THE STATISTICAL REPORTING SERVICE SYSTEM

By William E. Kibler [1/]

Basic Sampling Methodologies

During the past two decades the methodology for developing current agricultural statistics on crops, livestock and prices has shifted almost entirely from non-probability mail surveys to probability surveys using area and list sampling frames. With probability surveys, estimates for characteristics of interest can be generated without dependence on prior survey relationships or benchmark data such as the 5-year Census of Agriculture. Such surveys also provide the data necessary to derive sampling errors for evaluating the reliability of estimates generated and for optimizing sample designs and allocation of sampling units.

A basic requirement for any probability survey is a complete sampling frame which is an aggregation of the elements from which a sample can be selected. An area frame is the principal frame used for estimating major crop acreages, yields and production. This frame is made up of small geographic units of land called "segments" which may be sampled. It is constructed using the most current aerial photography available to classify (stratify) all land according to its current use.[1/] The stratification is based on extent and type of farming and can be described in four broad categories: (1) intensively cultivated areas where a significant portion of the land is under cultivation, (2) extensive agricultural areas used primarily for grazing and producing livestock, (3) highly developed land found in city residential, shipping and industrial areas, and (4) non-agricultural land such as parks, military reservations and other recreational areas.[2/]

As frames for individual States are periodically updated, by using additional materials such as satellite imagery, more sophisticated stratification procedures have been used. Examples include the addition of an agri-urban stratum which is used as a transition zone between the city and agricultural strata. Within the intensive agricultural stratum refinements have been made by including additional information such as soil type and topography to develop crop-specific strata. For example, a fruit/vegetable stratum in California, a dry land wheat stratum in Oregon and Washington, and rice, peanuts, wheat/sorghum, and cotton strata in Texas. Geographic stratification is sometimes used, in addition to the land use stratification, to separate differing agricultural areas. This is accomplished by grouping counties into type-of-farming districts.

About two decades ago research showed that an optimum size segment should include about two farms and be about one square mile in intensively cultivated areas, several square miles in extensively farmed areas and one-tenth square mile in industrial or urban areas. As additional refinements have been made in both the area sampling frame and sampling methodology over the past decade, segment size has generally been reduced to an average of about .7 square mile for the intensively cultivated strata. In many states .5 square mile segments are used. This, combined with increasingly sophisticated sample designs, has permitted

---

1/ Acting Administrator and Chairman of the Crop Reporting Board, Statistical Reporting Service, USDA, Washington, D.C., 1982.

significant reductions in data collection costs (up to 30 percent in some States) without adverse impact on the error level of the estimates from the June Enumerative Survey. This has been of primary benefit in helping to cover inflation costs in the absence of increased appropriations.

Two sampling methods are followed in selecting sample segments. A systematic-sample approach is used in some States with the frame units arrayed geographically to ensure proper dispersion over the area of interest. Since 1974 all new samples have used interpenetrating designs to provide flexibility in computing sampling variation and segment rotation.

Crop Sampling Methodology

A sample of 15,700 segments (about 350 per State) representing about 0.4 percent of the universe is selected and enumerated about June 1 to collect data on crops planted and livestock numbers as well as characteristics of farms. Both direct expansion and ratio estimators are used for aggregating sample data to State, regional and national totals. Survey data from each segment are expanded by the reciprocal of the probability of selection (typically a factor of about 250) to obtain the direct expansion estimate. A ratio estimate is also computed using current and previous years' data since about 80 percent of the segments are enumerated in successive years. This estimate is particularly useful in evaluating changes from year to year for identical segments. Sampling errors for acreage planted to major crops are about 2 percent at the national level, 3 to 4 percent at the regional level and 4 to 6 percent at the State level.

Sampling Errors from the 1981 June Enumerative Survey
for Planted Acreages at State, Regional and National Levels

Commodity	National	East North Central Region	Illinois	Mississippi
	- - - P e r c e n t - - -			
Corn	1.2	1.2	2.5	18.7
Soybeans	1.3	1.6	3.0	6.0
W. Wheat	1.6	3.2	6.6	15.6
Cotton	3.3	-	-	10.8
Sorghum	4.3	-	32.9	24.4

Followup objective yield surveys are made for wheat, corn, soybeans and cotton to update planted acres for fields actually planted after June 1, to collect information for forecasting yields during the growing season, and to estimate actual yields at harvest. These surveys provide information based directly on counts, measurements and weights of the crop made from small randomly selected plots in sample fields. Samples are designed to produce estimates of at harvest yield with sampling errors of 1 to 2 percent. Large nonprobability mail surveys are conducted to gather data for strengthening State and sub-State estimates for crops important to the State's agricultural sector and to support cooperative State-Federal programs. Samples for such surveys vary in size from 150,000 to 200,000 and operate

fairly effectively for disaggregating accurate annual benchmarks based on probability surveys.

Livestock Multiple Frame Sampling

Multiple frame sampling utilized more than one sampling frame to cover the universe of interest. The theory for multiple frame sampling was first developed in the 1960's. /3/ Its use has rapidly grown because of its distinct advantages in efficiency of costs in data collection and its ease of adaptation for specialized characteristics associated with a small portion of firms in a universe. The theoretical concepts of multiple frame sampling are basically the same as those for probability sampling concerning known probabilities and randomness of selection. In addition, two other characteristics must hold: (1) every element of the population must belong to at least one of the frames, and (2) it must be possible to specifically identify the frame(s) to which, if any, each selected sample unit belongs other than the one from which it was drawn. The use of the area frame as described earlier satisfies the first characteristic. The second characteristic requires the proper classification of each farm operator as to whether his name is included on the list(s) frame(s). Multiple frame sampling technology is used for rice, potatoes, quarterly hog surveys in 14 States, and semi-annual cattle surveys in 28 States.

With multiple frame sampling, data can be collected more efficiently by mail or telephone and more efficient sampling can be accomplished by stratification of the list by size of operation. A variety of list sources such as ASCS, State Farm Censuses, brand inspections, etc., is used in assembling list frames. However, due to rapid organizational and operational changes that occur, lists must be updated periodically to retain their advantages in sampling and cost efficiencies. There are also some complex operating problems associated with identifying and measuring overlap between the two frames (area and list) that increase nonsampling errors. Typical sampling errors for these multiple frame surveys for cattle and hogs are shown in the following table:

Sampling Errors for 1980 and 1981 Based on
Multiple Frame Surveys for Hogs and Cattle
at Various Geographic Levels

Survey	23 State Level	14 State Level	28 State Level	Iowa	Georgia
	- - - P e r c e n t - - -				
December 1, 1980 Hogs	2.1	2.3	--	3.5	11.9
June 1, 1981 Hogs	1.8	2.0	--	4.1	9.2
January 1, 1981 Cattle	--	--	1.3	3.6	4.9
July 1, 1980 Cattle	--	--	.8	3.5	6.5

New Probability Surveys for Prices

The area and list sampling frames described earlier are not suited for collecting current information on prices farmers receive for commodities they sell or prices

paid for inputs used in production. An indirect method is used to establish these frames and select appropriate samples. For obtaining prices received for grains, a list of all grain and oilseed elevators (about 14,000) is maintained from administrative records, available as a byproduct of licensing requirements. These are stratified by storage capacity and a probability sample of about 1 in 6 selected for surveying. Similar lists for cotton, peanut, and rice buyers serve as frames for these crops.

For commodities such as cattle, hogs, vegetables and fruits, a periodic point of sale survey is conducted to determine what portion of the total production is sold through each marketing channel such as auctions, dealers, commission firms, processors and packers. The universe list of these firms is then stratified by type of marketing channel and sampled using probabilities proportional to the channel's importance in the marketing of the commodity.

Firms sampled are surveyed monthly on about the 15th to collect actual quantities purchased and dollars paid farmers for each commodity during the previous month. These data are used to derive a self-weighted average published as the revised price received for the entire month. In addition, the exact price being paid to farmers about mid-month is obtained and published as the preliminary price as of the 15th of each month. The data on quantities purchased are used by the Economic Research Service (ERS) in estimating current cash receipts for aggregation and calculation of farm income. The typical entire monthly price received for corn has a sampling error of about 3 cents while the error for the aggregated 5-month average used for determining the level of deficiency payments is less than 1 cent.

In collecting data on prices paid for inputs used in commodity production, a periodic point of purchase survey is conducted to ascertain the portion of the various inputs that are bought through cooperatives, brokers or wholesalers, dealers or manufacturers. Lists of firms are assembled from phone directories, licensing bureaus, and the American Business Lists Inc., and classified by specific inputs sold. The listed firms are geographically grouped by counties to form a frame of primary sampling units. For primary sampling units selected to be surveyed, a second stage of sampling is performed to identify the individual firms to be included in the sample. The clustering by counties makes data collection more efficient by reducing travel. Much of the work requires personal interviews for establishing accurate specifications on inputs priced.

Reliability and Completeness of Principal Statistical Series

Many data users have requested that the Crop Reporting Board provide additional information on the sources of data used in establishing official estimates and measures of their reliability since social or economic costs of errors in forecasts can be significant./4/ Beginning in 1977, most major reports have included a general summary of survey procedures, comments about errors from sampling and non-sampling sources and typical sampling errors for surveys or Root Mean Square Errors for forecasts. The following is typical of the summaries provided for livestock reports:

> RELIABILITY AND ESTIMATING PROCEDURES: Primary data used in setting these hog estimates were obtained from a sample of farmers across the U.S. using probability surveys. Information was collected by mail,

telephone and personal interviews. Since all operations raising hogs were not included in the sample, survey estimates are subject to sampling variability. This variability, as measured by the relative standard error, is about two percent at the U.S. level for hog inventory. This means that chances are approximately 95 out of 100 that survey estimates will be within four percent of the complete coverage value if the same procedures were used to survey all producers. Survey estimates are also subject to non-sampling errors such as omissions, duplications, and mistakes in reporting, recording and processing the data. These errors cannot be measured directly, but they are minimized through rigid quality controls in the data collection process and a careful review of all reported data for consistency and reasonableness.

The sampling variability of survey estimates on intended farrowings is slightly larger than that for inventories. More important, actual farrowings may differ significantly from reported intentions due to unexpected economic and environmental conditions. These differences have exceeded four percent for about one-third of the quarterly pig crops during the last seven years.

In setting the inventory estimates, the Crop Reporting Board constructed a U.S. balance sheet using estimates on births, deaths and check data on slaughter, imports and exports. This balance sheet provided an additional check on survey inventory estimates./5/

Some users have commented that these have been useful in analyzing data but the numerical sampling or forecast errors have generally not been used extensively in modeling. In fact, some data users have completely ignored the cautions about intended farrowings and assumed that they will always represent what will occur during the next 6-month period. The table on page 7 illustrates the preliminary estimates for the inventory of all hogs and pigs based on sample survey data and the final estimate that was established after reevaluating all data when slaughter records became available six months later.

If we obtained perfect data collection the sampling errors would indicate that about 2 out of 3 of the estimates would require revisions of less than 2 percent and 19 out of 20 would require revisions of less than 4 percent. For the 20 estimates during this period, 18 required revisions of less than 2 percent and 19 required revisions of less than 4 percent. Hence, the sampling errors are reliable measures of the accuracy of the estimates. The same sample of producers is used to obtain data on farrowing intentions. Hence, the same statistical analysis can be applied to farrowing intentions using the table on page 8 .

These intentions forecasts also have a sampling error of about 2 percent. Note the very large deviations for some 6-month periods cannot be explained by statistical measures such as sampling errors. Thus, these deviations must be associated with either problems in acquiring accurate data on intentions from producers or changes in plans made by producers due to such things as weather, feed cost, market prices, or as a direct result of the published intentions report. Hence, analysts should use these data with much more caution than the inventory data.

Comparison of Preliminary and Final Estimates of U.S. Inventory of all Hogs and Pigs, 1971-1980

Year and Survey	All Hogs and Pigs Inventory		
	Preliminary Estimate Based on Survey	Final Estimate Based on Slaughter	Change
	1,000 Head		Percent
1971:			
June 1	66,070	65,718	-0.5
December 1	62,972	62,412	-0.9
1972:			
June 1	61,556	60,626	-1.5
December 1	61,502	59,017	-4.0
1973:			
June 1	60,271	59,571	-1.2
December 1	61,022	60,614	-0.7
1974:			
June 1	59,437	58,878	-0.9
December 1	55,062	54,693	-0.7
1975:			
June 1	48,165	47,860	-0.6
December 1	49,602	49,267	-0.7
1976:			
June 1	52,643	53,930	+2.4
December 1	55,085	54,934	-0.3
1977:			
June 1	54,100	54,460	+0.7
December 1	57,587	56,539	-1.8
1978:			
June 1	54,930	55,240	+0.6
December 1	59,860	60,356	+0.8
1979:			
June 1	64,890	65,020	+0.2
December 1	66,950	67,353	+0.6
1980:			
June 1	65,930	65,255	-1.0
December 1	64,520	1/	

1/ Subject to future minor revisions.

Comparison of Farrowing Intentions Forecast and Actual Farrowings That Occurred, U. S. Total, 1971-1980

Year and Period	Intentions Forecast	Actual Estimate	Change
	- - - - 1,000 Head - - - -		--Percent--
1971:			
December-May 1/	7,222	7,237	+0.21
June-November	6,265	6,339	+1.18
1972:			
December-May 1/	6,544	6,498	-0.70
June-November	6,005	5,973	-0.53
1973:			
December-May 1/	6,980	6,438	-7.77
June-November	5,979	5,869	-1.84
1974:			
December-May 1/	6,491	6,315	-2.72
June-November	5,760	5,476	-4.94
1975:			
December-May 1/	5,385	4,973	-7.65
June-November	4,730	4,952	+4.02
1976:			
December-May 1/	5,353	5,777	+7.92
June-November	5,811	5,850	+0.67
1977:			
December-May 1/	6,109	6,050	-0.97
June-November	6,144	6,009	-2.20
1978:			
December-May 1/	6,620	6,034	-8.86
June-November	6,247	6,398	+2.42
1979:			
December-May 1/	6,903	7,179	+4.00
June-November	7,419	7,306	-1.53
1980:			
December-May 1/	7,176	7,231	+0.77
June-November	6,716		
1981:			
December-May 1/	6,780 2/		

1/ December previous year.
2/ Latest estimates - subject to future revision.

The SRS has concluded four years' experience in using the Root Mean Square Error (RMSE)/6/ statistic as an indication of the reliability of crop production forecasts made during the growing season. The Root Mean Square Error is calculated on the basis of past forecasting performance. It is derived by averaging the squared deviations between monthly forecasts and the final estimate over a given period. The square root of these averages is the RMSE./7/ The assumptions necessary to make this statistical measure valid are (a) a normally distributed series of forecasts compared to the final estimates, and (b) factors affecting the current year's crop after the forecast date are not greatly different from those influencing crop forecasts during the historic reference period. For crops, 20 years of data are used and a t-value of 1.725 is used to compute the 90 percent interval compared to the normal distribution value of 1.645. Its performance has exceeded expectations as shown in the table on page 10 .

These data show that about 69 percent of the 189 forecasts made during the period have been within the expected 67 percent confidence interval and nearly 98 percent of the forecasts have been within the 90 percent confidence interval. Since the RMSE uses actual performance over a 20-year period to measure reliability, any improvements made in survey systems during recent years should make the measure a bit conservative when evaluating current year forecasts. An attractive concept of the RMSE statistic is that it can be derived well ahead of its expected use. The following table gives the RMSEs that have been or will be used in 1981 crop reports:

Root Mean Square Errors (Percent)
For 1981 Crop Production Forecast

(Derived from 1961-80 Forecasts as published by the Crop Reporting Board)

Crop	May	June	July	August	Sept.	Oct.	Nov.	Dec.
	- - - - - - - - - - - - Percent - - - - - - - - - - - -							
W. Wheat	6.7	5.9	3.4	1.3	--	--	--	8.0
O. Spring Wheat	--	--	10.2	6.7	3.2	3.1	--	--
D. Wheat	--	--	14.5	8.1	5.2	5.3	--	--
All Wheat	--	--	3.4	1.9	1.3	1.3	--	--
Barley	--	--	8.4	4.9	3.5	--	--	--
Oats	--	--	7.6	4.8	4.0	--	--	--
Corn	--	--	9.1	6.8	5.1	3.9	2.6	--
Sorghum	--	--	--	6.4	4.7	4.7	4.4	--
Feed Grains	--	--	--	5.5	4.1	3.1	2.1	--
Soybeans	--	--	--	5.1	4.2	3.3	2.6	--
Rice	--	--	--	4.2	3.2	2.6	2.3	--
Cotton	--	--	--	8.7	7.2	5.4	3.8	1.6

As should be expected, these data show that forecasts improve consistently as the growing season progresses and the RMSEs begin to approach the level of actual observed sampling error for at harvest estimates.

Data on Performance of the Root Mean Square Error Statistic
as an Indication of the Reliability of Crop Production Forecasts (1977-1980)

Forecast Period 1/	Total Number of Forecasts	67% Confidence Interval		90% Confidence Inter	
		Forecasts Within	Forecasts Outside	Forecasts Within	Forecas Outsid
			1977 Crop		
Early Season	23	20	3	23	0
Mid-To-Late Season	22	19	3	22	0
Total	45	39	6	45	0
			1978 Crop		
Early Season	25	16	9	24	1
Mid-To-Late Season	23	14	9	23	0
Total	48	30	18	47	1
			1979 Crop		
Early Season	25	15	10	22	3
Mid-To-Late Season	23	15	8	23	0
Total	48	30	18	45	3
			1980 Crop		
Early Season	25	14	11	25	0
Mid-To-Late Season	23	18	5	23	0
Total	48	32	16	48	0
GRAND TOTAL	189	131	58	185	4

1/ Early Season Forecast Months: December, May and June for Winter Wheat; July August for Durum, Other Spring and All Wheat, Corn, Barley and Oats; August a September for Soybeans, Cotton, Rice, Sorghum Grain and all Feed Grains (197E Mid-to-Late Season Forecast Months: July and August for Winter Wheat; Septen and October for Durum, Other Spring and All Wheat; September for Barley and C September, October and November for Corn; October and November for Soybeans, Sorghum, Feed Grains, Rice and Cotton.

Sampling Errors for 1980 Production Based on
Enumerative and Objective Yield Probability Surveys

Crop	Percent Sampling Error
All Wheat	2.2
Corn	1.6
Soybeans	1.9
Cotton	5.0

For less sophisticated data users the following type of table appears in each monthly crop report during the growing season:

Crop Production Forecasts

Ten-Year (1971-80) Record of Differences Between
First Monthly Forecasts and Final Estimate After Harvest

Crop and Month	Units	Quantity			Number of Years	
		Average	Smallest	Largest	Below Final Estimate	Above Final Estimate
July 1981						
Corn	Million Bu.	510	2	1,276	5	5
Oats	Million Bu.	45	8	92	5	5
Barley	Million Bu.	32	0	71	6	4
All Wheat	Million Bu.	54	2	143	5	5
Durum	Million Bu.	10	4	19	4	6
Other Spring	Million Bu.	31	3	97	6	4
Winter	Million Bu.	29	1	55	6	4
August 1981						
Rice	Million CWT	3	0	7	4	6
Soybeans	Million Bu.	78	1	165	6	4
Cotton	Thousand Bales	796	149	1,690	4	6
Sorghum	Million Bu.	34	2	78	5	5
Feed Grains	Million Tons	9	1	22	7	3

These are actual data that will appear in upcoming 1981 reports and are derived from records kept by the Crop Reporting Board to assist in analyzing its forecast record. For forecasts to maintain credibility they must be both reliable and objective. Producers always assert Crop Reporting Board forecasts are always too high. A simple measure such as the number of years forecasts are below or above the final estimates helps establish whether the Board errs consistently in an optimistic or conservative manner.

One topic that received much discussion in 1980 was differences in numbers of farms published by the Department of Agriculture and the Bureau of the Census. A task force was appointed to review the data and investigate whether (1) the Department survey procedures might be missing operations that qualified as farms, or (2) the Census might be tabulating duplicate records for the same farm or including operations that did not actually qualify as farms. As the result of this study, some modifications were made in survey procedures used by both agencies and many of the major differences observed in initial reviews were eliminated.

The number of farms as published differed about 2 percent for the 48 contiguous States with SRS estimates above Census in 10 States, essentially the same as Census in 12 States and below Census in 26 States./8/ Reasons for these differences include (1) the Census procedures include any farm operated during the year while SRS procedures measure only those operating on June 1; (2) the Census survey procedure is more likely to result in classification of marginal operations as farms; (3) Census coverage of special farms such as mink, nursery, fish operations, is more complete; (4) SRS's area frame screening procedures may not identify all urban farm operators or special farms; and (5) Census has more opportunity for duplicating partnerships, or counting landlords or tenants as operators. For the major north central region, the two series are at almost the same level.

Data Needs Identified for the 1980's

Numerous agricultural data needs have been identified by data users at Department of Agriculture and Bureau of the Census data users workshops, commodity organizations, advisory committees, and national or State policy officials./9/ These include:

(a) Earlier planting, farrowing or calving intents,
(b) Current feeding and feed conversion rates,
(c) Data on marketing trends,
(d) Calving rates for beef and dairy cows,
(e) Monthly sow farrowings,
(f) More detailed data on double cropping,
(g) Expanded sub-State and county data for livestock and poultry,
(h) Forecasts and broader geographic coverage for sunflower production,
(i) Data on quality and amount of forage production,
(j) County estimates for acreage, yield and production for double cropped and other cropping practices by commodity,
(k) More detailed data on farm and migrant labor,
(l) Improved coverage of aquaculture production, and
(m) Improved data on transportation costs.

All of these are very legitimate needs but must be prioritized with all existing series to match resources and needs. In making these decisions, a number of factor are considered. These include quality (reliability), timeliness, frequency, geographic and commodity coverage, data user support and industry acceptance of responsibility for providing the basic data.

Theoretical work in several areas has shown that social benefits accrue rapidly for commodity data until sampling errors reach about 2 percent. It also considers that sampling errors are not the only source of survey errors and that considerable

effort must be exerted to keep the level of nonsampling errors, which are difficult to measure, within the 2 percent bound. However, we must not overlook the fact that this generally becomes a threshold value in survey planning. The value of accuracy for certain types of information, such as prices received by farmers for crops covered by the Food and Agriculture Act of 1977, is more easily quantified. A change of 1 cent per bushel for the aggregate 5-month price period for corn could mean a $50 to $60 million difference in deficiency payments to farmers. Statistical data on such items as production, inventories, marketing, etc., can be used to develop a balance sheet for checking the reliability of data. The Crop Reporting Board develops these balance sheets also for use in generally reviewing statistical survey data. However, it does not discount survey data in establishing production or inventory in order to force the components to balance. To do so would imply an accuracy that doesn't exist. The Board policy is to accept an imbalance or residual of up to one percent which it feels reflects statistical (or non-statistical) errors that are present in each of the components used in the balance sheet.

For surveys during the growing season that relate to a specific date, such as August 1, the release date is set about 10 to 12 days later. Although there are numerous requests for earlier release of such data, its practicality and efficiency from the standpoint of costs are questionable. First, centering data collection on the first day of each month requires that data be collected the last few days of the previous month and the first few days of the new month. This leaves, at most, 7 or 8 working days to summarize, analyze and publish the results. Reviews of month-to-month changes in production forecasts over a period of 5 years show that they average less than 2 percent. This would imply that, on the average, changes that have occurred in the 10-day period between data collection and publication are generally very small. For larger surveys such as the hogs and pigs report and cattle or grain stocks reports, which must wait a few days after the first of each month for firms to close their books, the release date is about 3 weeks after the reference date. These surveys obtain inventories as of a point in time so the data are not subject to changes for the same reasons cited for crop production forecasts.

Forecasts of crop production are done on a monthly basis during the growing season with qualitative assessments on growing conditions provided by the Weekly Weather and Crop Bulletins. The Crop Reporting Board policy requires that survey data be collected from producers for each forecast. Hence, the added cost and the inability to measure changes that might occur for shorter intervals almost precludes the issuance of more frequent reports. Since it takes several days for each forecast to be totally reflected in the market and there is oftentimes 1 to 2 days' inactivity in the market immediately preceding major reports, there is some question whether more frequent reports would be beneficial. The established frequency of other weekly, monthly, quarterly or semi-annual reports is generally associated with the time required to produce the commodity. Much of the pronounced seasonality that once existed for milk, eggs, pigs, etc., has been somewhat eliminated by specialization and significant portions of the month-to-month changes are associated with differences in length of months rather than changes in actual production levels.

During recent years the Crop Reporting Board has followed a policy of providing current estimates of forecasts at the State level for approximately 95 percent of

production for each of the major crops and combining all other States into one aggregate total. A further refinement is to also include any State having at least 1 percent of U.S. total production. The geographic coverage is reviewed every 5 years and changed to reflect geographic shifts in production that occur through time. For the minor producing States, a single production forecast is made and carried forward through the growing season until an annual survey, using both probability and nonprobability surveys, is used to establish annual levels of production after harvest. This approach is cost-effective since it requires significant resources to collect data for commodities that are produced on a relatively small portion of all farms. It also carries some risks. This occurs when conditions or changes for the 5 percent omitted are considerably different from the other 95 percent.

New Federal Crop Insurance Corporation (FCIC) programs and rapid changes in cropping patterns are both increasing the need for county level data. Much of the current county data is financed by cooperative State funding and is based on nonprobability survey techniques. The current reliability of such data, although not scientifically measured, is thought to be 10 to 15 percent for major producing counties.

Cooperation in providing the basic data for the industry being surveyed is of utmost importance in maintaining the quality of data. Nonresponse rates of more than 10 percent are considered critical especially if they are confined to a particular segment of the universe. Special efforts are made to work with industry representatives to improve cooperation when nonresponse rates increase to this critical level.

Improvements Accomplished or Being Implemented

Area frame modifications made over the past decade that follow current land uses and provide for the control of segment size, have reduced sampling errors for major crops by about one-third even after sample sizes have been reduced about 15 percent. New work in this activity will focus on keeping these frames up-to-date in areas like the Mississippi Delta where significant land clearing continues and the Great Plains where rapid expansion of irrigation has introduced significant new cropland. Landsat imagery has been introduced into current area frame update procedures to compensate for the lack of up-to-date aerial photographs. Landsat data and technology are being researched as a method for refining area frame construction activities. /10/ Initial research results indicate that it might be possible to develop crop-specific stratification of intensively cultivated cropland by identifying special crops for areas 8 to 10 miles in size. /11/

Currently, efforts are underway to expand the list sampling frame (used in multiple frame sampling) to include as many farm operators as possible, with supplementary information on the commodities they produce and measures of their size. This frame will be more complete for large and specialized operations. Lists of farm operators from many sources have been assembled for developing as complete a list as possible. Despite extensive efforts to (1) identify duplication, and (2) eliminate names or operations with no agricultural operations, lists still generally contain more total names than there are farms and as many as one-third of those surveyed no longer operate farms. Thus, our experience in list development

is no different from that of the Census. It is very difficult to maintain a list frame that currently includes a large portion of the very small farm operators; however, the list frame with supplementary data is of utmost importance in developing more efficient data collection techniques and in improving the quality of statistical series at sub-State or county levels.

Research continues for making more extensive use of satellite data in improving crop acreage estimates at the sub-State and county levels. /12/ Relative efficiencies in terms of variability were about 2.5 for corn and soybeans compared with regular survey methodology. Relative efficiencies have to reach 4.0 or better to begin achieving the desired results. Although some county estimates had sampling errors as low as 7 to 9 percent, errors for a large number of counties still exceeded 20 percent. Recent research results on new estimators for using these same data show county estimates for major crops with mean square errors ranging from 5 to 18 percent with most being in the 5 to 10 percent range. /13/

Two very important data series that have not been shifted to a probability basis are the "on-farm" grain stocks survey and the fall acreage and production surveys designed to collect acreage utilization and yields for a large number of field crops. These will get careful consideration for future budget submissions.

* * *