending probability multiple frame estimation to 48 states.
- Fewer survey contacts, i.e., less respondent burden, for many producers without separate surveys by commodity.
- All data were collected on or after the survey reference date, eliminating the need for respondents to forecast events before they to occurred (e.g., hog farrowings, calves born, deaths, etc.).
- Nearly every operation sampled has positive data to report for some item on the integrated questionnaire.
- Data on farm and ranch characteristics can be obtained to provide additional estimates useful to the agricultural data user, e.g., estimates by size of operation for major States.
- Estimates for crop production and grain stocks were provided by the same respondents rather than from independent surveys as before integration.

INTEGRATED AREA SAMPLING: National area frame estimates, with extreme operator list samples, were continued in June and December

every year since the 1960's. Area probability indications were used in conjunction with multiple frame estimates. The December area frame survey was then discontinued as an independent survey after December 1986. The contribution from this survey was judged inconsistent with the cost. The June area frame survey continued to provide area frame mid-year estimates to complement the multiple frame indication. It supports a full one-year cycle of multiple frame surveys by providing a sample of those operators not represented by the list frame.

The area frame sample remains the foundation of multiple frame probability surveys whether or not they are integrated. Some survey efficiencies were achieved for area nonoverlap sampling under the integrated survey procedure. Independent multiple frame surveys necessitated separate nonoverlap domains for each commodity. Currently, one common subset of area frame operators comprise the nonoverlap domain for all crop, hog, and grain stock multiple frame estimates. A different nonoverlap domain is defined for the cattle, sheep, and goats integrated survey.

In addition, the NASS agricultural economic surveys that encompass an annual Farm Cost and Returns and Cost of Production Survey as well as the quarterly Agricultural Labor Surveys are supported by one common nonoverlap domain.

Prior to 1985, independent list samples were selected for each commodity included in several Cost of Production surveys and separately for a Farm Production Expenditure Survey. This was changed so one list population of farmers and ranchers provided for the combined needs of these annual surveys through one common sample. These were the first NASS surveys to be integrated. The integrated survey was renamed the Farm Costs and Returns Survey (FCRS). Quarterly Agricultural Labor Surveys share the same list population, hence the same area nonoverlap domain, as the FCRS but stratification and sample selection are independent of the annual survey.

Some independent multiple frame surveys that have not been integrated remain. Examples of these include surveys for aquaculture and chickens. Nonoverlap domains continue to be individually defined for these surveys. Each tract operator in the June area frame base survey is evaluated against the list populations for the above five surveys (3 integrated and 2 independent) and coded as overlap or nonoverlap for each survey program. This is far fewer nonoverlap domains to manage than would be necessary had NASS continued to conduct independent surveys.

Nonoverlap domain estimates for all multiple frame surveys following the June survey are based on a "frozen domain" concept (Bosecker, 1984). The probabilities of selection for each farm and ranch operator appearing in the June area frame base survey are known. These established probabilities are frozen for nonoverlap

operators throughout the following year and combined with the reciprocals of any subsampling fractions that are applicable. The nonoverlap status of an operator cannot change during the survey year since the list frame, as it was sampled, is held constant. However, the occurrence of new farm operators after the base survey is possible so substitution rules have been incorporated to permit these operators to also have a chance of selection.

Sampling from the nonoverlap domains as defined in the base survey is done after restratifying the agricultural tract operators according to the reported items of interest. (Procedures and variance calculations are described by Kott, 1988.) Separate nonoverlap samples are selected for each of the integrated and independent probability surveys. Many of the same operators are nonoverlap for all the surveys because they do not appear anywhere on the list. To alleviate respondent burden on nonoverlap operators, the area frame replicated sample has been partitioned so that different replications are serving the crop and livestock surveys versus the agricultural economic surveys. A representation of the allocation of area frame replications to surveys is shown in Illustration 2.

ILLUSTRATION 2

Allocation of Area Frame Replications to Surveys

-- SURVEYS --	REPLICATION YEARS IN SAMPLE				
	FIRST	SECOND	THIRD	FOURTH	LAST
June Area	X	X	X	X	X
July C/S/G[1/]	X	X	X		
July Labor				X	X
Sept. H/C/S[2/]	X	X	X		
Oct. Labor				X	X
Dec. H/C/S	X	X	X		
Jan. C/S/G	X	X	X		
Jan. Labor				X	X
FCRS[3/]				X	X
March H/C/S	X	X	X		
April Labor				X	X

1/ C/S/G is Cattle/Sheep/Goats Integrated Survey
2/ H/C/S is Hogs/Crops/Stocks Integrated Survey
3/ FCRS is Farm Costs and Returns Integrated Survey

INTEGRATED LIST SAMPLING: The existence of a national list sampling frame system, providing simultaneous access to many classification data items associated with each record, made integrated sampling possible for NASS.

Each name on the list frame is a sampling unit. The NASS list frame is currently composed of 2.5 million names. Of these , about 1.8 million are eligible for list sampling nationwide. Each name has associated address information and most records contain telephone number and Social Security Number identifiers. Altogether there are more than 40 fields on the record devoted to information that identifies the sampling unit as an individual entity. In addition, the latest available inventory data associated with the farm or ranch is maintained. This "control" data includes items like acres in the farm, cropland acres, acreage for individual crops, inventory numbers for various livestock species, grain storage capacity, number of hired workers, etc. In 1989, an <u>average</u> of nine control data fields was associated with each of the 2.5 million names on the list frame, active and inactive. The longest individual record for control data had a total of 63 separate data fields.

In addition to identifier information and control data, fields are maintained on the record to designate the strata in each survey for which the record is classified, the survey samples for which the record was selected, the replication number assigned for each sample, any applicable special handling needs, a survey performance record (times contacted versus times responded), and the NASS releases desired by the individual. An abbreviated overview of how the list frame features are used in a one year integrated survey cycle is shown in Illustration 3.

Integration necessitated sampling in a multivariate environment in place of independent univariate samples. Stratification by size alone for each commodity was replaced by stratification based on both size and type of operation. This procedure ensures adequate representation in the sample for multiple commodities of interest.

A stratification priority system was devised to group agricultural operations by size and type according to frequency of occurrence in the population. Large operations for a variable of interest, especially relatively rare commodities, are fewer in number and therefore are given higher priority in the stratification process. Small farms with commonly produced commodities, fall through to the lowest priority stratum. An example of list stratification, demonstrating lowest to highest priority, is shown for the hogs/crops/stocks integrated survey in Illustration 4. A similar example is shown for an integrated cattle/sheep/goats sample in Illustration 5.

ILLUSTRATION 3

Yearly List Sampling Frame Cycle for Integrated Surveys

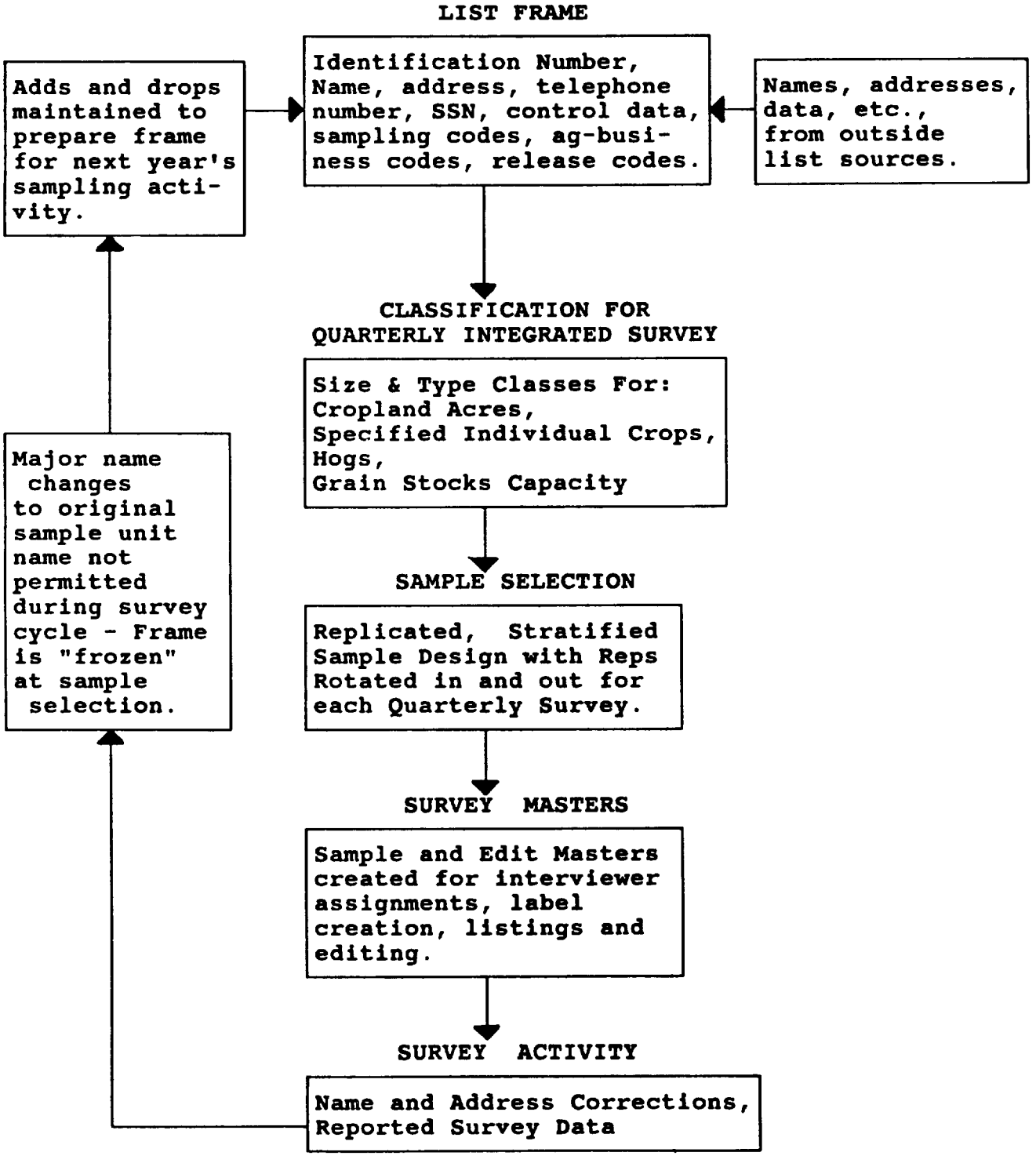


ILLUSTRATION 4:

Example Stratification and Sample for Integrated Hogs/Crops/Stocks Survey In One State

STRATA	DESCRIPTION	POPULATION	SAMPLE
61	Cropland 1-299	13147	240
63	Capacity 1-14,999	11855	250
64	Cropland 300-999	12061	530
66	Capacity 15K-49,999	3102	200
69	Hogs 1-99	2655	160
70	Cropland 1000-5999	3893	390
72	Capacity 50K-399,999	1429	310
75	Hogs 100-199	789	100
76	Hogs 200-499	835	220
78	Dry Beans 50+	127	55
80	Hogs 600-1499	310	125
82	Hogs 1500-3999	84	35
Total		**50287**	**2615**

ILLUSTRATION 5:

Example Stratification and Sample for Integrated Cattle/Sheep/Goats Survey In One State

STRATA	DESCRIPTION	POPULATION	SAMPLE
2	Cattle 1-99	48078	650
4	Cattle 100-199	13710	500
5	Goats 1-499	2314	120
6	Sheep 1-599	5151	370
10	Dairy 50-199	1866	270
16	Cattle 500-2999	2659	450
18	Goats 500-2499	221	65
19	Sheep 600-2499	508	155
22	Dairy 200-499	417	208
30	Dairy 500-999	82	82
31	Cattle 3000-9999	152	152
33	Goats 2500+	147	147
34	Sheep 2500-4999	134	134
Total		**75439**	**3303**

Sample allocation among strata is based on an algorithm described in Bethel (1986), "An Algorithm for Multivariate Surveys". Target precision levels are established by commodity, then samples are allocated to meet or surpass these specified levels.

List samples are selected in replicates to facilitate rotation among survey periods. Both level and change indications are desired so combinations of new and old replicates are used to assure comparable quarter to quarter reports for a portion of the sample as well as infusing new reporters. Rotation eases respondent burden over the cycle of the surveys and provides a different perspective on the appropriate inventory levels by increasing the total number of reports during the one year period.

For example, 11 replicates are selected by stratum for the four quarterly hogs/crops/stocks integrated surveys. Five replicates provide the desired sample size. Replicates 1-5 are used in the June Survey, replicates 3-7 in the September Survey, replicates 4-6, 8 and 9 in December and 6-8, 10 and 11 in March. Therefore, except for the initial survey, 40 percent of the sample is new each quarter and the remainder maintains comparability with one or more previous quarters.

INTEGRATED MULTIPLE FRAME RESULTS: After integration, 48 states were included in the national multiple frame survey program for major crop and livestock estimates. Previously, independent multiple frame surveys could only be afforded in major producing states (e.g., 10 quarterly states for hogs, 18 states semi-annually for cattle, 23 states quarterly for grain stocks, etc.).

Probability surveys for all states utilizing independent samples for hogs, cattle, sheep, and grain stocks, and rice and potatoes in applicable states, would have required a list sample size approximately 10 percent greater than was necessary after integration. Area Frame nonoverlap subsamples were 17 percent smaller after integration of surveys. Sample size comparisons between alternative survey programs are shown in Illustration 6. At current survey costs per interview, the reduced list and area samples result in a savings in excess of $250,000.

Direct comparisons in six mid-western states that previously had separate multiple frame surveys for hogs, cattle, sheep, and grain stocks showed these states experienced a 15 percent decrease in sample sizes between 1985 (year before integration began) and 1988 (integrated survey program after developmental period).

ILLUSTRATION 6:

Sample Sizes for Independent Multiple Frame Surveys Versus Integrated Surveys for 48 States

Survey Period	Independent Surveys 1/		H/C/S Integrated Surveys 2/		C/S/G Integrated Surveys 3/	
	LIST	NOL	LIST	NOL	LIST	NOL
June-July	88,200	4/	70,600	4/	15,000	7,300
September	72,400	11,400	64,200	11,000	---	---
Dec.-Jan.	146,300	32,600	69,400	10,500	57,900	9,200
March	72,400	9,000	61,700	6,000	---	---
Total	379,300	53,000	265,900	27,500	72,900	16,500
Total List	379,300		---- 338,800 ----			
Total Area		53,000	---- 44,000 ----			

1/ Separate surveys for hogs, grain stocks, cattle, sheep, rice and potatoes. Estimated sample sizes for 48 state program.

2/ Integrated surveys for hogs, crop acreage and production (including rice and potatoes) and grain stocks.

3/ Integrated surveys for cattle, sheep and goats (July survey specifically for cattle at National level while January survey measures all three species at State level).

4/ Nonoverlap (NOL) in June is a component of the base area frame survey rather than a separate subsample attached to the integrated surveys as in other months.

Coefficients of variation (C.V.'s) for four commodities are presented in Illustration 7 for survey years before and after integration. Cattle, hogs and rice each had independent multiple frame surveys prior to integration in December 1986. Winter wheat (seeded acres) is included because the integrated survey estimate replaced an area frame probability estimate. Precision levels have been maintained or improved for hogs, rice and winter wheat after integration and with reduced sample sizes. Cattle C.V.'s have gone up in recent years due to increased variance contribution from the area frame nonoverlap domain. List frame C.V.'s for cattle have remained consistent with pre-integrated survey levels.

Illustrations 6 and 7 demonstrate the progress toward the goals stated in the introduction to this paper. Uniform probability survey procedures have been instituted throughout the contiguous United States for all major crop and livestock estimates rather than selected commodities in selected states. Statistical reliability has been preserved at pre-integration levels while the number of producers sampled has decreased.

ILLUSTRATION 7:
Coefficients of Variation for Selected Commodities from Probability Surveys 1981 to 1989

YEAR	Hogs1/ C.V.%	Cattle2/ C.V.%	Rice3/ C.V.%	Winter wheat Seeding4/ C.V.%
1981	2.0	1.8	2.3	2.3
1982	2.0	1.0	2.2	2.6
1983	3.2	1.0	3.1	2.3
1984	1.9	0.9	3.0	2.4
1985	2.1	0.8	3.5	2.4
1986*	**1.8**	**0.8**	**3.0**	**2.2**
1987	**2.1**	**0.9**	**3.1**	**1.3**
1988	**1.9**	**1.3**	**3.0**	**1.5**
1989		**1.6**	**2.0**	

* Integrated survey coefficients are presented in bold type.

1/ 10 State multiple frame coefficients for December surveys.

2/ 28 State multiple frame coefficients for January Surveys.

3/ 6 State multiple frame coefficients for June survey.
Note: 1981 and 1982 coefficients were adjusted downward for outliers.

4/ 48 State probability survey coefficients for December. Prior to integrated survey in 1986, coefficients are based on subsample of area frame tracts.

RESEARCH AND ANALYSIS: Several research and analytical projects were undertaken to evaluate the impact that integration of surveys and a change in data collection periods had on survey indications. Not all of the studies were published in formal NASS Staff Reports. The issues studied (and associated documentation reference) related to the integration process included the following:

1. **data collection costs** for 1984 integrated surveys (Kleweno, January 1986);

2. **post-stratification of hog data** from integrated sample design to pre-integration hog strata (Geuder, memo to NASS Livestock Branch, March 1985);

3. **sample design** for 1985 integrated surveys (Bethel, May 1985);

4. **integrated questionnaire design** (questionnaire pretest report, February 1986);

5. **comparison of hog estimates**, integrated vs. single purpose questionnaires, 9-state parallel surveys (Geuder, unpublished analysis, June and September 1985);

6. **comparison of rice estimates**, integrated vs. single purpose questionnaires (Brown, April 1986);

7. **effects of a change in data collection reference dates**, June 1986,9-state parallel surveys (Pafford, April 1987);

8. **rotation group comparison** of estimates from new and previously contacted respondents, 1984 Illinois integrated survey data (Mergerson, September 1986);

9. **data imputation** (Atkinson, February 1988);

10. **sample design evaluation** for December 1986 and January 1987 integrated surveys (Battaglia, June 1988).

No statistically significant differences were found in the studies comparing data from integrated questionnaires with data from commodity specific questionnaires. Likewise, the study concerned with the change in data collection dates showed no statistical significance in differences between survey indications. Before surveys were combined, data collection had straddled the first of the month reference date and questions were asked as of the time of the interview. Upon integration of surveys, data collection was conducted in a two week period beginning on the reference date and all questions referred to inventory as of the reference date.

Statistical significance was reported by Mergerson (1986) for survey indications in Illinois between first time respondents

(rotated in replications) and previously contacted respondents (retained replications). New replications produced higher estimates than did those retained from a previous survey. Replicated sampling with rotation between quarters, to reduce respondent burden, had been a long standing feature of the independent multiple frame surveys. The integrated program continued sample rotation. Since estimates of change between survey periods are produced from replicates retained in the sample, further study is needed on the potential impact of the observed rotation group effect.

Prior to integration the NASS procedure for handling missing data was a reweighting process. Original selection probabilities were adjusted by stratum to compensate for missing reports. No record by record imputation program existed. When a single commodity had been stratified by size and the entire questionnaire related to that commodity, the implicit assumption that nonrespondent data were similar to respondent data by stratum was acceptable. Also, there was very little item imputation needed because the entire questionnaire was either complete or missing.

With the advent of integration, strata based on a priority arrangement among several variables no longer provided the same homogeneity by size for individual commodities. Differences for a given commodity within a stratum could be greater between respondents and nonrespondents than before. Moreover, some sections of the integrated questionnaire might be complete while others were incomplete. The only other experience for NASS in dealing with questionnaires containing different sections for several commodities was in the area frame. Manual imputation was required for missing data in those questionnaires.

A new automated system was developed to handle missing data for integrated questionnaires (Atkinson, 1988). Partially completed questionnaires were likely to occur more often given an integrated questionnaire consisting of separate sections devoted to various commodities. Computer imputation provided a practical and consistent solution relative to the following alternatives: (a) loss of valid data in partially completed questionnaires, (b) separate summaries by commodity section with differing expansion factors, or (c) manual imputation. The existence of the imputation program also permitted an evaluation of the impact of missing reports on survey indications without assuming nonrespondent means were the same as respondent means for each stratum. Imputation could be based on supplementary information for each commodity. This became more important after integration because strata were based on individual, dominating variables that may or may not be well correlated with all other variables in the questionnaire.

Sample design analysis is a continuous process. Early work by Bethel (May 1985 Staff Report) on the integrated survey design evaluated alternative stratification based on simultaneous use of

two or three variables to define some of the strata. The goal was to reduce the number of strata originally used in the three test states. Reduced numbers of strata have since been implemented but without multiple variables in any given stratum. The most comprehensive analysis to date is included in the report by Battaglia (1988). Target precision levels were identified by state and evaluations of both the list and area frame components were presented to achieve the desired multiple frame precision. An update to this report based on data for more years under an integrated sample design is being prepared.

Further research is planned for the integrated survey program in the areas of alternative estimators and quality assurance. Reinterview studies have been conducted to identify sources of nonsampling error. Differences in estimates for crop acreage between the area frame survey and the integrated multiple frame survey in June are being investigated. Incorporation of integrated survey data into the NASS county estimates program is also of prime interest. Much work remains to be done to get maximum benefit from the integration of agricultural surveys.

SUMMARY: Integration of crop and livestock surveys in the National Agricultural Statistics Service (NASS) was initiated in 1984 beginning with three states (Arizona, Illinois and Tennessee). All 48 conterminous states had been incorporated into the present sample design and survey reference dates beginning with the June 1988 survey. The intervening developmental years were necessary for evaluating the impact of integration and changes in reference dates on survey estimates. A change in the reference date for mid-year cattle estimates was cancelled after a brief trial period.

Major benefits derived from integration include cost efficiencies, respondent burden reduction, and improved analytical capabilities. Forty-eight states were incorporated into multiple frame probability sampling methodology through the integrated program. Individual surveys by commodity would not have been cost effective and individual producers may have been contacted repeatedly for a succession of commodity specific surveys. Questionnaires that are more comprehensive concerning the characteristics of a farm operation provide opportunity to increase knowledge about the structure of agriculture.

On the other hand, integration has increased the complexity of sampling and lengthened questionnaires. Stratification using several variables to serve multiple survey requirements is challenging but has been proven successful. Longer questionnaires are judged to be less burdensome on respondents than several independent contacts. In addition, respondents are more likely to feel they have made a contribution by supplying positive data for some items on the questionnaire rather than reporting no data for a questionnaire devoted to a specific commodity.

The integrated survey program is now a integral part of NASS methodology. Those data needs that fit the established reference dates and can be met by incorporating the target population into the integrated sampling scheme will likely become additions to the existing integrated survey vehicle. Limitations to further integration will be avoidance of overlong questionnaires and the prevention of sample degradation for existing program needs due to incompatible special requirements. Analysis continues in order to achieve maximum effectiveness through integration.

REFERENCES:

1. Agricultural Surveys Supervising and Editing Manual.

2. Atkinson, A. D., "The Scope and Effect of Imputation in Quarterly Agricultural Surveys," Staff Report, February 1988.

3. Battaglia, R. J., "Methodology to Evaluate Sample Designs of the Quarterly Agricultural Survey," Staff Report, June 1988.

4. Bethel, J., "An Algorithm for Multivariate Surveys," Staff Report, January 1986.

5. _____, "Sample Design for the 1985 ISP/JES," Staff Report, May 1985.

6. Bosecker, R. R., "List Dominant and Frozen Domain Procedures," Staff Report, July 1984.

7. Brown, L. E., "Comparison of Rice Survey Indications Between Rice Multiple Frame and ISP Surveys," Staff Report, April 1986.

8. Cochran, R. S., "Multiple Frame Sample Surveys," Proceedings of the Social Science Section of the American Statistical Association Meeting, Chicago, Illinois, 1964.

9. Data Collection Branch, "Summary Report: Bridge Pretest Survey - December 1, 1985," Staff Report, February 1986.

10. "Framework for the Future," Statistical Reporting Service, Special Task Force Report, March 1983.

11. Hartley, H. O., "Multiple Fame Surveys," Proceedings of the Social Science Section of the American Statistical Association Meeting Minneapolis, Minnesota, 1962.

12. King, A. J. and R. J. Jensen, "The Master Sample Agriculture," Reprint in Journal of the American Statistical Association, Vol. 40: pp. 38-56, 1945.

13. Kleweno, D. G., "Data Collection Costs for the 1984 Integrated Survey Program," Staff Report, January 1986.

14. Kott, P. S., "Estimating the Nonoverlap Variance Component for Multiple Frame Agricultural Surveys," NASS Staff Report, May 1988.

15. Mergerson, J. W., "Integrated Survey Program: Rotation Effects Analysis Summaries, Illinois 1984-1985," Staff Report, September 1986.

16. Nealon, J. P., "Review of the Multiple and Area Frame Estimators," Staff Report, March 1984.

17. Pafford, B., "Changing Survey Reference Dates: An Analysis of the June 1986 Bridge Survey," Staff Report, April 1987.

18. "Survey Reference Dates Report," Statistical Reporting Service, Special Task Force Report, July 1985.

☆ U.S. Government Printing Office 1990 - 261 161 20090