United States
Department of
Agriculture

National
Agricultural
Statistics
Service

Research Division

SRB Research Report
Number SRB-93-11

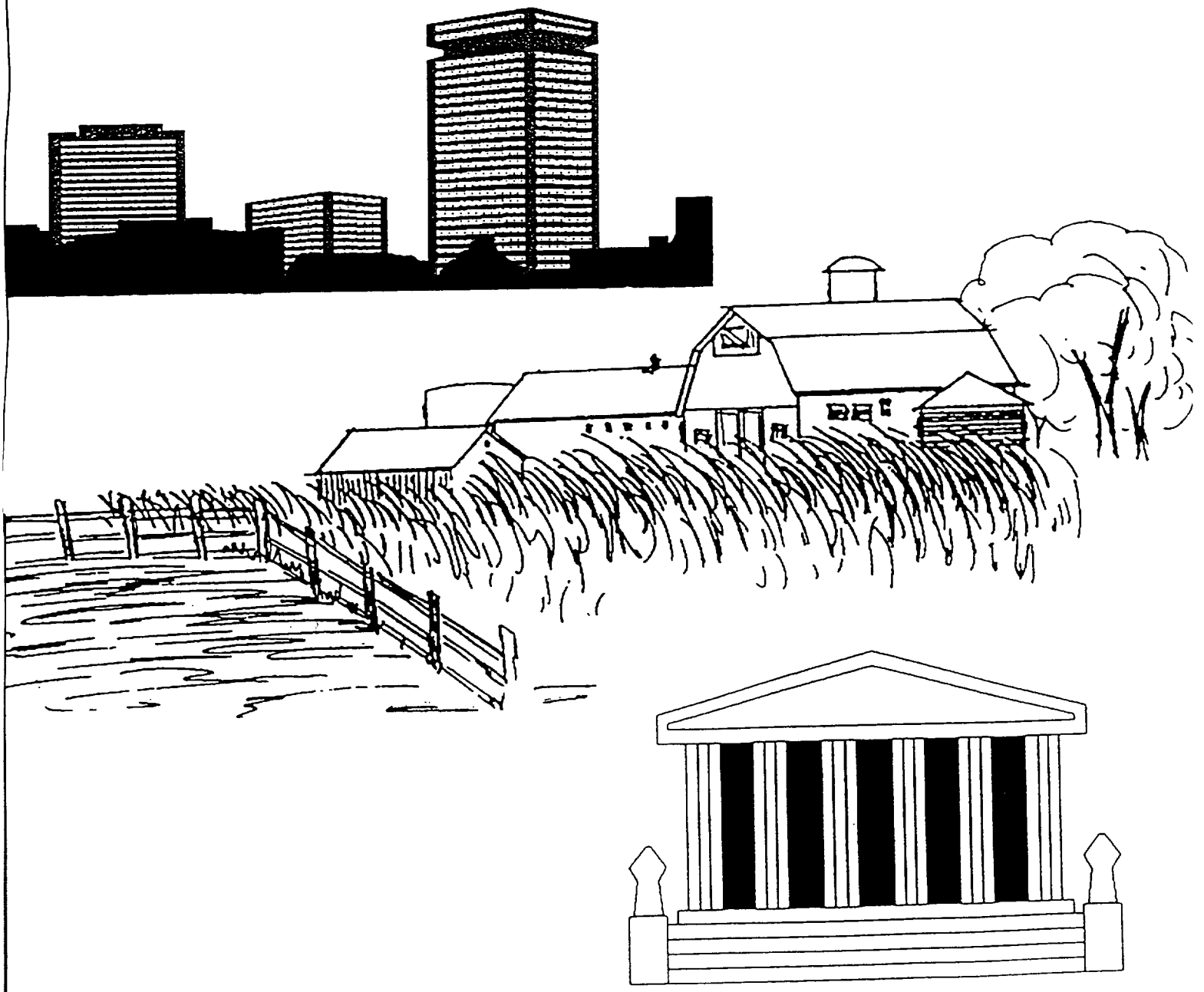
November 1993

SURVEY METHODS FOR BUSINESSES, FARMS, AND INSTITUTIONS

INTERNATIONAL CONFERENCE ON ESTABLISHMENT SURVEYS

Part II

NASS Participants



SURVEY METHODS FOR BUSINESSES, FARMS, AND INSTITUTIONS by Participants, National Agricultural Statistics Service, U.S. Department of Agriculture, Washington, DC 20250-2000, November 1993, Part 2 of 2, Report No. SRB-93-11.

ABSTRACT

Part 2 of this report is a compilation of all invited and monograph papers presented at the International Conference on Establishment Surveys in Buffalo, New York, June 28-30, 1993. These were requested for the meeting due to there general content regarding NASS policy, program and procedures. These are also ordered by general subject matter. Several of these will be printed as separate and more detailed research reports. Part 1 of this report included reports of specific aspects of NASS research.

ACKNOWLEDGMENTS

This report has been the culmination of the efforts of many individuals in NASS. Special thanks go to Yolanda Grant and Shawna McClain for their many hours of assistance in preparing for the conference and this report.

TABLE OF CONTENTS

THE EVOLUTION OF AGRICULTURAL DATA COLLECTION IN THE UNITED STATES[1]

Rich Allen[2]
Deputy Administrator for Programs
United States Department of Agriculture
National Agricultural Statistics Service

Vicki J. Huggins[2]
Branch Chief
United States Census Bureau
Agriculture Division

Ruth Ann Killion[2]
Division Chief
United States Census Bureau
Statistical Research Division

INTRODUCTION

Farms represent a special subset of establishments. They, more than any other type of establishment, combine business and family considerations. Many farms are operated by a single person but others involve quite complicated ownership and operation arrangements. Agriculture includes many different forms of products and production practices which often change over time.

This chapter traces the development of agricultural data collection in the United States from 1790 to the present time. The discussion should be helpful to others for developing or improving collection of agricultural data. The goal is to summarize major issues and developments. For more detailed reading of historical developments, please see references *Bidwell and Falconer (1925)*, *Brooks (1977)*, *National Agricultural Statistics Service (1989a)*, *Statistical Reporting Service (1969)*, *Taylor and Taylor (1952)*, and *Wright and Hunt (1900)*.

Major changes in methodology are described with explanations of factors which might have allowed or demanded the changes. Methodology for both agriculture censuses and the current agricultural survey programs have changed extensively as the United States evolved

[1] The authors have benefitted from several reviews at various stages of the writing and editing process. Reviewers included Doug Miller and Jim Liefer of the Census Bureau; Kenn Inskeep of the Evangelical Lutheran Church in America; Nanjamma Chinnappa and George Andrusiak of Statistics Canada; Bill Arends, Jerry Clampet, Bob Schooley, Paul Bascom, and Lue Yang of the National Agricultural Statistics Service. The authors also want to recognize Mary Ann Higgs for her tireless administrative support and Priscilla Simms, Marsha Milburn, and Hazel Beaton for the additional typing and graphics assistance.

[2] The views expressed are attributable to the authors and do not necessarily reflect those of the Census Bureau or of the National Agricultural Statistics Service.

from an overwhelmingly rural and agricultural nation to one which is highly urbanized with only a small fraction of the population involved in agriculture.

Uses of Agricultural Statistics

Early requirements of agricultural statistics in the U.S. stemmed from farmers' need to know about current and improved farming practices for subsistence reasons. Farmers were primarily concerned with increasing production to improve their own family's food and fiber supplies and to purchase goods that they could not produce. Local and state farming societies were interested in farming methods since farming products and practices tended to be similar by geography. Gradually, the need for national level agricultural statistics increased as better marketing channels opened, production expanded westward into new farming territories, production levels and commercialization increased, and pricing arrangements evolved.

For many years published agricultural statistics were designed largely to provide government with production information, and this remains a major purpose. These data are essential for administering farm income support and disaster relief programs and developing new legislation. Since much agricultural production is exported, current statistics enable orderly marketing in both domestic and foreign trade channels.

Many national and state programs utilize combinations of census and current statistics for allocating funds for extension services, research, and soil conservation projects. Private industry uses production data for facility construction and marketing decisions. Purchasers of farm products, suppliers of farm inputs and services, producers, and producer organizations need good information.

Present Structure of Official Agriculture Data Collection

The agriculture census program of the Census Bureau and the current statistics program of the National Agricultural Statistics Service (NASS)[3] complement each other quite well. For example, data collected by NASS on prices, production levels, marketing patterns, inventories, and production expenses are used in farm income estimates which are part of the U.S. national accounts. When census data are available, farm income estimation model components are revised and benchmarked to current census levels.

The Census Bureau is located within the Department of Commerce while NASS is an agency in the Department of Agriculture (USDA). Since the two sets of statistics are created by different organizations, there is no internal pressure that the census must support what the current statistics have been indicating. There is a good working arrangement and cooperation between NASS and the Agriculture Division of the Census Bureau. NASS aids in census mail list creation, and its probability area frame survey data are important

[3] NASS was previously called the Statistical Reporting Service (SRS). NASS is used throughout this chapter for consistency.

indicators of list coverage. NASS staff assist in reviewing state and county census data during the editing phase. The two organizations also participate in joint research projects.

Census Bureau Current Program

Census data collection, currently conducted by the Census Bureau every five years for years ending in two and seven, represents the only effort to collect a broad set of data from all farms. Farm operation and family characteristics are collected along with data on acreage, land use, irrigation, crops, vegetables, fruits and nuts, nursery and greenhouse products, livestock and poultry, value of sales, type of organization, farm workers, and characteristics of operators. The census provides the only set of small area data for farm characteristics and minor production items.

The agriculture census is currently a mailout/mailback enumeration of all farms and ranches. Response is mandatory and extensive follow-up procedures are conducted by mail, certified mail, and telephone to improve response rates. Additional data on items such as use of commercial fertilizers; use of chemicals, machinery, and equipment; expenditures; market value of land and buildings; and farm related income are collected from an approximate 30 percent sample. A stratified, systematic sample design at the county level is utilized to ensure reliable county level statistics.

Following each agriculture census, the Census Bureau has generally conducted one or more follow-on surveys to collect detailed data on relatively narrow areas of interest without adding greatly to overall respondent burden. Past surveys collected information on topics such as land ownership, farm expenditures, farm finances, farm and ranch irrigation practices, and farm energy use. There also have been censuses of horticulture specialties. See Table 3 for a chronological listing.

National Agricultural Statistics Service Current Program

The United States agricultural estimating system for current statistics utilizes voluntary sample surveys to measure production and inventories of major crop and livestock items along with information on prices, labor usage, and disposition of livestock. Expected production levels of major crops are forecasted monthly, starting three or four months before harvest. Because many statistics can have an impact on volumes and prices for commodity futures and cash market trading, strict procedures and laws govern the preparation and release of these statistics to the public.

Many different survey vehicles are used, with some programs conducted weekly, monthly, and quarterly in addition to surveys conducted on an annual or semi-annual basis. U.S. and state level estimates are normally produced but county estimates are created for items needed to administer federal and state farm programs. Most surveys are based on probability samples but some utilize designated panels of reporters and others are based on contact of individuals who can answer for an entire segment of a specific population. A few surveys contact agribusinesses such as grain elevators, food processors, seed companies, livestock packing houses, and feed companies when they can report data more efficiently

than farmers. Administrative data such as exports, imports, and marketings are utilized either directly or as a check on estimation levels.

Nearly 400 statistical reports are issued from NASS Headquarters each year on a schedule published before the year starts and more than 9,000 reports are issued by State Statistical Offices. Six of the current statistics series are included in the Principal Economic Indicators Series of the U.S. *Statistical Reporting Service (1983)*

Farm Definition

The definition of a *farm* used in the United States has changed over time. The Secretary of Commerce sets the definition after collaboration and agreement with all relevant agencies (Census Bureau, Bureau of Economic Analysis, USDA, etc.) and advisory groups. The definition is not codified in U.S. law. Since 1975, the definition has been *any place which sells, or normally would sell, $1,000 or more of agricultural products in the reference year.* Because of the relatively low value threshold, individuals who keep just a few head of cattle or sell some fruits and vegetables qualify even if they do not regard themselves as farmers. The greatest difficulty in conducting a complete census or representative survey is in accounting for the vast number of small farms. In the 1987 census, approximately 50 percent of all farms had less than $10,000 in total value of agricultural products sold, 60 percent had fewer than 180 acres of land and 30 percent had fewer than 50 acres. *Bureau of the Census (1989).*

Table 1 provides the farm definitions used during different periods of U.S. history. Although defining a farm appears to be a simple task, it requires a tremendous amount of discussion, analysis, and communication to satisfy federal, congressional, and private concerns. For example, for the 1974 census, USDA and Commerce both supported raising the threshold to $1,000 because of increased prices and changes in the structure of agricultural operations. However, Congress disagreed and in 1974 census all-farm preliminary reports were published using the same farm definition as the 1959-1969 censuses. Only the final reports were based on the new definition with a threshold of $1,000. *Bureau of the Census (1979)*

Definition of Counties and Townships

The major geographic administrative subdivision within a state is the *county*. There are over 3,300 counties within the 50 states of the U.S. Their size varies depending upon the physical size, geography, and population density of the states. Many counties are divided into *townships* of about 36 square miles each. County and township reports on agriculture were important at various times in history and counties remain important units for publication of both census and current agricultural statistics.

Table 1 The U.S. Farm Definition Throughout History

Period(s)	Definition
1850-1860	No acreage requirement, but a minimum of $100 of total value of agricultural products sold (TVP).
1870-1890	A minimum of 3 acres or $500 TVP.
1900	Acreage and minimum sales requirements removed.
1910-1920	A minimum of 3 acres, with $250 or more TVP or required full-time services of at least one person.
1925-1945	A minimum of 3 acres, with $250 or more TVP.
1950-1954	A minimum of 3 acres or $150 or more TVP. If a place had sharecroppers or other tenants, land assigned to each was treated as a separate farm. Land retained and worked by the landlord was considered a separate farm.
1959-1974	Any place with 10 acres or more, and with $50 or more TVP, or any place with less than 10 acres, but at least $250 TVP. If sales were not reported, average prices were applied to reported estimates of harvested crops and livestock produced to arrive at estimated sales values.
1978-Present	Any place that had, or normally would have had, $1,000 or more TVP.

AGRICULTURAL STATISTICS BEFORE 1880

In 1790, the United States was essentially an area along the Atlantic Ocean with an average width of only 255 miles from the coast and a population of less than four million people, over 90 percent involved with agriculture. Tobacco was the principal agricultural export and, valued at $4.4 million, accounted for about 44 percent of total exports. Farm operations in the northeastern states centered around villages, were often large, multiple family plantations in southern states, and were isolated farmsteads in middle states. *Economic Research Service (1993)*

Before 1840, the only agricultural information available for extensive areas came through asking knowledgeable individuals for their assessment of current conditions and supplies of crops and livestock. An interesting sidelight in American history is that President George Washington is perhaps responsible for the first agricultural survey. Washington wrote to friends in several states in 1791 listing agricultural questions that he wanted them to answer. This is the first known occasion when an actual survey form was used to collect expert opinions. Collection of expert opinions was the major source of agricultural statistics for the next 100 years, except for the agriculture censuses. *Statistical Reporting Service (1969)*

There were few advances in the collection of agricultural statistics during the first third of the 19th century because agriculture was mostly for subsistence. The need for more broad based collection of agriculture statistics was not apparent. An exception was the South which was more commercially oriented and needed information for marketing. Some individual

states and agricultural societies collected and provided agricultural information on a small scale to local farms and businesses.

By the middle of the 19th century, the country was growing rapidly and expanding westward, agricultural surpluses existed in some areas, and state efforts to collect agricultural information were uncoordinated. The need for national level data became more obvious.

Archibald Russell, a young scholar who prepared a book of proposals for collecting U.S. resource data with the 1840 population census, wrote that an economic understanding of the U.S. could be greatly helped by agricultural data since agriculture was the leading business. In 1838, President Martin Van Buren recommended to Congress that the 1840 Census of Population be expanded to collect information "...in relation to mines, agriculture, commerce, manufacturers, and schools..." As a result, 37 questions about agriculture were included with the 1840 population census. At that time, slightly over 70 percent of the U.S. population reported that they were engaged in agriculture. All information was collected and recorded by personal interview. The questions concentrated on production without collecting area harvested or yield which limited comparisons or linking of other information to the census. *Wright and Hunt (1900)*

Table 2 shows the range of questions and the increase in the number of questions relating to agriculture in the decennial years of 1840 to 1890.

In 1839, the U.S. Congress appropriated $1,000 to the Patent Office for collecting agricultural statistics on new varieties of seeds such as wheat and other agricultural information. It was hoped that providing annual statistics would guard against monopolies or exorbitant prices. The Patent Office issued its first report for 1841 by linking back to the 1840 census. Since census estimates of crop areas harvested were not available, estimates of population change in each state and territory were used as indicators of production changes. Other information came from agricultural societies, agricultural papers, and knowledgeable individuals.

The Patent Office attempted to track crop failures or other unusual conditions in making its annual reports. However, in 1849 the Patent Office came under new leadership which did not support agricultural statistics except for prices and the first continual series of production estimates ended. Even with an outcry to reinstate the crop reports, no further attempt was made to provide annual production estimates. Orange Judd, editor of the American Agriculturist journal, described the Patent Office as a "seed store in Washington." *Statistical Reporting Service (1969)*

Table 2 Number of Agriculture Census Questions, 1840 to 1890

Items of Inquiry	1840	1850	1860	1870	1880	1890
Name of person conducting farm		1	1	1	1	1
Color of person conducting farm						1
Tenure					3	3
Acres of land		2	2	3	4	5
Acres irrigated						1
Value of implements, machinery, and livestock		3	3	3	3	3
Cost of building and repairing fences					1	1
Cost of fertilizer purchased					1	1
Wages paid for farm labor				1	1	1
Weeks of hired labor upon farm					1	2
Estimated value of all farm productions				1	1	1
Forest products	7			1	2	2
Grass lands and forage crops	1	3	3	3	5	22
Sugar	1	3	3	3	8	17
Cereals: barley, buckwheat, corn, oats, etc.	6	6	6	7	12	27
Rice	1	1	1	1	2	3
Tobacco	1	1	1	1	2	4
Peas and beans		1	1	1	2	4
Peanuts						3
Hops	1	1	1	1	2	4
Fiber: cotton, flax, hemp, broomcorn, wool	3	6	7	5	12	19
Horses, mules, asses, sheep, goats, dogs	2	3	3	3	10	22
Neat cattle [a] and their products	1	3	3	3	8	10
Dairy products	1	2	2	3	3	11
Swine	1	1	1	1	1	4
Poultry and eggs	1				3	8
Beeswax and honey	1	1	2	2	2	4
Onions						4
Potatoes	1	2	2	2	4	6
Nurseries, Orchards, and Vineyards	5	2	2	2	12	42
Other	3	4	4	4	1	19
TOTAL	37	46	48	52	[b] 108	[b] 55

[a] Neat cattle includes working oxen, beef cattle, and dairy cows.

[b] Number of inquiries or details called for in the general schedule of agriculture only, additional inquiries on special schedules of agriculture, not common to the general schedule or other special schedules, are not included. *Wright and Hunt (1900)*

Because of complaints about many errors found in the 1840 census, the 1850 census included some major changes in organization and data content. Legislation was passed to clearly define the duties of persons employed by the census and the consequences of neglecting their duties. The general organizational structure initiated in the 1850 census has continued through present day censuses. The Census Bureau established a farm definition as any place that had $100 or more in value of sales of agricultural products. A temporary census board within the Department of the Interior was established to oversee the conduct of the 1850 census, a procedure which was followed until 1900. *Wright and Hunt (1900)*

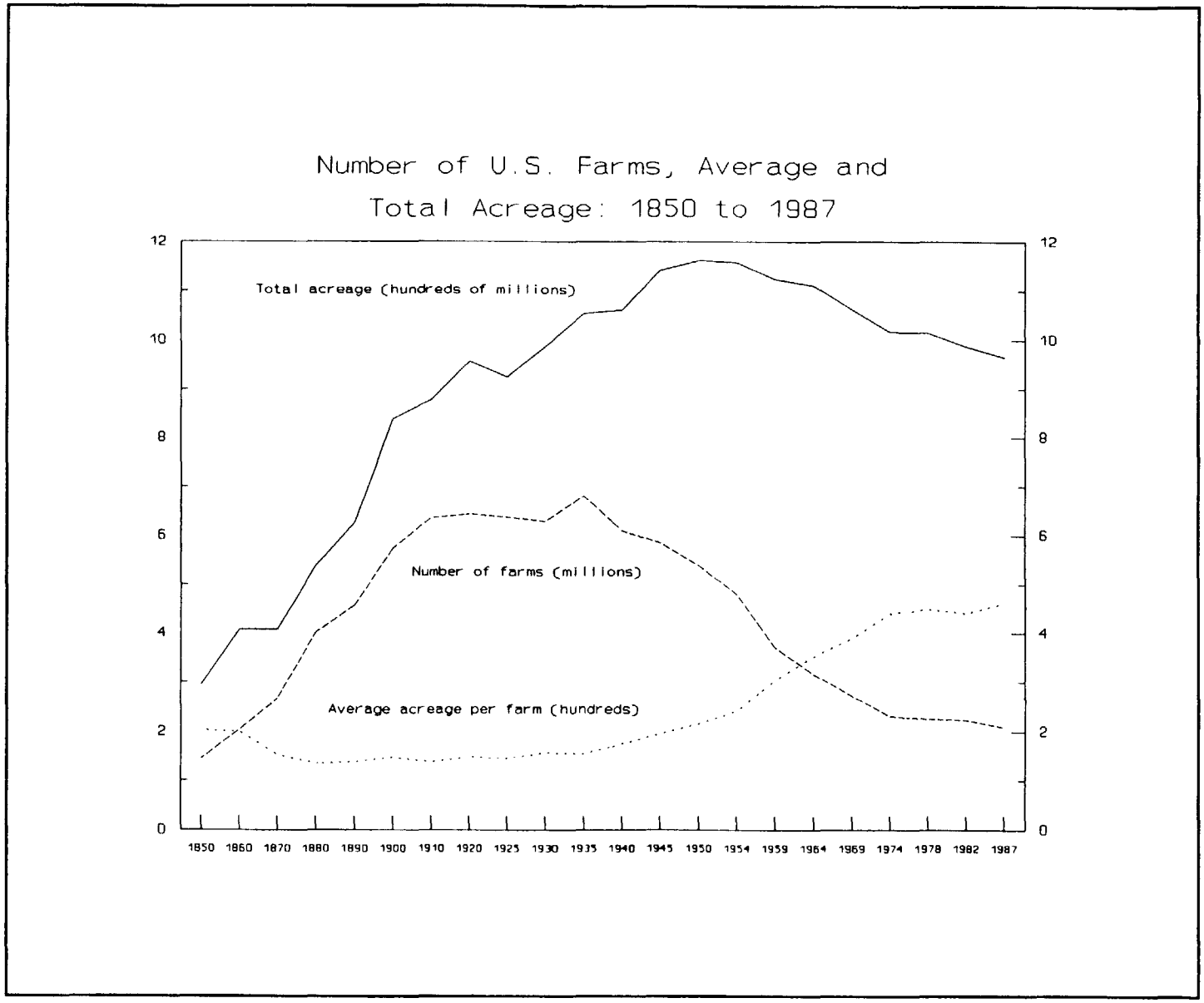
By 1860, commercial corn and wheat production was well underway with the west developing these crops rapidly. The Northern States began developing other agriculture interests such as dairying and feed crops. Cotton surpassed tobacco as the major agricultural export crop with its $100 million or so average value accounting for half of all export value. *Economic Research Service (1993)*

There was increased foreign migration into the country as a result of the potato famine in Ireland and the German Revolution of 1848 and many immigrants were lured into agriculture. Scientific invention was applied to create and improve farm machinery. The nation became increasingly more dependent on national and international markets as it moved from subsistence farming toward commercialization. *U.S. Census Office (1902)*

The average farm size in 1850 was just over 200 acres. However, because of a law called the Homestead Act which made 160 acres of new agricultural land available to settlers, and the breaking up of southern plantations during the 1860's, the average farm size declined and it was not until 1950 that it again exceeded 200 acres. See Graph 1.

Between 1849 and 1862, when there were no federally collected annual statistics, agricultural societies continued to publish their "interpretations" or estimates. A new Commissioner of Patents in 1856 encouraged governors and other prominent individuals to make estimates for their areas. However, none of these efforts resulted in a consistent collection of statistics.

The present day program of current agricultural statistics traces its development to events about 1860. In 1862, Orange Judd asked each town to select a person to fill out a monthly form on crop area and crop prospects. Respondents were asked to evaluate the current month's data relative to a base of ten as an average crop and to reply in whole numbers, with each unit above or below ten indicating a 10 percent departure. Mr. Judd received between 1,000 and 1,500 responses for various months in 1862. This established the pattern of monthly information used by USDA to the present day. Mr. Judd's initiatives and foresight broke critical ground for ensuring orderly economic commerce of agriculture.


Number of U.S. Farms, Average and
Total Acreage: 1850 to 1987
Total acreage (hundreds of millions)
Number of farms (millions)
Average acreage per farm (hundreds)
0
2
4
6
8
10
12
1850 1860 1870 1880 1890 1900 1910 1920 1925 1930 1935 1940 1945 1950 1954 1959 1964 1969 1974 1978 1982 1987

Mr. Judd's survey was discontinued in 1863 when Congress established USDA which started collection of statistics. This was in response to a grass roots clamor throughout the nation, swelled by the agriculture press, for better data than were being produced by the Patent Office and the decennial census. USDA's first report, using the methods of Orange Judd, covered estimates for 1859 and 1862 based on 1860 census figures (crop year 1859) and data received on questionnaires sent to every county during the winter of 1862-63.

The Office of the Statistician in USDA was created in 1863 and monthly reports of crop conditions were started immediately. One reporter and five assistants per average size county were selected to provide local agriculture assessments. Reports were sent by mail to Washington, D.C. on a designated day each month. The purpose of monthly crop reports was espoused by USDA in 1863 as follows: "Ignorance of the state of our crops invariably leads to speculation, in which oftentimes, the farmer does not obtain just prices, and by which the consumer is not benefitted...the true condition of these crops should be made known. Such knowledge, while it tends to discourage speculation, gives to commerce a more uniform and consequently, a more healthy action." *Statistical Reporting Service (1969)*

While the early emphasis was on crop production, cattle, hog, and sheep numbers and values were estimated starting on January 1, 1867, along with wheat and corn prices for 1866. Price data continued to be annual estimates up until 1908 when monthly prices were instituted. *National Agricultural Statistics Service (1989b)*

There was little change in the agricultural data collected in the 1860 and 1870 population censuses. There was increased call for information on the acreage of crops such as wheat, barley, and oats, but this information continued to be omitted from the census until 1880. There were some inquiries in the 1860 census relating to "what crops are short," and "usual average crop," that may have been useful in crop prediction for intercensal years, but these questions were dropped in the 1870 census. The first statistical atlas was published based on the 1870 results which showed, for example, the geographic distribution of crop production. [See Table 2 for the description of questions in the 1860 and 1870 censuses.]

AGRICULTURAL STATISTICS FROM 1880 TO 1915

By 1880, large cattle operations and wheat production were established on the Great Plains. Only 50 percent of the population was employed in agriculture but agriculture still accounted for over 75 percent of all U.S. exports. Improved tools for production and harvesting had been developed and were in wide use. *Economic Research Service (1993)*

The census of 1880 marked a turning point in agricultural statistics. Cooperation between the census and current statistics programs was at a peak. J.R. Dodge, Chief, Division of Statistics in USDA, served as Chief Statistician for development of the 1880 census. He was given considerable freedom in revising the census questionnaire and added questions on area harvested and yield. These data were important as a census benchmark as well as to assist in intercensal predictions. The basic questionnaire had 108 questions about agriculture while special schedules contained a total of 1,572 questions. *Statistical Reporting Service (1969)*

The 1880 results were presented in both cartographic and tabular forms. The 1880 census asked comprehensive questions on renting arrangements and mechanization. One other significant feature, which aided interpretation of current statistics surveys, was collection of information on average and largest cereal crop yields by "region" or "locality." *Wright and Hunt (1900)*

Periodic reports being issued by USDA were aided by the 1880 census information on acreage and yields and a shift was made to estimates of actual yield at the end of the growing season. In 1884, reporters were asked to consider 100 as condition of a full crop, not an average crop. Another important improvement in procedures was the 1888 initiation of state weighted averages based on county acreage, since counties varied greatly in area and production potential.

Bronsen C. Keller, an economic and social scholar, organized an association with two other men from St. Louis. These three men had a large impact on the 1890 census by sending a letter in 1889 to people around the country asking them to insist that indebtedness information be collected. They believed that indebtedness was leading to loss of farms through foreclosure and hence decreasing ownership. Through this grass roots effort, five questions on indebtedness were collected on the population census and the results were classified by farm and nonfarm homes.

As agricultural statistics got more attention and the number of users increased, the accuracy of monthly crop reports began to be questioned. In 1895 the National Board of Trade passed a resolution stating "Whereas the monthly and yearly crop reports of the U.S. Department of Agriculture have in recent years been confusing, misleading and manifestly erroneous in important particulars...that if the crop reporting services...is to be continued...every needful effort be made for ensuring the fullest degree of efficiency...completeness and accuracy of the data..." *Taylor and Taylor (1952)*

The National Board of Trade established a Committee on Crop Reports which made several proposals for improvement of agricultural statistics. Based on its recommendations a law was passed by Congress in 1909 making it a crime to divulge any information ahead of a scheduled release. Official township reporting was recommended and, by 1896, there were 28,000 township reporters, 9,000-10,000 county reporters, and 6,000-7,000 assistants to state statistical agents. Reports were also received from 15,000 grain dealers, millers, and elevator operators and 123,000 farmers. Interpreting the vast amount of information received was a difficult task. In 1905, the Crop Reporting Board was established to improve interpretation. This Board, consisting of state statisticians and experienced headquarters statisticians assisting the Chief Statistician in reviewing indications and setting estimates, has continued for major reports since.

The 1900 agriculture census general schedule was similar to 1890, but added questions on tenure, total value of farm buildings, and ownership of rented farms. Congress provided guidelines for publication which required that data be tabulated by race and gender of farm operators. There were several processing innovations introduced for this census: punch cards and electronic tabulating machines were adopted and, because of the large number of

crop cards, a sorting machine was developed. A new ten-key keypunch machine was used for farm census cards 20 years before it was used for the population census.

There were large differences between USDA estimates and the 1900 census results for crop acreage and production. In almost all instances, USDA estimates were significantly lower than census results. Statistics for the number of acres for two of the nation's largest crops, corn and wheat, differed by 16 percent and 18 percent, respectively. A committee of inquiry was set up to investigate this and found that the census provided a low base in 1890 and that the USDA estimates for subsequent years included accumulated error from faulty yearly percent change ratios applied to the census base. The committee recommended improvements in training of census enumerators and clerks, improved editing procedures, and more verification during processing and tabulation of the data. They also recommended that a census of population be conducted every five years, especially for the collection of agricultural information. *Statistical Reporting Service (1969)*

In 1902, a permanent Bureau of the Census was established in the Interior Department. It was transferred to the new Department of Commerce and Labor in 1903. When that Department split in 1913, the Census Bureau was placed in the Department of Commerce. Establishment of a permanent bureau created a more stable environment for the census program which promoted better planning, comparability between censuses, evaluation opportunities, more time for systems development, and a basis for producing additional statistics upon demand.

Specialized censuses on irrigation and on drained land were added to the agriculture program. These two censuses remained as part of the program through 1950. Table 3 lists special censuses and surveys that have been conducted by the Census Bureau since 1890. *Bureau of the Census (1987)*

AGRICULTURAL STATISTICS FROM WORLD WAR I TO WORLD WAR II

By 1915, gasoline powered tractors and combines were being developed for the extensively farmed areas. New varieties and disease resistant strains of plants were being developed. The average value of U.S. agricultural exports was approaching $2 billion a year, about 45 percent of all exports. Only about 30 percent of the population was now engaged in agriculture but the need for more commercial agricultural information was growing. There was increased awareness of the importance of marketing and a pressing need for reliable information on supplies of food and fiber. *Economic Research Service (1993)*

Table 3 Census Bureau Special Censuses and Surveys by Reference Year

Year	Title(s)	Year	Title(s)
1890	Census of Horticulture	1956	Farm Mortgage Indebtedness
1905	Cotton Ginnings [a/]	1959	Census of Horticulture, Census of Irrigation, Irrigation in Humid Areas, Census of Drained Land
1910	Census of Irrigation		
1920	Census of Irrigation, Census of Drained Land	1964	Survey of Farm Workers, Survey of Hired Farm Workers, Survey of Farm Indebtedness
1930	Census of Horticulture, Census of Irrigation, Census of Drained Land, Farm Mortgage Indebtedness [b/]	1965	Survey of Nonfarm Income and Source
		1969	Census of Horticulture, Census of Irrigation, Census of Drained Land
1935	Survey of Part-Time Farming, Farm Mortgage Indebtedness		
1940	Census of Irrigation, Census of Drained Land, Farm Mortgage Indebtedness	1970	Agriculture Finance Survey
		1979	Census of Horticulture, Farm Finance Survey, Survey of Farm Energy Use, Farm and Ranch Irrigation [b/]
1945	Farm Mortgage Indebtedness		
1950	Census of Irrigation, Census of Drained Land, Census of Horticulture, Farm Mortgage Indebtedness	1984	Farm and Ranch Irrigation
		1988	Farm and Ranch Irrigation, Census of Horticulture, Agricultural Economics and Land Ownership Survey
1955	Irrigation in Humid Areas [b/]		

a/ The Cotton Ginnings Survey collected data twice a month through the ginnings season every year from 1905 up to the present. The Survey was conducted by the Bureau of Census until 1991; now conducted by National Agricultural Statistics Service.

b/ The titles of the Survey of Farm Mortgage Indebtedness, the Survey of Irrigation in Humid Areas, and Farm and Ranch Irrigation Survey have been shortened to better fit the table format.

Several important improvements to current agricultural statistics were made between 1910 and 1920. Official annual estimates of the numbers of farms by state began in 1910. Monthly condition information was converted to yield forecasts starting in 1911 for crops like wheat, oats, corn, and tobacco. Once the validity of the monthly forecasts was proven, more market volatile crops such as cotton were added. The *par* procedure was developed to interpret yield by adjusting the ten-year average yield per acre by the ratio of current condition to average condition. As agricultural statisticians searched for more mathematically based yield forecasting procedures, regression models based on year-to-year relationships were developed. By 1929, regression models were in common use for monthly yield forecasts. *Statistical Reporting Service (1969)*

An important factor in heightening interest in agricultural statistics was the onset of World War I in 1917. More information was needed on available food and feed supplies at county, as well as state levels. By 1920, state and national estimates on 29 crops were being produced, compared to 13 crops ten years earlier, and condition reports were being issued on 44 crops -- about double that in 1910. *Statistical Reporting Service (1969)*

The chief of the Bureau of Statistics (USDA) reported in 1919 of the requests for agricultural data during World War I that "A vast amount of information was compiled and furnished in response to inquiries received by telephone, telegraph, letter, or personal call of representatives of the Food Administration, the War Trade Board, the War Industries Board, the Military Intelligence Office of the War Department, the Tariff Commission, the Federal Trade Commission, the Council of National Defense and other departments of federal and state governments, congress and private individuals...". Following the armistice of November 11, 1918, the demand for special agriculture information declined, but demand for food shipment to war torn Europe continued. There was a general reluctance to discontinue the products and services provided by USDA during the war, so many of them continued. *Statistical Reporting Services (1969)*

Many states developed various statistics programs which meant that different state and federal estimates might be published for the same commodity and respondents might be contacted by multiple organizations for the same information. Because of the emphasis on more agricultural statistics during World War I and respondent burden issues, state and federal organizations in the state of Wisconsin agreed in 1917 to share expenses of data collection and publication. Many states quickly followed suit and today agreements exist in every state involving State Departments of Agriculture or state agricultural universities or both. These agreements are truly unique within the United States. State funding provides for special publications, surveys, or other services not covered by the federal program. All employees (state or federal) work under the NASS selected state statistician who usually has a State Department of Agriculture title and state government duties.

By 1918, to improve reliability of data, sampling was shifted largely from panels reporting for their locality to panels reporting for individual farms. This provided a more clearly defined basis for comparison of production levels. This sampling of farmers in every township became the monthly Farm Report Survey. Special livestock reporter lists were also established. In 1922, looking for methods to improve livestock statistics, a proposal was adopted to utilize free delivery of mail in rural areas to provide a broader base of reporters for major reports. Rural mail carriers received a supply of card type questionnaires to drop off at a sample of farms along mail delivery routes. Completed responses were forwarded to the state statistician by the Postal Service.

Rural carrier cards collected data for only the current year. Before that time, information was requested for both current and past year and a current compared to historic interpretation was made. The main basis for interpretation of rural carrier acreage surveys was calculation of ratios of acreage

of specific crops to total cropland or to land in farms. However, ratios were usually biased on the high side since progressive farms tended to be sampled and farmers without row crops often did not report. The approach utilized to adjust for bias was the *ratio relative* calculation: current year ratio of a crop to total land divided by the previous year's ratio to estimate true change. An effort was made by 1928 to look at reports from the same farm on subsequent surveys. Matching provided valuable information on year-to-year changes, but was extremely time consuming and difficult.

From the beginning, there had been complaints that an agriculture census every ten years was not adequate for an industry with such large fluctuations. Yearly estimates for crop and livestock items by USDA relied on the agriculture censuses for a new base or benchmark each ten years. Occasionally, due to the methodologies for projecting yearly changes and any initial bias in census figures, annual estimates differed widely from the data for the same year available later in farm census findings -- for example, the previously described differences observed in 1900. Data users engaged in agriculture research, farm management, and business investments needed more current census information on agriculture activities. *Statistical Reporting Service (1969)*

Proposals were floated for annual agriculture censuses or at least a census every five years. In 1909, Congress mandated that the Department of Commerce conduct a mid-decade agriculture census. However, because of World War I preparations, the first five-year Census of Agriculture was not conducted until 1925. The five-year census greatly assisted the annual crop and livestock estimates and provided improved information for decision making.

In addition to providing measures of production, the census of agriculture provides data on the effects of technological changes on agriculture and on social and economic characteristics of farm operators. A significant amount of new content was added to the census questionnaire in 1920. Farm operators were asked if they had gas or electric lighting in their homes and if they owned tractors, automobiles, or trucks. Additional questions in 1930 and 1940 queried farmers about the kinds of roads adjoining their farms, whether telephones were available, and the presence of new equipment such as combines and milk machines. The presentation of data by type of farm starting in 1930 was a valuable contribution to the analysis of agriculture production. It provided a basis for much discussion and planning on the needs of farms in the early 1930's.

Socioeconomic questions on topics such as hired farm labor, farm versus nonfarm employment, income, race, and tenure of farm operators were asked in all 20th Century censuses. In 1920, for example, the census found that 61 percent of the rural population and 30 percent of the total population were engaged in farming. *Bureau of the Census (1983)*

Other expansions of statistics were made around 1925. The number of crops included on monthly farm reports nearly doubled. Questions on milk cows and milk production, hens and layers on farms, numbers of eggs produced, and farm labor were added to the monthly farm report. The percentage of U.S. farms having various livestock species is shown in Table 4 with comparisons at intervals of about ten years. As percentages of farms with milk cows and chickens declined, those questions were removed from the monthly farm report in the 1970's and specific surveys were developed for these data. *Bureau of Agricultural Economics (1933)*

Table 4 Percent of Farms with Livestock, Censuses of 1910 to 1987

Item	1910	1920	1930	1940	1950	1959	1969	1978	1987
All Cattle	83.1	83.1	76.4	79.4	75.5	72.1	63.0	59.6	56.3
Milk Cows	80.8	69.2	70.8	76.2	67.8	48.3	20.8	13.8	9.7
Hogs and Pigs	68.4	75.2	56.2	61.8	56.0	49.8	25.1	19.7	11.7
Chickens	87.7	90.5	85.4	84.5	78.3	58.5	17.3	10.7	6.9

The 1930's brought the next major changes in agricultural estimates. During the period of extremely dry weather, floods in the South, and critical economic conditions in the U.S. called The Great Depression, there was overproduction of some agricultural products, particularly hogs, resulting in very low prices. Federal farm relief was demanded to help pull the country out of the agricultural slump that began after World War I.

Many government emergency programs were established to provide financial support to farmers. One 1933 program called for controlling the supply of hogs by selective destruction of a portion of supply. Thus, good information on supplies was essential. In less than two years, 90 additional professional staff members were hired for State Statistical Offices specifically to develop hog estimates county by county. *Brooks (1977)*

The 1933 Agricultural Adjustment Act and its successor, the Soil Conservation and Domestic Allotment Act of 1936, were critical milestones in the government's approach to agricultural policy and to the statistical work necessary to support it. USDA was given unprecedented authority and funds to alleviate distress situations in agriculture. *Statistical Reporting Service (1969)*

One shortcoming of all procedures used through the 1930's was that up-to-date lists of farms were not available and no other probability sampling frame existed. In 1938, research to divide the entire land area of the U.S. into sampling units began. The area sampling approach showed promise and, in 1943, USDA and the Bureau of the Census jointly funded work at Iowa State College (now Iowa State University) to create the "*master sample of agriculture.*" This master sample created segments of land with definite boundaries which contained an average of four farms per segment. The master sample was first used to measure coverage in the 1945 Agriculture Census.

The use of sampling in the agriculture census was stimulated by World War II to reduce cost and time limitations since special statistics were needed during that period. In the 1940 Census of Agriculture, data were tabulated separately for large and small farms to identify their contribution to production levels and assist in food supply decisions. Sampling as an enumeration methodology was introduced in the 1945 census, when county-level data were collected through a conventional all-farms canvass, while selected data at various geographic levels were obtained by sampling. *Bureau of the Census (1979)*

AGRICULTURAL STATISTICS FROM WORLD WAR II TO THE PRESENT TIME

The years surrounding World War II saw some of the largest productivity changes in United States agriculture. The sad condition of agriculture in the early part of the century began improving. The surplus food problem began to vanish and programs to increase production ensued. Farm production reached a high during the war, despite labor loss and difficulty in obtaining machinery.

Commercial fertilizer use tripled in a 20-year period, along with continued improvements in hybrid seeds. Use of irrigation also increased as the country set war time production goals. The structure of farms changed as *vertical integration* -- the ownership of multiple stages of the production, marketing, and distribution functions by one organization -- in the poultry industry started and new marketing techniques such as frozen foods shifted production patterns. In 1950, only about one-eighth of the labor force was farmers; this was down to 2.6 percent by 1990. Agricultural exports are currently about 15 percent of all U.S. export value. *Economic Research Service (1993)*

After the turn of the 20th Century, data users began requesting information in addition to production quantities and sales by product. In determining census of agriculture content, two contradictory issues had to be balanced: demand by data users for more detailed data and the need to keep respondent burden to a minimum to encourage adequate response. Experiments to tailor report forms to reflect different characteristics of farm operations in various regions were introduced during the 1940's and 1950's. From the 1945 to the 1959 censuses, questions were added to identify emerging farm operational patterns such as landlord-tenant operations or multiple operations owned by corporations. Technical improvements in processing also continued. Mechanical editing of data captured on punch cards began in 1940, followed by development of modern computer technology, improved the timing for publication and controlled the enormous processing responsibilities. The world's first general-purpose electronic computer, the UNIVAC system, developed to the Census Bureau's specifications and installed in 1951, was used for part of the 1950 population census, and then to process 1954 agriculture census data. *Bureau of the Census (1983)*

Until 1950, agriculture censuses used personal enumeration--farm to farm canvassing. Drawbacks were delays due to bad weather, smaller pools of census enumerators over time, and difficulty in locating absentee farm operators. For the 1950 census, the Bureau introduced mail questionnaires with questions phrased as if they were being asked by an interviewer. Questionnaires were delivered to rural route box holders, who were asked to complete the report forms and hold them until an enumerator came. This moderately successful system was used through the 1964 census.

Throughout history, the agriculture census was taken in conjunction with the decennial census, but in 1950, the agriculture program split off and mid-decade censuses were taken independent of the decennial census. In the 1970's the timing of the census of agriculture was changed to coincide with all other U.S. economic censuses in years ending in two and seven.

The use of sampling techniques in the census of agriculture program expanded with the introduction of random samples for follow-on surveys of farms with specific characteristics. Following the 1954 agriculture census, a mail sample survey of farm expenditures was

conducted and follow-on surveys such as irrigation and horticultural specialties have been included in every subsequent census of agriculture.

As the use of agricultural statistics grew after World War II, data users requested more current information. Several states had developed cooperative arrangements with the Weather Bureau and with the Federal-State Extension Service to pick up informed opinions on crop progress and fieldwork operations each week to supplement the monthly *Crop Production* reports. By 1958, this popular "weekly weather crop" approach was expanded to all states with submission of state summaries to NASS headquarters for a national release.

Other than weekly weather crop, the emphasis of agricultural statistics has been on developing probability based methodology to improve the quality and stability of estimates and forecasts. In 1957, a long range plan for improving USDA agricultural statistics was presented which called for development of a scientifically distributed area frame sample of farms to strengthen state and national crop and livestock estimates. Since 1964, a June Enumerative Survey[4] has been conducted in all states except Alaska. This survey, which yields sampling errors of 1 percent or less for U.S. estimates of major crops, became the backbone of all improved crop and livestock estimating procedures.

Area frame sampling was extremely successful for crop estimates, but did not provide the same efficiency for livestock numbers because they can vary tremendously (from zero to many thousands) in relatively small land holdings. One means of stabilizing estimates and sampling errors was creation of lists of large livestock producers who are surveyed with certainty. Because of the high costs of personal interviewing, the area frame sample is fully enumerated only once a year in June.

Another livestock survey approach, the probability mail survey, was tried in the mid-1960's. All available information for a livestock species such as hogs was used to create a list sampling frame for that species. Samples were drawn by strata which improved the stability of estimates, since lists were not complete, data expansions did not cover total production and this method was abandoned.

Ongoing internal and external research provided an improved solution in the early 1970's to livestock estimation difficulties. H.O. Hartley at Texas A&M University developed multiple frame sampling which utilized the relatively low cost of list sampling with complete universe coverage of the area frame survey. Area frame sampling is explored in more detail in *Vogel and Kott* chapter of this monograph. The June Enumerative Survey was a natural vehicle for determining completeness of a list frame. Once the base area frame survey had been conducted each year, subsequent mail or telephone surveys would include samples from the list frame with supplements derived from the area frame. In addition to livestock, the multiple frame approach was tried for grain stocks, farm labor, production of specialty crops, and was adapted in the mid-1970's to economic surveys of farm operators. Multiple frame surveys were successful in improving consistency of estimates.

---

[4] Currently called the June Agricultural Survey.

The Census Bureau first introduced a mailout/mailback enumeration procedure in the 1969 agriculture census. This method of enumeration was more cost effective and allowed farmers to complete questionnaires at their convenience, permitted unhurried access to records, and gave respondents a chance to review and correct forms before turning them into the Bureau. To ensure good response rates, six or seven mail follow-ups, as well as telephone enumeration of large farms, were conducted. *Bureau of the Census (1992)*

This approach has several problems including development of a complete mailing list and ensuring complete and timely response. Identifying small farm operators is especially a problem since they constantly enter and exit the universe and are not adequately covered by administrative lists. There is no single list source that identifies all farms. There are sources such as government farm program records, farm tax forms, State Department of Agriculture livestock inspection lists, etc., which contain farm operator names but also contain names of individuals such as landlords who are not farm operators. Some operators do not show up on any list.

Budget efficiencies, as well as the convenience of mailout/mailback, outweigh the drawbacks. The Census Bureau has evaluated coverage for each census of Agriculture since 1945. Net coverage error for number of farms has generally ranged from 85 to 93 percent. And, coverage of agricultural production has consistently been above 95 percent. See the ICES Proceedings paper by *Clark and Vacca* for more information on coverage measures for the Census and NASS agricultural programs. Despite problems with a mail census, overall coverage obtained is only marginally lower than personal enumeration conducted prior to 1969.

Major list frame development for the agriculture census program began prior to the 1969 mailout/mailback census and a major mail list frame development effort began at NASS in 1976 to support a mailout/mailback mode of data collection for their current surveys. The primary difference between the two mail list programs is that NASS built their mail list once with a capability of routine updates and maintenance whereas Census has developed a current mail list of farm operators for conduct of each upcoming census. Both mail list programs include development of computer software routines to convert and standardize name forms from multiple lists; matching all portions of names, address, and identifiers across records; prediction of the probability of farm or nonfarm status based on combinations of data sources; and creation of outputs for sampling and list maintenance purposes. The NASS mail list is a source for the census mail list as are Internal Revenue Service farm tax records.

Both agencies constantly strive to improve their mail list by using mathematical modelling to improve match success rates and reduce duplication. It is estimated that at least 20 percent of active name records on a state's list frame at NASS change in some way each year, demonstrating the high volatility in the farm universe.

Emphasis on probability survey techniques had a significant effect on data collection methods. Funds were not available for extensive personal interview followup of nonrespondents so both agencies began using telephone calls for most follow-up in the late 1970's. To improve the quality of telephone interviewing, the agencies started researching

the use of Computer Assisted Telephone Interviewing (CATI) with interactive editing about 1980. See the *Werking and Clayton* Chapter for more discussion of CATI.

Another probability methodology improvement introduced by USDA was development of procedures to determine crop yield and production by in-field visits, counts, and observations. Since the 1960's these objective yield surveys have been conducted for corn, wheat, soybeans, and cotton and procedures have been developed for a wide range of tree and field crops. These surveys depend on forecasting of number of "fruit" (ears of corn, bolls of cotton, number of hazelnuts, etc.) to be present at harvest plus a forecast of weight per fruit. Forecast models utilize historic information for the same time period and maturity stage. Objective yield surveys have been extremely successful but they are expensive since monthly on-site visits are required and they are only utilized in major producing states, usually covering 75-80 percent of U.S. production for a given crop.

Since 1972, NASS has utilized aerospace remote sensing as a data source. NASS became a leader in the automatic classification of full satellite scenes of digital data involving many million pieces of information. The June Enumerative Survey area frame segments are an ideal sample of data for training computer discrimination models and for judging the precision of classifications. If cloud free imagery can be obtained, classification of the satellite data after training usually yields sampling errors equivalent to increasing the ground based sample by three to five times. However, the satellite data can not provide information for acreage determination earlier than conventional means. Thus, the value to NASS is for review of season ending estimates of planted and harvested acreage. *Statistical Reporting Service (1983).*

AGRICULTURAL STATISTICS FOR THE 1990's AND BEYOND

The development of agricultural statistics over the years has provided many innovations in the field of statistics and data collection. From early concepts such as obtaining reports as variances from a norm, through design of keypunch and sorting equipment, matching cases for developing change estimates, the seminal work in area samples at Iowa State University, continuing with research into list frame development, use of questionnaire design techniques to improve quality of data, use of multiple frame samples and estimation, to being in the forefront with techniques such as interactive editing, the agriculture data collection community has made many contributions.

The benefits to basic statistical theory and data processing techniques will continue as agriculture data collectors address today's issues. These issues can be divided into three basic areas: management concerns, technological developments and societal changes. Both the Census Bureau and NASS, as well as other groups collecting agriculture data, will face these challenges.

Management concerns include such matters as controlling costs, ensuring appropriate coverage levels, increasing data quality, and ensuring respondent confidentiality while providing maximum data to the users. As U.S. federal budgets become more restrictive, agencies have to determine more cost effective and cost reducing measures while providing increasingly convincing arguments of the need for data collection in the agriculture sector.

As more and more data collection efforts depend on complete list frames, both U.S. agencies need to keep abreast of constantly changing operating and marketing arrangements, while maintaining substantial evaluation programs to determine the completeness of lists. The application of many federal and state laws depend on accurate agricultural sector data. Both agencies need to keep a vigilant eye on developments in order to ensure the quality, timeliness, and comparability of data. This includes such efforts as ensuring that customer (legislatures, universities, market researchers, etc.) needs are met.

The confidentiality of respondent data is of utmost importance. The goodwill of data reporters depends on the agencies ensuring that individual data will not be publicly disclosed. The Census Bureau, in particular, is developing an extensive body of theory related to cell suppression theory which protects respondent data in tabular data presentation. Application of the theory can be difficult but is necessary to ensure respondent confidence and, hence, cooperation. The counterbalance to protecting respondent data is providing useful data to the many users. The more data are suppressed to ensure respondent confidentiality, the less data are available for users. This poses a constant dilemma, since many of the users are paying for the data collection. See the *Larry Cox* Chapter on Disclosure as well as the proceedings papers by *Colleen Sullivan* and the Panel Discussion on Disclosure. The future holds many policy development and statistical innovation challenges.

The area of agricultural technological development is most challenging because it is harder to predict the issues. As technological advances are made in the agricultural community, such as sustainable agricultural practices, more direct marketing of commodities like fruits and vegetables, or growth in the number of farmers producing new crops for very specific markets, it is important to gather data about the changes. This is a task made more difficult by the need to reach some consensus about definitions before attempting measurement.

Technological advances are also occurring in the fields of data collection and statistical estimation. For example, how will data collection activities change as more and more respondents use computers and computer driven communications and technology? Data collectors have an obligation to keep up with -- or, preferably, ahead of -- such trends.

Societal issues run the gamut from maintaining acceptable response rates in a diverse society to changes in the characteristics of farms as they become larger and more small farmers leave the field. Both agencies are currently devoting efforts in the areas of questionnaire design, respondent burden, and customer needs. For example, NASS holds yearly data user meetings around the country, focusing on different types of statistics each year, and the Bureau of the Census has increased it's presence at agricultural related meetings. Again, the issue of customer expectations is difficult; meeting the needs of data user customers while not requiring too much of an ever shrinking community of data provider customers is a challenge that will require perseverance and creativity on the part of both agencies. The issues of respondent burden may force innovative solutions involving special contacts and wider use of database techniques.

The goal of both U.S. agencies is to be quality oriented in the future. The needs for data on the agricultural sector will not diminish as the federal, state, local, and private

sectors strive to meet their mandates. Planning, research, marketing, and management of farm and rural programs in this country will continue to depend on quality data collected through innovative techniques. Both agencies are determined to meet the challenges of the future, as they have in the past!

REFERENCES

Bidwell, P.W. and Falconer, J.I., *History of Agriculture in the Northern U.S.*, Carnegie Institution of Washington, 1925.

Brooks, E.M., *As We Recall: The Growth of Agricultural Estimates, 1933-1961*, Washington, D.C., Statistical Reporting Service, U.S. Department of Agriculture, 1977.

Bureau of Agricultural Economics, *The Crop and Livestock Reporting Service of the United States*, Miscellaneous Publication No. 171, Washington, D.C., U.S. Department of Agriculture, 1933.

Bureau of the Census, *1974 Census of Agriculture Procedural History*, Department of Commerce, Volume 4, Part 4, 1979.

Bureau of the Census, *1978 Census of Agriculture Procedural History*, Department of Commerce, Volume 5, Part 4, 1983.

Bureau of the Census, *1982 Census of Agriculture - History*, Department of Commerce, Volume 2, Part 4, 1987.

Bureau of the Census, *1987 Census of Agriculture - History*, Department of Commerce, Volume 2, Part 4, 1992.

Bureau of the Census, *1987 Census of Agriculture U.S. Summary and State Data*, U.S. Department of Commerce, Volume 1, Part 51, 1989.

Economic Research Service, *A History of American Agriculture, 1776-1990*, U.S. Department of Agriculture, 1993.

National Agricultural Statistics Service, *The History of Survey Methods in Agriculture (1863-1989)*, U.S. Department of Agriculture, 1989a.

National Agricultural Statistics Service, *Agricultural Production and Prices - 125 Years, A Historical Review*, U.S. Department of Agriculture, 1989b.

Statistical Reporting Service, *Framework for the Future*, Washington, D.C., U.S. Department of Agriculture, 1983.

Statistical Reporting Service, *Scope and Methods of the Statistical Reporting Service*, Miscellaneous Publication No. 1308, Washington, D.C., U.S. Department of Agriculture, 1983.

Statistical Reporting Service, *The Story of U.S. Agricultural Estimates*, Miscellaneous Publication No. 1088, Washington, D.C., U.S. Department of Agriculture, 1969.

Taylor, H.C. and Taylor, A.D., *The Story of Agricultural Economics in the United States, 1840-1932*, The Iowa State College Press, Ames, Iowa, 1952.

U.S. Census Office, *Census Reports: Twelfth Census of the U.S.: Farms, Livestock, and Animal Products,* U.S. Government Printing Office: Washington, D.C., Volume V, Part 1, 1902.

Wright, C.D. and Hunt, W.C., *History and Growth of the U.S. Census: 1790-1890*, U.S. Government Printing Office, 1900.

MULTIPLE FRAME ESTABLISHMENT SURVEYS

Frederic A. Vogel
National Agricultural Statistics Service

Phillip S. Kott
National Agricultural Statistics Service

1 INTRODUCTION

Sample surveys of economic establishments are usually designed to provide estimates of characteristics such as total sales, expenditures, number of workers, and inventories for a population of interest. Basic principles from finite population sampling theory apply regardless of the specific sampling design used. A population of elements must be defined (not always a trivial task in establishment surveys, see Colledge chapter and Nijhowne chapter), and a sampling frame must be constructed from which a sample of elements can ultimately be drawn. Every element in the population must have a known probability of selection. Ideally, the frame should also be complete; that is, the selection probability of every element in the population should be positive.

The general term "element" is used to mean a statistical unit in the sense of the Colledge and Nijhowne chapters. In many establishment surveys, but certainly not all (see Subsection 3.1), the population of elements is identical to the population of establishments of interest.

Populations of establishments, whether of farms, retail stores, factories, buildings, schools, or governments often possess common characteristics that impact on the choice of sampling frame and overall sample design. Among these characteristics are skewed distributions, diversity of variables of interest, and changing population membership.

Because establishment populations are skew, efficient sample designs for their surveys demand the use of list frames that incorporate known or projected measures of size for each element in the population (see Sigman and Monsour chapter). Unfortunately, when the variables of interest are diverse and weakly correlated, a single list frame with one measure of size will not suffice. Moreover, the changing nature of the population causes any list frame or combination of list frames to become outdated quickly and therefore incomplete in terms of coverage of the population.

One way to assure the completeness of the frame is to use an area frame covering the entire population of interest so that all the elements in the population and all potential future elements will be located somewhere within the area frame. Whereas a list frame is a list of the elements in the population, an area frame is a collection of geographical units or area segments. In list frame designs for establishment surveys, the sampling units are usually the

establishments themselves, and stratified single stage sample designs are commonly used to select establishments directly from the list. In area frame designs, area segments are the sampling units, and these units are often selected using stratified multistage designs. Correspondence rules are needed to link the establishment in the population to the area segments in the frame in a unique manner.

Although area frames ensure complete coverage, they do not generally lead to efficient sampling designs because area segments are essentially clusters of elements. Segment sizes (in terms of numbers of constituent elements) are usually unequal and unknown at the time of the sample design. In fact, there is no guarantee that an area segment contains any establishments at all. Large area sample sizes will be needed to overcome problems with populations containing rare items (variables) and with skewed distributions.

Area frames are most useful for general purpose surveys covering a wide spectrum of items that are fairly evenly distributed geographically or when the sizes (as defined above) of the area segments are available or can be estimated reasonably well. That is why they are used so often in demographic surveys where population censuses provide adequate measures of the size of an area segment. It should also be noted that an area frame has a long life span. It only needs to be updated when geographic features have changed to the point that it becomes difficult to associate population elements with sampled area segments or when the availability of more up-to-date information allows for the development of improved sampling designs. This occurs often in agricultural surveys with the use of recent aerial photographs and/or satellite data.

Area frame establishment surveys are generally more expensive than list frame surveys of comparable sample size. This is because area frames are usually much more costly to develop. In addition, sampled establishments from an area frame often have to be personally enumerated, while sampled elements from a list frame can be enumerated by telephone or electronically.

A *multiple frame* survey uses a combination of frames. The primary reason for using multiple frame sampling for establishment surveys is to utilize the strengths of one frame to offset the weaknesses of the other. In principle, the theory of multiple frame sampling can be applied to the use of more than one list frame (see, for example, Bankier 1986). The main focus here, however, will be on combining the use of list and area sampling frames. Area frame sampling assures completeness, while list frame sampling can be designed efficiently for large and rare items.

Section 2 outlines the theory of sampling from multiple frames. Section 3 discusses the most common use of multiple frame sampling in which there is a single list and a single area frame. Section 4 addresses subsampling from an area frame. Section 5 reviews some practical problems with overlapping frames. Section 6 briefly discusses future directions in multiple frame methodology.

2 FUNDAMENTALS OF MULTIPLE FRAME SAMPLING

This section develops some theory for sampling and estimation from multiple frames. Each frame consists of a set of mutually exclusive primary sampling units, and there exists a one-to-one or many-to-one mapping from the elements in the population of interest to the primary sampling units in a particular frame. The mapping need not be complete for each frame (i.e., some population elements may not be associated with any sampling unit for a particular frame). For simplicity, we will speak of an element belonging to a frame when it maps onto a sampling unit in the frame (sometimes these elements are referred to as the *ultimate* sampling units of the sampling frame). After one or more stages or phases of random sampling, a sample of elements is drawn independently from each frame.

Two important assumptions are made in this section. They are:

Completeness - Every element in the population of interest must belong to at least one frame; and

Identifiability - It should be possible to determine whether or not a sampled element from one frame belongs to any other frame and hence could also have been sampled from it.

The completeness assumption is satisfied whenever an area sampling frame covering the entire population of interest is one of the multiple frames. The identifiability assumption, while simple in theory, poses most of the operational difficulties in implementing a multiple frame survey. This is because it is not always a trivial matter to ascertain whether an element sampled from one frame is also contained in another frame (see Section 5).

The basic theory of multiple frame sampling was developed by Hartley (1962) and extended by Cochran (1965). Hartley divided the population into mutually exclusive domains defined by the sampling frames and their intersections. For example, if there are two sampling frames, A and B, there are three possible domains:

Domain (a) containing elements belonging only to Frame A;

Domain (b) containing elements belonging only to Frame B; and

Domain (ab) containing elements belonging to both Frames A and B.

With k frames, there will be $2^k - 1$ domains (recall the completeness assumption precludes the existence of a domain without any members from at least one frame).

Let us focus attention on the two frame example introduced above to clarify some of the issues involved in estimation using multiple frames. Suppose we are interested in estimating a population total T. It is possible to decompose T as

$$T = T_a + T_b + T_{ab}, \tag{1}$$

where T_d is the population total in domain d ($d = a$, b, or ab). Attention in this chapter will be principally focused on estimating population totals. Extensions to other population parameters are briefly discussed in 3.3.

When $d = a$ or b, one can estimate T_d with the Horvitz-Thompson expansion estimator,

$$t_d = \sum_{i \in S_d} e_i y_i, \tag{2}$$

where S_d is the set of sampled elements in domain d, e_i is the expansion factor (the inverse of the selection probability), and y_i is the item value of interest for element i. This estimator is unbiased under the sampling design.

A continuum of unbiased estimators for T_{ab} is given by

$$\begin{aligned} t_{ab(p)} &= p \sum_{i \in S_{ab}^A} e_i y_i + (1-p) \sum_{i \in S_{ab}^B} e_i y_i \\ &= p t_{ab}^A + (1-p) t_{ab}^B, \end{aligned} \tag{3}$$

where $0 \le p \le 1$, S_{ab}^G is the set of elements in domain ab sampled from frame G ($G = A$ or B), and t_{ab}^G is the Horvitz-Thompson estimator for T_{ab} based on the frame G sample. The limits on p assure that $t_{ab(p)}$ will be non-negative whenever the y_i are non-negative.

We now have a continuum of unbiased estimators for T:

$$t_{(p)} = t_a + t_b + t_{ab(p)}, \tag{4}$$

where $0 \le p \le 1$. The sampling design is independent across frames but not necessarily across domains. Consequently, The variance of t_p is

$$\begin{aligned} \mathrm{Var}(t_{(p)}) &= \mathrm{Var}(t_a) + \mathrm{Var}(t_b) + 2p\mathrm{Cov}(t_a, t_{ab}^A) \\ &\quad + 2(1-p)\mathrm{Cov}(t_b, t_{ab}^B) \\ &\quad + p^2 \mathrm{Var}(t_{ab}^A) + (1-p)^2 \mathrm{Var}(t_{ab}^B). \end{aligned} \tag{5}$$

It is reasonable to choose p so that the variance of $t_{(p)}$ is minimized. The *optimal* (i.e., variance minimizing) p is

$$p = \frac{\mathrm{Var}(t_{ab}^{B}) - \mathrm{Cov}(t_a, t_{ab}^{A}) + \mathrm{Cov}(t_b, t_{ab}^{B})}{\mathrm{Var}(t_{ab}^{B}) + \mathrm{Var}(t_{ab}^{A})} . \qquad (6)$$

If the two co-variance terms in (6) were zero (or equal to each other), p would take on the form:

$$p = \frac{\mathrm{Var}(t_{ab}^{B})}{\mathrm{Var}(t_{ab}^{B}) + \mathrm{Var}(t_{ab}^{A})} , \qquad (7)$$

which always lies between zero and one. As one would expect, the size of p is directly related to the precision of t_{ab}^{A} relative to that of t_{ab}^{B}. The more relatively precise the Frame A sample is in estimating domain ab, the more weight t_{ab}^{A} is given in the estimation of T_{ab}. The expansion factors, e_i, in the above expressions can be either unconditional or conditional. That is to say, they may reflect either the original probabilities of selection or the recomputed selection probabilities within the domains. For example, if the sample design in Frame A were simple random sampling without replacement, then the unconditional expansion factor for every sampled element from the frame would be N_A/n_A, where N_A and n_A are the respective sizes of the population and sample in Frame A, while the conditional expansion factors for the sampled elements in the intersection of domain d and Frame A would be $N_{A(d)}/n_{A(d)}$, where $N_{A(d)}$ and $n_{A(d)}$ are the respective sizes of the population and sample in this intersection.

There are theoretical reasons for preferring estimation with conditional rather than unconditional expansion factors (see Rao 1985). In many applications, however, especially those involving an area frame (where population sizes are often unknown), conditional selection probabilities will be either impossible to calculate or impractical. Consequently, unconditional inference will have to suffice.

Although one can, *in principle*, choose p so that the variance (conditional or unconditional) is minimized, this is usually impossible to do in practice because the component variance and co-variance terms in equation (5) are unknown. They could be estimated from the sample, but then the choice of p would not really minimize the variance of $t_{(p)}$ but the *estimated* variance of $t_{(p)}$. As a consequence, this estimated variance would be biased downward.

Even if the distinction between the variance-minimizing p and the estimated variance-minimizing p could be ignored, the following inconvenience remains: the optimal p can vary from survey item to survey item. This point as demonstrated by Armstrong (1979) with Canadian farm data.

A popular alternative is to eschew issues of optimality or near optimality and fix the

value of p in advance, most commonly at zero. One can then estimate the components on the right hand side of equation (5) in an unbiased fashion and create an unbiased estimator for the variance of $t_{(p)}$. This approach would be biased if p were estimated from current data instead of being fixed, because the estimated variance of $t_{(p)}$ is no longer a linear combination of unbiased estimators.

It is possible to develop a more direct Horvitz-Thompson estimator for T in equation (1) by treating the samples from the two frames as a single sample and then computing the probability of selection for each sampled element (the sum of its probability of selection from each frame minus the product of these two terms). There is no reason to believe, however, that the resultant estimator will have less variance than $t_{(p)}$ when a near optimal p is used. Moreover, estimating the variance of this alternative will often be quite difficult.

Finally, one can sometimes improve on $t_{(p)}$ by using auxiliary information. The interested reader is directed to Fuller and Burnmeister (1972), Bosecker and Ford (1976), Bankier (1986), Skinner (1991), and Rao and Skinner (1993).

3 THE DOMINANT SPECIAL CASE: ONE LIST FRAME, ONE AREA FRAME

The U.S. Census Bureau's 1949 Sample Survey of Retail Stores (Hansen et al. 1953) was one of the first establishment surveys that used a multiple frame technique. It employed a single list frame and a single area frame. This two frame approach is still used in the Bureau's Monthly Retail Trade Survey. Nevertheless, it is fair to say that the approach currently finds its widest use in surveys of agriculture. See Vogel (1975) and Julien and Maranda (1990) for respective discussions of U.S. and Canadian multiple frame agricultural surveys.

3.1 Area Sampling

The subject of list sampling for establishment surveys has been reviewed in some detail in the Sigman and Monsour chapter. Area sampling concepts for establishment surveys require some additional discussion.

The basic concepts of area frame sampling are simple. The total area to be surveyed is first stratified by geography or other known characteristics that are related to the variables of interest (population density is one such characteristic for a retail survey, farm density is one for an agriculture survey). A sample of segments -- usually compact blocks of land -- is then selected within each stratum, perhaps using a multistage sampling design.

Gallego and Delincé (1993) discuss sampling designs where points are randomly selected within sampled area segments. In this section, however, we focus on designs for which the last stage of sampling is the selections of area segments. When employing such designs, rules need to be developed to uniquely link (map) the elements in the population with the sampled segments. The area expansion factor for each element is simply the inverse of the selection probability of the segment with which it is linked.

Although the basic concepts are simple, the successful application of area frame sampling can become very complex, especially if used in a multiple frame environment. The theory and application of area frame sampling are well documented and include Jessen (1942), King and Jessen (1945), Houseman (1975), Nealon and Cotter (1987), and Bush and House (1993). Fecso et al. (1986) contains a good list of references covering area frame sample design issues.

The link between population elements and sampled segments is complicated in those area frame surveys (e.g., agricultural surveys) where the physical location of establishments (farms) may cross segment boundaries. There are three approaches to redefining the elements of an area frame frequently used in this situation.

In the *closed* (or tract) approach, an element is defined as the intersection of an establishment and an area segment (i.e., the portion of the establishment that lies within the area segment). Response errors occur when the reporting unit, usually the establishment, is not able to measure and report on that portion of its business that lies within the selected segment's boundaries.

The closed approach is statistically robust and relatively efficient for items that are generally distributed evenly over wide areas, such as crop acreages. Outliers can be controlled by the size of the segment for many of these items. For example, if the segment size is 600 acres, the maximum acres for any crop in the segment is 600 acres even though a single establishment located in the segment may account for many times that number of crop acres in its entire operation.

In the *open* approach, an element is defined to be an establishment and is linked only with that segment containing its headquarters. This can lead to difficulties with large establishments with complicated corporate or partnership structures. For example, many medical practices involve several doctors with offices in more than one location and identifying the "headquarters" may not be simple. Complex counting rules must be devised that ensure such establishments do not have a chance of being linked with more than a single area segment.

A sampled segment will likely contain fewer elements with larger item values when using the open approach to defining elements as opposed to a closed approach. As a result, the former is often less statistically efficient than the alternative. A potential advantage of the open approach, however, is that once the establishment has been defined, it may be easier for the respondent to report values for the entire establishment rather than for that portion of the establishment located within an area segment. Segment boundaries will have little intrinsic meaning to many respondents.

In theory, the definition of the element in the *weighted* approach is the same as in the closed approach: the portion of an establishment that lies within an area segment. The difference is in the data values attributed to the element. Data values for an establishment that is the "parent" of several elements (each in a different area segment) are prorated using fixed weights defined so that they sum to unity. For agricultural surveys, it is common to

use the fraction of the establishment's (farm's) land within a sampled segment as the weight for the element in that segment.

The weighted approach leads to less sampling variability than the open approach because large operations are spread across many segments. One major problem with this approach, especially when it is applied to agricultural surveys, is that the weights themselves are prone to measurement errors (see Nealon 1984, p. 19). In principle, the weights for all the elements with the same parent establishment should sum to unity. In practice, since only one of these elements is likely to be selected in an area sample, pains must be taken to assure that the weight used to prorate establishment values to sampled elements are not systematically larger or smaller than they should be.

Nealon (1984) provides a theoretical and empirical examination of these three approaches for area frame and multiple frame surveys conducted by the U.S's National Agricultural Statistics Service (NASS). The results favor the weighted approach. Houseman (1975) contains a more in depth discussion of the three approaches. Faulkenberry and Garoui (1991) explore additional approaches.

3.2 Estimation

Suppose two frames are to be used, one list and one area frame, and a total is to be estimated. Using the notation of Section 2, let A denote the area frame and B the list frame. Observe that domain b is empty because all elements should be located somewhere on the area frame. Consequently, $T_b = t_b = 0$.

In almost all applications of the two frame design, p in equations (3) and (4) is set equal to zero, rendering $t = t_a + t_{ab}^B$. This value for p, besides being convenient for variance estimation purposes, is often close to optimal (depending on the efficiency of the list frame stratification) because the variance of the list frame estimator of domain ab (i.e., t_{ab}^B) is almost always considerably smaller than the variance of the area frame estimator of this domain (t_{ab}^A) (recall Section 1). Consequently, equation (7) suggests that p should be very small. Depending on the content of the list frame, $\mathrm{Cov}(t_a, t_{ab}^A)$ in equation (6) may actually be positive because t_a and t_{ab}^A are based on the same sampled segments. This would make the optimal p even smaller (note: $\mathrm{Cov}(t_b, t_{ab}^B)$ in (6) must be zero because t_b is always zero).

A little renaming to simplify the notation: let $t_L = t_{ab(0)} = t_{ab}^B$, since it is the Horvitz-Thompson expansion from the list frame of domain ab (the overlap domain); and let $t_N = t_a$, since it is the area expansion for the nonoverlap domain (using either the closed, open, or weighted approach to element definition).

The estimator for the total, $t = t_L + t_N$, is called the *screening multiple frame estimator* because elements in the area sample are "screened" and only those in the nonoverlap domain are enumerated. Its variance has the obvious form:

$$\mathrm{Var}(t) = \mathrm{Var}(t_L) + \mathrm{Var}(t_N). \qquad (8)$$

Estimating Var(t_L) is straightforward when, as is usually the case, t_L is based on a single stage stratified list sample. Estimating Var(t_N), however, can be a bit more complicated. Recall that a sample of segments was selected from the area frame. In many cases the design used is roughly equivalent to stratified simple random sampling (see Kott 1989). In practice, sampling fractions are so low within strata that the distinction between with and without replacement sampling can be ignored. Suppose there are H strata and n_h sampled segments within stratum h. If X_{hj} is the sum of the expanded values of all nonoverlap elements in segment j of stratum h (i.e., $e_h \Sigma y_i$, where e_h is the expansion factor for all sampled elements in the stratum h, y_i is the value of element i, and the summation is over all nonoverlap elements in segment j), then

$$\text{var}(t_N) = \sum_{h=1}^{H} n_h/(n_h - 1) \{ \sum_{j=1}^{n_h} X_{hj}^2 - [\sum_{j=1}^{n_h} X_{hj}]^2/n_h \} \qquad (9)$$

is an unbiased estimator for Var(t_N) whenever finite population correction can be ignored. Although equation (9) appears straightforward, one needs to be aware that X_{hj} will be zero for many sampled segments either because there are no elements linked with the segment or because the elements linked with the segment are also on the list frame.

3.3 Estimating Means, Ratios, and Regression Coefficients

The area sample is composed of elements, while the population itself is composed of establishments. Although this difference poses no problems when the goal is estimating population totals, it can when estimating population ratios -- which include population means -- and regression coefficients.

Estimating population ratios and regression coefficients with a multiple frame sample is a relatively trivial matter when the open approach to element definition in the area frame is used (and, as in the last subsection, $p = 0$). When an establishment selected from the area frame is in the overlap domain, the establishment's area frame data values are ignored in both the numerator and denominator of a ratio estimator. An analogous technique is used for estimating population regression coefficients. As in equation (9), however, estimates of variance need to reflect the existence of those sampled segments that do not contribute to the estimation of the parameter itself.

When the weighted or closed approach is used, estimating ratios is again straightforward since either approach can be used to estimate the appropriate total in both the numerator and the denominator. It is unclear, however, how one can estimate population regression coefficients with the closed approach. The weighted approach is discussed below.

A finite population regression coefficient has the form:

$$\mathbf{B} = \left(\sum_{k \in P} \mathbf{x}_k' \mathbf{x}_k \right)^{-1} \sum_{k \in P} \mathbf{x}_k' y_k , \qquad (10)$$

where

k denotes an establishment,
P denotes the population of establishments,
y_k is a value attached to establishment k, and
$\mathbf{x}_k$ is a row vector of values attached to k.

Without loss of generality, we will assume that the elements of the list frame sample are, in fact, establishments. In what follows, an establishment is said to be "associated" with a sampled element when it is sampled from the list frame or when it is the parent of a sampled element from the area frame.

A reasonable estimator for **B** that uses the weighted approach to element definition is

$$\mathbf{b} = \left(\sum_{k \in S} e_k w_k d_k \mathbf{x}_k' \mathbf{x}_k \right)^{-1} \sum_{k \in S} e_k w_k d_k \mathbf{x}_k' y_k , \qquad (11)$$

where

S is set of establishments associated with sampled elements,
e_k is the expansion factor for the element associated with k,
w_k is the weight attributed to the element associated with k if is sampled from the area frame, one otherwise, and
d_k is 0 if k is associated with an element sampled from the area frame that is also on the list frame, one otherwise.

Two observations about equation (11) are in order. One, the set S may, in principle, include particular establishments more than once due either to list/area overlap or to a particular establishment being in two sampled area segments. The d_k and w_k terms assure us that neither the y-values nor the **x**-values for such establishments will be double-counted.

Two, if we call $g_k = e_k w_k d_k$ the adjusted expansion factor, then **b** can be expressed as $\mathbf{b} = (\Sigma_{k \in S} g_k \mathbf{x}_k' \mathbf{x}_k)^{-1} \Sigma_{k \in S} g_k \mathbf{x}_k' y_k$. In this form, one can use a survey analysis software package like SUDAAN (Shah et al. 1991) to calculate **b** and estimate its variance.

4 SUBSAMPLING FROM AN AREA FRAME

In this section, we explore two methods of subsampling from an area frame sample. What distinguishes these methods from simple multistage area sampling is that each incorporates the development of a list frame for subsampling.

In the method described in 4.1, one starts with an area sample of primary sampling units (PSU) and then uses a multiple frame estimation strategy (list and area) within each PSU. In the second method described in 4.2, one starts with an area sample of elements using the weighted method of element definition and then draws a subsample of these elements. Since it is not required that subsampling be done independently within the original sampling units, this is a *two-phase* sampling design (see Särndal et al. 1991, Chapter 9).

4.1 Using Multiple Frame Designs Within Area PSU's

There are many situations where the cost of developing a list frame is prohibitively high. On the other hand, the variance of an estimator based on a pure area frame sample is prone to be unacceptably large.

This was the situation faced by the designers of the U.S. Commercial Buildings Energy Consumption Survey (CBECS) which measures the consumption of energy in commercial buildings (Energy Information Administration 1989). Their solution was to use a multi-stage area sample and to create list frames of large commercial buildings within sampled PSU's. This approach reduced the potential size of the expansion factors for such buildings.

Let j denote a PSU, and T_j the first stage expanded values of interest from all elements in the PSU. A list and area multiple frame estimation strategy can be constructed to estimate each T_j. Since the elements in the CBECS were buildings that never crossed segment boundaries, the linking of elements to the area segments was straightforward.

The estimator developed for the CBECS survey did not estimate each T_j with a member of the family in equation (4). Instead, the CBECS estimator has a form equivalent to

$$t_j' = t_j + e_{1j}\left(\sum_{k \epsilon P_j'} y_k - \sum_{k \epsilon S_j'} e_{Lk} y_k \right), \qquad (12)$$

where
t_j is the screening estimator ($t_{jL} + t_{jN}$) for T_j,
e_{1j} is the first stage (PSU) expansion factor,
P_j' is the set of all elements subsampled from the area frame that are also on the list frame in PSU j,
S_j' is the set of all elements subsampled on both the list and area frames of PSU j, and
e_{Lk} is the inverse of the probability that element k is subsampled from the list frame in j.
It is not difficult to show that t_j' is unbiased (Chu 1987).
Let $d_j = t_j' - t_j$. The variance of t_j' conditioned on the first stage selection of PSU j is

$$\mathrm{Var}_2(t_j') = \mathrm{Var}_2(t_j) + \mathrm{Var}_2(d_j) + 2\,\mathrm{Cov}_2(t_j, d_j), \qquad (13)$$

where the subscript 2 denotes the conditional nature of the variance (co-variance) operator. Since the co-variance term in equation (13) is likely to be negative, it is possible for t_j' to have less variance that t_j.

Now consider unbiased estimators having the general form:

$$t_{jc} = t_j + cd_j ,$$

where c is a constant. Observe that $t_{jc} = t_j$ when $c = 0$, and $t_{jc} = t_j'$ when $c = 1$. The value of c may be chosen so that the variance of t_{jc} is minimized, although that may not be easy to do in practice.

Suppose there are H strata in the original area frame and n_h sampled PSU's within stratum h. A variance estimator for $t_c = \Sigma_h \Sigma_j t_{jc}$ is

$$\text{var}(t_c) = \sum_{h=1}^{H} n_h/(n_h - 1) \{ \sum_{j=1}^{n_h} t_{jc}^2 - [\sum_{h=1}^{n_h} t_{jc}]^2/n_h \}. \qquad (14)$$

This estimator is unbiased when the first stage sample is drawn with replacement and is nearly unbiased in many practical applications when PSU's are selected without replacement. Goldberg and Gargiullo (1988) discuss variance estimation for surveys (like the CBECS) in which some PSU's are selected with certainty.

4.2 Using Lists Across Area Segments

Suppose one has previously drawn a sample of segments from an area frame for a particular multiple frame (list and area) survey and now wishes to mount a new, smaller survey to estimate different population parameters. This happens frequently in agricultural surveys where different parameters are of interest at different times of the year and particular parameters may change over time. Suppose further that the same list frame will be employed for the new survey so that the nonoverlap domain is effectively fixed. Although this supposition is not necessary, it greatly simplifies the exposition.

In order to reduce the sample size for the new survey, the elements of the area frame sample can be subsampled. Since list sampling is generally more efficient than area sampling, it makes sense to use list sampling techniques to draw such a subsample.

Elements sampled from the area frame can be separated into two groups: those that overlap with the list frame and those that do not. As discussed in 3.2, the values of overlap elements are ignored when estimating T with the screening estimator, $t = t_L + t_N$. This means that these elements are attributed values of 0. Since we know their values in advance of subsampling, we can think of any subsample as containing all originally sampled overlap elements (these elements do, however, need to be in the original sample so that we can determine whether or not they are in the overlap domain). The only question then is how to subsample from the sampled nonoverlap elements.

Let e_{1i} be the original expansion factor of element i from the area sample, w_i be the weight attributed to i in the weighted approach to element definition. In the open approach,

w_i is unity. Let y_i be the value of interest for the establishment associated with i. Define $z_i = e_{1i}w_iy_i$. The goal of subsampling from the area frame sample is to estimate

$$t_N' = \sum_{i \epsilon N_A} z_i , \qquad (15)$$

where N_A is the set of nonoverlap elements in the original area sample. In the open approach, N_A is restricted to elements with headquarters in the original area sample.

List sampling designs to estimates parameters like t_N' in equation (15) in an unbiased and efficient manner are the subject of the Sigman and Monsour chapter. There is no requirement that the original area sampling design be reflected at all in the subsampling design used to estimate t_N'. At the U.S. Department of Agriculture, for example, area sampled nonoverlap elements from the large June Agriculture Survey are restratified for "follow-on" surveys in subsequent months. Those elements believed to have similar z_i values are placed in the same subsampling stratum irrespective of their original area stratum or their original expansion factor (except so far as e_{1i} impacts on z_i), and a without replacement simple random sample is drawn within each new stratum.

An attractive alternative to stratified simple random sampling is probability proportional to size (pps) sampling, which can also incorporate stratification. Since there are likely to be a number or values of interest in the new survey and because y_i values are unknown before enumeration anyway, one may choose to treat the $e_{1i}w_i$ as the measures of size for pps sampling. Alternatively, data from the original survey can be used to determine reasonable measures of size.

A Horvitz-Thompson expansion estimator for t_N', and thus T_N, is

$$t_N = \sum_{i \epsilon N_2} e_{2i}z_i \qquad (16)$$

$$= \sum g_i z_i ,$$

where
e_{2i} is the inverse of element i's probability of being subsampled given it is in the original area sample,
N_2 is the set of elements subsampled from the nonoverlap domain of the area frame, and
$g_i = e_{2i}e_{1i}w_i$.
It is interesting to note than when unstratified probability proportional to $e_{1i}w_i$ sampling is used in the subsampling, g_i is identical for all subsampled elements.

The variance of the two phase estimator described above is the sum of the variance of t_N' in equation (15) as an estimator of T_N (which was discussed last section) plus the expected value of the variance of t_N in equation (17) as an estimator for t_N' (Cochran 1978). An *estimator* for this variance is more difficult to express. Särndal and Swensson (1987) provide a general formulation.

Kott (1990) explores the important special case where the first phase is a stratified simple random sample of area segments and the second phase is a (re)stratified simple random sampling of elements. It turns out that if x_{hj} is the sum of the fully expanded values of all nonoverlap elements in area segment j of stratum h and there are n_h sampled segments in stratum h ($h = 1, \ldots, H$), then

$$\text{var}(t_N) = \sum_{h=1}^{H} n_h/(n_h - 1) \{ \sum_{j=1}^{n_h} x_{hj}^2 - [\sum_{j=1}^{n_h} x_{hj}]^2/n_h \}, \qquad (17)$$

is biased upward as an estimator for the variance of t_N.

It is well known that equation (17) is unbiased under two-stage sampling when the first stage sampling fractions can be ignored. This is because the x_{hj} are (virtually) independent. In the design discussed in Kott (1990), the x_{hj} tend to be inversely correlated: when one segment has more elements than expected -- and thus a larger x_{hj} value -- another segment would tend to have fewer elements than expected. Such inverse correlations tend to depress the variance of $t = \Sigma^H \Sigma_j x_{hj}$, while increasing the value of the right hand side of equation (17). This informal reasoning suggests that equation (17) is likely to provide a conservative variance estimator for two phase estimation strategies more complicated than the one formally treated in Kott (1990).

5 PROBLEMS IN SURVEYS WITH OVERLAPPING FRAMES

As noted in previous sections, an area frame provides complete coverage of the population, but it can lead to inefficient estimation when the population of interest has a skewed distribution. A list frame, by contrast, can provide the means for more efficient estimation, but is often incomplete.

When a list of establishments is used along with an area frame in a multiple frame environment, the list need not be complete. It should, however, contain those large and rare establishments which are the bane of area frame surveys.

5.1 List Frame Issues

All establishments on a list frame must be completely identified by name, address, etc. Operations that are large, have complex management structures, and/or are scattered over different locations must be so identified. This is the identifiability assumption described in Section 2. In a two frame (list and area) sampling design one must be able to determine whether an establishment (or portion of an establishment) sampled from the area frame could also have been selected from the list frame; otherwise, intractable nonsampling errors will result. It must be realized, however, that the identifiability requirement can greatly increase the cost of list development.

When developing a list frame of establishments, more than one source of

establishment names are often available. The survey designer must decide whether to use each list as a separate frame in a multiple frame survey or to combine the lists prior to sampling and create a single list frame for sampling purposes.

Suppose the former approach is chosen with two list frames and a single area frame. The population can then be divided into four mutually exclusive domains: (1) establishment on neither list frame, (2) establishments on the first list frame but not the second, (3) establishments on the second list frame but not the first, and (4) establishments on both list frames.

One can see that the need to identify all domains when there are two or more list frames greatly complicates the survey and estimation process. Many statistical agencies have therefore decided that it is usually more practical to combine all lists and remove duplication prior to sampling. This can still be a significant undertaking requiring the use of record linkage methodologies that may be prone to errors. These methodologies are discussed in the Winkler chapter.

5.2 Overlap Detection Issues

The reliance upon matching names on the list frame to sampled elements from the area frame to determine overlap complicates the survey process. Whether an establishment is surveyed via the list or area frame, it is necessary to determine the primary business name, other business names, and the names of individuals who are also associated with the business. Rules of association of individuals' names with establishments must be defined. For example, an area frame unit may contain the residence of Dr. *X* associated with the medical practice of Drs. *X*, *Y*, and *Z*. Should the association of the establishment with an area frame element be based on the residence of each name or on the location of the business itself? This will depend upon the counting rules used to determine the overlap between the frames. This example becomes more complicated if there is a Dr. *T* involved in the practice, but the list frame only identifies the practice involving Drs. *X*, *Y*, and *Z*. If counting rules are used that would allow Dr. *T* to report for the entire practice of Drs. *X*, *Y*, and *Z*, then it must be possible to identify the practice as overlap with the list frame. This situation occurs often in agriculture where an individual associated with a partnership can report for the entire partnership, but the name matching process fails to properly link it with a name on the list frame, resulting in an upward bias. A more detailed discussion of these problems appear in Vogel (1975). The survey instrument for both frames must be carefully designed to identify and link names with establishments so that the overlap domain can be properly determined.

Resources need to be available to re-interview "questionable links" so that the domain determination is correct. The domain determination has been shown by several studies to be the single largest source of non-sampling errors in multiple frame surveys (Nealon 1984).

5.3 Estimation Difficulties

A two frame (list and area) sampling design will generally yield more efficient and robust estimators than an area frame design by itself. Nevertheless, outliers can still occur

when the list frame is not constructed carefully or is not up to date. An establishment missing from the list frame that is sampled in the nonoverlap domain of the area frame can have a much larger expansion factor then it would have had as a list sample (since the costs associated with an area frame are larger, its sampling fractions are usually smaller). Five such establishments in a 1992 U.S. Department of Agriculture survey accounted for six percent of the national estimate.

Although the desire to eliminate potential outliers provides a powerful argument for including as many establishment names as possible on the list frame, there is an equally compelling reason for limiting the list frame to only larger establishments. The smaller the list, the easier it is to check overlap. Moreover, the incremental costs of adding enumerated elements to the nonoverlap domain are relatively small since, a, the fixed cost of developing an area frame has already been expended, and, b, each element linked to a sampled area segment must be contacted to determine its overlap status whether or not it eventually needs to be enumerated.

The use of a multiple frame design can make it more difficult to measure change over time. Ratio estimators are usually an efficient means of estimating change when a portion of the sample in the reference (denominator) time period remains in the sample in the comparison (numerator) period. Since establishments may move between the overlap and nonoverlap domains as their structures change or as the list frame is updated, the efficiency of ratio estimators of change in multiple frame surveys is reduced. Extreme caution must also be exercised in multiple frame designs to ensure that establishments found through the area frame sample are not added to the list frame during the duration of the design since that can bias estimation (by effectively changing selection probabilities). Such additions to the list frame should only be allowed when an entirely new sample is selected from both frames. The independence of sample selection between the area and list frames must be maintained for the area frame to properly estimate for the incompleteness of the list frame (see Vogel and Rockwell 1977).

6 THE FUTURE OF MULTIPLE FRAME SURVEYS

There are two opposite directions in which multiple frame surveys are heading. The first is the elimination of the area frame altogether. Because of the cost of maintaining an area frame, Canada models the incompleteness of its list frame of retail and wholesale establishments rather than measure list undercoverage with an area sample (Sande 1986). The U.S. Census Bureau may soon follow suit and abandon the area sample component of its Monthly Retail Trade Survey (Konschnik et al. 1991).

This pattern could be reversed in Italy, however, where lists of business establishments are viewed as unreliable. Petrucci and Pratesi (1993) discuss the possibility of introducing multiple frame business surveys in that country.

The future use of area frames both by themselves and in a multiple frame environment is more secure for surveys of agriculture and construction. Here, we find the

other direction in which multiple frame surveys are heading: more complicated estimators. NASS is currently experimenting with regression estimation for the second phase of area sampling and with outlier resistant estimators. One intriguing idea is to employ time series techniques (see, for example, Scott et al. 1977) to dampen the effects of outliers from the area frame.

Chapman (1993) discusses using area cluster sampling within a list frame to reduce the cost of a survey conducted by personal interview. The Garrett and Harter chapter reviews an innovative use of area frame methods to update list samples in a repeated survey. A network sampling design linking workers and employers that starts with an area frame sample is discussed in the A. Johnson chapter.

REFERENCES

Armstrong, B. (1978), "Test of Multiple Frame Sampling Techniques for Agricultural Surveys: New Brunswick," *Survey Methodology*, **5**, pp. 178-184.

Bankier, M.D. (1986), "Estimators Based on Several Stratified Samples With Applications To Multiple Frame Surveys," *Journal of the American Statistical Association*, **81**, pp. 1074-1079.

Bosecker, R.R. and Ford, B.L (1976), "Multiple Frame Estimation With Stratified Overlap Domain," *Proceedings of the Social Statistics Section*, American Statistical Association, 1976, Part I, pp. 219-224.

Bush, J. and House. C. (1993), "The Area Frame: A Sampling Base for Establishment Surveys," *Proceedings of the International Conference on Establishment Surveys*, American Statistical Association, pp. ?-?.

Chapman, D.W. (1993), "Cluster Sampling for Personal-Visit Establishment Surveys," *Proceedings of the International Conference on Establishment Surveys*, American Statistical Association, pp. ?-?.

Chu, A. (1987), "A Proof that the Assignment of Conditional Weights Will Produce Unbiased Estimates" in *Weighting Procedures for CBECS III*, unpublished report, Rockville, Maryland: Westat.

Cochran J.G. (1977), *Sampling Techniques*, New York: Wiley.

Cochran, R.S. (1965), *Theory and Applications of Multiple Frames Surveys*, unpublished Ph.D. dissertation, Ames, Iowa: Iowa State University.

Energy Information Administration (1989), *Commercial Buildings Characteristics 1989*, Washington, DC: Energy Information Administration.

Faulkenberry, G.D. and Garoui, A. (1991), "Estimation a Population Total Using an Area Frame, *Journal of the American Statistical Association*, **86**, pp. 445-449.

Fecso, R., Tortora R.D., and Vogel, F.A. (1986), "Sampling Frames for Agriculture in the United States, *Journal of Official Statistics*, **2**, pp. 279-292.

Fuller, W.A. and Burnmeister, L.F. (1972), "Estimates for Samples Selected From Two Overlapping Frames," *Proceedings of the Social Statistics Section*, American Statistical Association, pp. 245-249.

Gallego, F.J. and Delincé, J. (1993), "Sampling Frames Using Area Samples in Europe," *Proceedings of the International Conference on Establishment Survey*, American Statistical Association, pp. ?-?.

Goldberg, M.L. and Gargiullo, P.M. (1988), "Variance Estimation Using Pseudostrata for a List-Supplemented Area Probability Sample, *Proceedings of the Survey Research Section*, American Statistical Association, pp. 479-484.

Hansen, M.H., Hurwitz, W.N., and Madow, W.A. (1953), *Sample Survey Methods and Theory*, Volume 1, pp. 516-556, New York: Wiley.

Hartley, H.O. (1962), "Multiple frame surveys," *Proceedings of the Social Statistics Section*, American Statistical Association, pp. 203-206.

Hartley, H.O. (1974), "Multiple frame methodology and selected applications," *Sankhyá*, Ser. C, **36**, pp. 99-118.

Houseman, E (1975), *Area Frame Sampling in Agriculture*, SRS Report No. 20, Washington, DC: Statistical Reporting Service, U.S.D.A.

Jessen R.J. (1942), *Statistical Investigation of a Sample Survey for Obtaining Farm Facts*, Research Bulletin 304, Ames, Iowa: Iowa State College of Agriculture and Mechanical Arts.

Julien, C. and Maranda, F. (1990), "Sample Design of the 1988 National Farm Survey," *Survey Methodology*, **16**, pp. 116-129.

King, A.J. and Jessen , R.J. (1945), "The Master Sample of Agriculture, *Journal of the American Statistical Association*, **40**, pp. 38-56.

Konschnik, C.A., King, C.S., Dahl, S.A. (1991), "Reassessment of the Use of an Area Sample for the Monthly Retail Trade Survey," *Proceedings of the Survey Research Section*, American Statistical Association, pp. 208-213.

Kott, P.S. (1989), "The Variance of Direct Expansions from a Common Area Sampling Design," *Journal of Official Statistics*, **5**, pp. 183-186.

Kott, P.S. (1990), "Variance Estimation when a First Phase Area Sample is Restratified," *Survey Methodology*, **16**, pp. 99-104.

Lund, R.E. (1968), "Estimators in Multiple Frame Surveys," *Proceedings of the Social Statistics Section*, American Statistical Association, pp. 282-288.

Nealon, J.P. (1984), *Review of the Multiple and Area Frame Estimators*, SF&SRB Staff Report 80, Washington, DC: Statistical Reporting Service, U.S.D.A.

Nealon. J.P., And Cotter. J. (1987), *Area Frame Design for Agricultural Surveys*, Washington, DC: National Agricultural Statistics Service.

Petrucci, A. and Pratesi, M. "Listing Frames and Maps in Area Sampling Surveys on Establishments and Firms (1993)," *Proceedings of the International Conference on*

Establishment Survey, American Statistical Association, pp. ?-?.

Rao, J.N.K. (1985), "Conditional Inference in Survey Sampling," *Survey Methodology*, Vol. **11**, pp. 15-31.

Rao, J.N.K. and Skinner, C.J. (1993), "Estimation in Dual Frame Surveys With Complex Designs," unpublished paper.

Sande, I. (1986), "Business Survey Redesign Project: Frame Options for the Monthly Wholesale Retail Trade Survey Redesign," unpublished memorandum, Ottawa, Canada: Statistics Canada.

Särndal, C.E. and Swensson, B. (1987), "A General View of Estimation for Two Phases of Selection with Application to Two-phase Sampling and Nonresponse," *International Statistical Review*, **55**, 1987, pp. 279-294.

Särndal, C.E., Swensson, B., and Wretman, J. (1991), *Model Assisted Survey Sampling*, New York: Springer-Verlag.

Scott, A.J., Smith, T.M.F. And Jones, R.G. (1977), "The Application of Time Series Methods to the Analysis of Repeated Surveys," *International Statistical Review*, **45**, 13-28.

Shah, B.V., Barnwell, B.G., Hunt, P.H., and LaVange, L.M. (1991), *SUDAAN™ User's Manual*, Research Triangle Park, NC: Research Triangle Institute.

Skinner, C.J. (1991), "On the Efficiency of Raking Ratio Estimation for Multiple Frame Surveys," *Journal of the American Statistical Association*, **86**, pp. 779-784.

Vogel, F.A. (1973), "An Application of a Two-stage Multiple Frame Sampling Design," *Proceedings of the Social Statistics Section*, American Statistical Association, pp. 617-622.

Vogel, F.A. (1975), "Surveys with Overlapping Frames - Problems in Application," *Proceedings of the Social Statistics Section*, American Statistical Association, pp. 694-699.

Vogel, F.A., and Rockwell, D. (1977), *Fiddling with Area Frame Information in List Development and Maintenance*, SF Staff Report 77-01, Washington, DC: Statistical Reporting Service, U.S.D.A.

EDITING SYSTEMS AND SOFTWARE

Mark Pierzchala
National Agricultural Statistics Service

INTRODUCTION

Software development progresses so quickly that it defies any attempt to track it in a book such as this one. There is not a large enough market place for data editing software systems that they would be discussed in the monthly or weekly trade press. The best way to learn about a particular system is to obtain a copy of the software and test it, attend presentations, visit the development site, and question current users. Systems do not perform all the same functions. This chapter describes a few general approaches taken, why they are taken, how to evaluate available software, and the results of some documented implementations. The information presented here can also be used as a starting point for development of software if no suitable systems are available. Mention of a system in this chapter does not infer recommendation nor that it is generally available. Also there are systems in use that are not mentioned.

1 HISTORIC AND FUTURE DEVELOPMENT OF EDITING SYSTEMS

Almost all editing and imputation techniques and strategies are manifested through computer systems. For many establishment surveys, problems with data still must be detected and treated in the post-collection editing stage. The treatment of each form can be very time consuming and often must be done by highly skilled and highly paid subject matter specialists. Based on its (non-probability) sample of United States surveys (of all types), the Federal Committee on Statistical Methodology (FCSM) (1990) found that all error correction is done by analysts or clerks in 60% of the surveys . For some surveys the respondent may be contacted numerous times over long periods of time, increasing respondent and agency burden. In the collecting agency, either many people are assigned to the job, the job is done inadequately, or (more positively) well designed computer tools and editing strategies make the editing process more productive while allowing the agency to achieve or maintain quality estimates.

1.1 Historic Justification

An excellent summary of an early experience with a mainframe batch editing system in the U.S. Bureau of Labor Statistics is given by Stuart (1966). Reasons given for moving from hand-based editing to a computer-based editing system are much the same as those given today: to speed up data review, to relieve the human data editor from much tedium, to do things that people cannot do such as apply the same edit logic consistently to all forms, to automate imputation techniques, to allow more edits to be applied (but see Granquist Chapter ?? about the desirability of this), and to focus valuable staff time on records of high impact or with high probability of error. He further elaborates on kinds of review that the computer can do; for example, checks for internal (micro) consistency, external (macro) consistency (e.g., behavior

of a firm against other firms in a class), and of matching reports. Also mentioned are the possible integration of the edit review with analysis and summary and the need for a general system that would help ease the production of the edit review programs. This short descriptive paper is worth reading because all of the topics discussed are still of major concern today.

The early to mid 1970s saw the advent of Computer Assisted Data Collection Methods. These began with Computer Assisted Telephone Interviewing (CATI) (see Werking and Clayton, Chapter ??). Later, advances in portable computing facilitated the introduction of Computer Assisted Personal Interviewing (CAPI). Recently, the availability of computers in the respondents' offices has allowed the development of electronic self-reporting systems (Heath, Proceedings Chapter ??, Hundepool, Proceedings Chapter ??). New technologies have led to new issues, including which kinds of edits and how many edits to do during an interview as opposed to post-collection review (Eklund, Proceedings Chapter ??); design of computer screen interfaces for interviewers, respondents, and data editors; and how to handle mixed mode data collection (e.g., paper and CATI collection in the same survey) (Woefle, Proceedings Chapter ??).

1.2 Major Trends in Software Development

Current opportunities in development and design of editing systems stem from advances in personal computer hardware, data base technology, modular and generalized system development methods, networking, graphical computer screen displays, and telecommunications. Some trends in editing system development can be discerned by inspection of available systems and by a review of the literature. Most trends listed already appear in two or more systems.

- More edits are invoked in the data collection stage by Computer Assisted Interviewing (CAI) or electronic self-reporting. This reduces callbacks and the need to review data after collection.
- Direct computer-to-computer data transfer from establishment to collecting agency (Electronic Data Interchange or EDI) is being tested. Telephone touchtone collection is operational in a few surveys, as is reception of data by facsimile. Data collection by automated voice recognition is being explored. These technologies will speed up data collection and be more convenient for the respondent, but for most of them edits will not be invoked interactively at the time of collection.
- For post collection data review, if human intervention is required it should be interactive. This speeds up review and allows forms to be cleaned in one treatment, reducing cyclical processing inherent in inappropriate batch methods.
- Batch processing should be reserved for file-based activities that do not require human intervention, for example finding edit failures or making computations.
- Some batch-based generalized edit and imputation (Fellegi and Holt) systems perform tasks previously reserved for human review including automatically finding errors, choosing items for deletion, and effecting changes.
- Top-down (macro or statistical) editing is very useful where relatively few offending reports must be found from a massive file and treated quickly.
- Integration of survey tasks within one system speeds applications development and maintenance, reduces or eliminates some tasks, and facilitates data flow between

tasks. For example, Computer Assisted Data Collection reduces or eliminates the need for data entry.

- *Meta-data* (descriptive information about survey data) are formalized and generated in order to document a survey and also to ease data transfer from one system to another. In the future, some editing systems will write directly (not export) to different file formats so different packages can process the same data set.
- Where data are collected on paper, editing should be done after data entry or perhaps during data entry, but not before data entry. This allows the agency to determine the impact of editing on estimates.
- Systems are designed to plan and manage multi-mode data collection, (e.g. a combination of paper and Computer Assisted data collection) and to optimize resource allocation between modes.
- *Selective editing* strategies (manual treatment of some reports, automated treatment of others, with automated determination of which report gets which treatment) are being successfully tested and used, as well as strategies for selective sampling and follow-up in order to reduce respondent burden.
- In the future, cross-survey coordination will be implemented in order to avoid asking the same questions of an establishment for different surveys.
- Generalized systems are built in order to avoid duplication of programming effort as large development costs are spread among many surveys. They are constructed in modules to make maintenance and updating easier.
- Systems are becoming more portable with the advent of the micro-computer, and widely used operating systems are making it is easier to investigate the usefulness of systems and approaches to problems.
- Systems are more polished because the end users must interact directly with them. As a result end users are becoming more demanding.
- Systems can be implemented by non-programmers. This gives more control to subject matter specialists and methodologists.

Many positive results are already beginning to appear through the early implementation of these trends, some of which are discussed below. The coming years should see even more improvements in productivity, timeliness, and data quality.

2 FUNCTIONS AND FEATURES OF EDITING SYSTEMS

When an organization buys or develops an editing system it is necessary to have a methodical way of evaluating it. There should be two parts of such an evaluation: first is a profile of the functionality of the system; second is an **honest** statement of agency needs regarding specific features.

2.1 Functional Profile of Editing Systems

There are many functions that an editing system might be called upon to perform and different ways to perform the same functions. No system performs all functions so a profile

characterizes what a system does and how it does it. These systems often perform complementary tasks. For example, in Statistics Canada, the DC2 system (an interactive integration system) can be used for data entry, data collection, and interactive editing of reports of major impact while GEIS (an automated batch system) can then clear up many reports of minimal impact without human intervention.

Functional profiles of three kinds of systems representing much development work in the past years are displayed in Table 1. These are interactive integration, Fellegi and Holt, and top-down (macro or statistical) editing systems. An *interactive integration* system performs many different survey tasks, combining data editing with other functions such as data entry and interviewing. Data flow smoothly from one process to another, and programming code is applied across modules, for example, for interviewing, interactive editing, and data entry instruments. *Fellegi and Holt* systems automate some tasks formerly reserved for people. In the case where many records will have small impact on aggregated estimates, the system is trusted to not only detect errors but to decide which action should be taken and then take it. All actions comply with stated methodological principles. A *top-down* editing system (also known as macro or statistical editing) gives the editor a better look at the data starting from preliminary calculations of aggregates and provides a way to interactively trace outliers at the aggregate level to individual reports. See Granquist (1987) for a description of the main features of such a system.

[Insert Table 1 here]

For establishment surveys, examples of interactive integration systems include Blaise from the Netherlands Central Bureau of Statistics and DC2 from Statistics Canada. Examples of Fellegi and Holt systems include GEIS from Statistics Canada and SPEER from the U.S. Bureau of the Census. Examples of macro-editing systems include ARIES from the U.S. Bureau of Labor Statistics and the ***gred*** system from the New Zealand Department of Statistics. These systems will be discussed further in Section 3. Other systems do not fit into this functional categorization. For example a system used primarily to collect data through respondent electronic self-reporting is the PC Electronic Data Reporting Option (PEDRO) system from the U.S. Department of Energy (Heath, Proceedings Chapter ??).

2.1 Lists of Editing System Features

Once it is decided which kind of system is required then it is necessary to build a list of required and desired features. A list of editing system features was put together by the United States Federal Committee on Statistical Methodology (FCSM, 1990), Subcommittee on Data Editing in Federal Statistical Agencies. The FCSM list was originally based on an article written by Cotton (1988) where he evaluated four editing systems. This list covers features for all kinds of systems for all kind of surveys. An advantage of this list is that it appears in the Committee's report on editing in Federal Statistical Agencies along with a Glossary of Terms, a chapter describing the role of editing systems, some useful case studies, a somewhat detailed description of three editing systems, and short descriptions of several other systems. The FCSM list and glossary include brief explanations of the more esoteric words and phrases which infest this field of study. The FCSM list is meant to be modified for specific agency needs. For example, it was one starting point for the National Agricultural Statistics Service (NASS, 1992) list of features

for interactive survey processing including data editing, CAI, data entry, and survey management.

2.2 Use of Lists of Features of Editing Systems

The first step in modifying an existing list is to evaluate the agency's survey program, current and expected. The evaluation should focus on the agency's difficult surveys, noting specifically the difficult or unique aspects of such surveys. For example, the surveys may require a complicated data set structure, production of many similar but different versions of questionnaires, or extensive referral to external data files for line-by-line data checking. A good place to start is to gather a representative example of questionnaires, making sure that all difficult data processing and data collection situations are covered.

There should be a vision of where the agency is heading with its survey data processing system. The vision should encompass methodological and productivity goals and make some statement about how people are to work and how data are to be processed and passed to analysis and summary. This vision will help establish the relative importance of each feature. One strength of NASS's list of features (1992) is that each feature's relative priority of is stated.

The more detailed the list becomes, the more chance there is for disagreement on the evaluation committee. On the other hand, it is desirable to be as complete as possible as no one wants a system that cannot handle an important set of surveys that depend on one or a few critical features. The NASS list ended up being 400 items long and is probably a good compromise between thoroughness and workability. It should be seen as very positive if a candidate system has much more capability than required by the list of features. It will be almost impossible to anticipate every situation. The more powerful the system's language, data handling, and utilities the better it will be suited for unanticipated surveys.

Using the Profile and Lists

Editing systems that are under consideration for procurement should be pre-screened against a functional profile and the major headings in the list of features. The objective is to avoid spending too much time on systems that clearly do not meet specifications. Further evaluation of an editing system against the list should be done by someone in the agency with experience in the system. This may mean that the agency will have to invest in training in the system. However, major capabilities or limitations can be missed if the evaluation is done by hearsay or by relying too heavily on the developer for information. To the extent possible, the system should be programmed for the most difficult agency surveys or at least the difficult parts of them. The system under consideration should be evaluated against all features on the list. Most features will require a simple yes or no; others may require some comment as to how well the system satisfies the criterion.

Other non-technical evaluation considerations are (historic) rate of system development, responsiveness of the developer to clients' needs including the need for support, stability of the developer, and the adaptability of the system for future needs. Finally, the system must be evaluated according to whether it can handle the agency's survey program as a whole. For

example, even if a system can handle each of the agency's surveys singly considered, it may not be able to handle all of them together. This may occur if development and maintenance time for each survey is excessive due to a weak programming language which precludes the possibility of bringing more surveys into the system.

2.3 Major Features of Establishment Surveys Editing Systems

There are about 250 features listed in the FCSM (1990) list of editing system features under 8 major headings: General Features, Survey Management, Systems Items, Edit Writing, Types of Edits, Edit Rule Analysis, Data Review and Correction, and Support, Updates, and Training.

General Features

Included here are things that may immediately eliminate systems that cannot meet basic needs. General features include types of data handled (categorical, numeric, text), operating systems, kinds of computer platform (mainframe, PC, Local Area Network (LAN)), other software needed (data base, compiler), functionality (data collection, editing, analysis, summary, imputation), how it edits (batch, interactive), and availability (public domain, sold, or licensed). No system handles all tasks of all surveys and operates on all platforms and operating systems. Choices have to be made. Platform considerations become less important as power on the desktop increases and rapidly becomes less expensive. The computing paradigm is shifting from mainframe file-based activity to desktop record-based activity (Bethlehem and Keller, 1991). Even the under-appreciated DOS operating system on a LAN has enough power to handle any establishment survey as long as the system is targeted for that environment. For example, an interactive editing system will handle a survey one record at a time instead of the whole file at a time as a batch editing system would do. A personal computer can handle one record at a time for any survey. Interactive systems tend to be based on micro computers or work station platforms though there may be a connection to a data base server. Fellegi and Holt systems, because they work on large files and automate tasks previously reserved for humans, are used mostly on mainframe systems though there may be a work station aspect as well.

Systems based on the Fellegi and Holt methodology are restricted as to type of data that can be handled. Both the GEIS and the SPEER systems operate on real data but not on categorical data. This is due to the mathematical algorithms used for edit rule and error analysis.

Survey Management

Survey management features include the production of various kinds of reports such as lists of forms not yet received, validation of identification variables against a sample master, lists of records that need further treatment, and a statement about the impact of editing and imputation on survey estimates. In addition, the survey manager should be able to take actions based on the reports; for example, to call up a record and fix it, add or delete records, send a form back for follow-up, mail out reminders, and broaden or tighten edit limits. An editing system either must be able to handle survey management or it must be embedded within other systems that handle management. If the system handles survey management there is less work

in preparation and in production.

Systems Items

Features include types of data storage file formats, whether data can be read in or out and with which formats, how the system is put together (modular or not), whether it is interpreted or compiled, whether audit trails and log files are kept, whether it can read external data for edit check bounds, and whether it allows calls to sub-routines or has a macro capability. A well designed and executed system can be adapted to new surveys much more quickly than can a sloppily built one.

Edit Writing

This section is important to the subject matter specialist in preparing the system for a survey. Important features are whether a specialist can enter edits directly into the system or whether a programmer (or a specially-trained person) must do it instead. The former will allow the subject matter specialist greater control. Also included are formatting features for data display such as whether historical or calculated values can appear in an edit and edit message, whether it is possible to choose variables that are to be flagged, whether edit messages are in a spoken language or in a code (e.g., Error 333), and whether the edit writer has control over the user interface on a computer screen.

Types of Edits

Types of edits allowed in a system are as important as the kinds of data that are allowed. Some kinds of features are easily understood such as whether the system allows edits to contain complex conditional statements, edit checks against historical data, and levels of edit failure (e.g., fatal versus suspicious). Definitions of other features are more obscure; for example, ratio edits, linear edits, and consistency checks (however these are briefly explained in the FCSM (1990) list). Systems that cannot handle all kinds of edits must be embedded in or used with other systems, to perform for example, an additivity check. However, there may be enough value added by the use of such a subsystem (e.g., great imputation options) to make its use worthwhile.

Fellegi and Holt systems are constrained as to the type of edits that they can handle. GEIS in Statistics Canada is based on linear edits (mathematical representation is a straight line or plane) while the SPEER system is based on ratio edits. However, these systems are not as constrained as it might seem, as it is usually possible to recast other kinds of edits in an appropriate way. These limitations are a consequence of the mathematical basis of such systems.

Edit Rule Analysis

Edit rule analysis is unique to Fellegi and Holt systems. It allows the subject matter specialist the ability to evaluate the set of edits that are explicitly written by checking for redundancy or contradictions, by evaluating unstated edits that are implied by the explicit edits, and by checking the most extreme data that can pass edits without invoking an edit failure. In

Fellegi and Holt methodology a set of edits will be jointly used to define a valid range for imputations. The edit rule analysis is necessary to ensure that a good record can be defined. If not, every record will be in error.

Data Review and Correction

Covered here is how a person interacts with the system to review data and make corrections, (if applicable), for example, whether changes are made interactively on a computer screen or on paper which then have to be keyed in. Also covered is which kinds of imputation are allowed, for example automated and/or manual updates, and for automated imputations whether formula based and/or donor based methods are allowed (see Kovar and Whitridge, Chapter ?? for a discussion of imputation in establishment surveys).

Support, Updates, and Training

Included here is whether the system is available to external users. If it is available to others then there is a matter of support, obtaining explicitly needed features, training, documentation, and very importantly, whether the system is being continually updated and with what speed. Vendor commitment to using the system in its own organization is important, as is the kind of laboratory the vendor organization represents. This includes the number and variety of surveys that are handled and how successfully they are implemented.

3 GENERALIZED SYSTEMS

Many statistical agencies have spent considerable resources in developing generalized systems. In generalized systems, conventions are standardized, programmed, tested, and then applied to many different surveys. The system has to be adapted for each survey but the common system tasks are already handled, relieving the subject matter departments of developing the same things. Fellegi and Holt editing systems are described first. These systems are designed to perform much of or all of the subject matter review and to utilize the subject matter specialist in preparing the system to automatically handle bad and suspicious forms. All of this is done according to a set of guiding methodological principles. A second class of edit systems, known as interactive integration systems, integrate editing with data collection and high speed data entry among other things. The integration is twofold, first in the production of instruments to handle related tasks, and second in the use of those instruments to execute those tasks. The idea is to state specifications once and to use them in different ways. The third class of editing system described is advanced top-down (macro or statistical) editing with interactive graphical or tabular query and tracing to offending reports.

3.1 Fellegi and Holt Systems

Fellegi and Holt (1979) proposed an editing methodology whereby a computer program running in batch can inspect forms, detect problems, determine which items to replace, and then make imputations. Certain methodological principles are enforced by such programs: 1) each record satisfies all edits, 2) as few changes are made as possible, 3) editing and imputation must

be part of the same process, and 4) imputations must retain the structure of the data. The principles have been applied by two statistical bureaus to the treatment of continuous data. At Statistics Canada, the Generalized Edit and Imputation System (GEIS) is now in operation and was used to process the Census of Agriculture in 1991 among other surveys. The SPEER system from the United States Bureau of the Census has been used on several economic surveys since the early 1980s. These systems are described in Pierzchala (1990a and 1990b) and in FCSM (1990).

Role of the Subject Matter Specialist

Fellegi and Holt systems require subject matter specialists to shift much of their work from post-collection hand processing to pre-collection specification of edit and imputation rules. Both GEIS and SPEER have edit analysis capability that gives the specialist another perspective on edit rule specification. For example in GEIS, edit analysis consists of finding conflicting and redundant edit rules, generation of implied edits from two or more explicitly stated edits, and the creation of "extremal records". These *extremal records* are artificially generated records which demonstrate to the specialist the worst possible records that can pass through the system untouched. The implied edits allow the specialist to determine if the data will be constrained in an unintended way. After iterative specification, the resulting edit rules are applied automatically following data collection. Though one of the goals of the Fellegi and Holt approach is to eliminate post-collection review by subject matter specialists, this is not fully accomplished in establishment surveys. In the GEIS system data coding and callbacks will be carried out before the system is applied. When GEIS is applied, almost all remaining problems are cleared up in batch. However, there may be a few records that GEIS cannot handle and which must be reviewed by specialists (Legault and Roumelis, 1992). The SPEER system has combined the Fellegi and Holt methodology with interactive processing whereby the specialist can inspect and treat referred forms. The value of these systems is to consistently treat each form, to provide several agency-sanctioned imputation options, and to standardize procedures both within and between surveys while adhering to the stated methodological principles.

GEIS in Statistics Canada

GEIS may be regarded as an imputation system based on user-specified edits (Kovar, 1993). It can be adapted relatively quickly to most new applications, though there will usually be pre- and post-GEIS modules to build and run. Imputations are objective and reproducible. Imputation options include hot deck methods as well as several model-based methods that can incorporate trend adjustments. Many reports and audit trails are produced which help measure the impact of imputations on the data. The system is used in several different ways in Statistics Canada. For example, it is used in mass imputation in administrative data subsamples, as part of selective editing, and as an easy way to evaluate different imputation strategies. Mass imputation is used when incomplete administrative data are available for a population and a subsample of administrative records is surveyed to gather information for missing fields (Kovar, 1993). Unsampled records are then mass imputed in order to generate complete rectangular data sets, which is especially useful for ad hoc data requests (Kovar and Whitridge, Chapter ??). Mass imputation applications include the Census of Construction, the Annual Motor Carrier Freight Survey, Agriculture Tax Data program, and the Agriculture Whole Farm Database

Project. Selective editing is a strategy that is useful in highly skewed populations where relatively few reports have large impact on estimates. The high impact reports are dealt with manually while smaller firms of lesser impact are dealt with automatically by GEIS. A score function is defined to channel reports to the appropriate processing method. The Annual Survey of Manufactures has adopted the selective editing strategy using GEIS (Boucher, Simard, and Gosselin, Proceedings Chapter ??), and a similar use is being considered in the Survey of Employment, Payrolls and Hours. The system is also being used in the Motor Carrier Freight Surveys, and the income tax data acquisition program.

The largest use of GEIS to date is in the 1991 Agricultural Census (Roumelis and Legault, 1992). GEIS found quality donor records for imputation which had a reasonable impact on the estimates while ensuring that imputations satisfied all edits simultaneously. It was felt that edit and imputation methodology were considerably improved over previous censuses. On the other hand in order to treat 280,000 records there were about 3,500 job submissions for GEIS and 1,500 submissions for surrounding jobs. Much of this was due to breaking up the form into 16 edit groups (due to limitations of numbers of variables) while the population was divided into 52 imputation regions (similar farms within regions). Streamlining job submission procedures would not be straightforward because users wanted to look at output between each submission. Steps taken included: a) data preparation, b) fine tuning of edits (to avoid unwanted effects of imputation and to speed performance), c) simultaneous edit and imputation in GEIS, d) post-GEIS processing treatment of the few records GEIS could not handle, and e) imputation of total refusals. Computer use increased somewhat over the 1986 census while there was a slight decrease in human resources used. It is felt that the effort made for the 1991 census would be paid back in future censuses as the same modules will be used again.

SPEER in the U. S. Bureau of the Census

SPEER is used in several large establishment surveys and censuses, for example the Annual Survey of Manufactures, the Census of Manufactures, the Enterprise Summary Report and the Auxiliary Establishment Report of the Economic Censuses, and the Census of Construction (Draper, Petkunas, and Greenberg, 1990). In the Bureau of the Census, the SPEER system does not necessarily reduce the amount of human resources spent on editing a survey but allows the specialists to see more forms within the allotted time. SPEER refers selected records to work stations where they can be treated interactively. For selected records, data editors review actions taken in batch processing by SPEER and either confirm the validity of those actions or override them. As economic data are highly skewed and imputations can have a large impact on some statistics there is a need for some review capability. The on-line capabilities of SPEER are based on the same programs as the batch modules and ensure that editor actions are edited according to Fellegi and Holt principles as data are changed. The implementation of on-line capability addresses a particularly difficult processing situation in the Bureau of the Census where paper forms are received in Jeffersonville, Indiana and edited in Suitland, Maryland. This situation has traditionally led to many delays in the treatment of forms. By implementing an interactive capability, the batch cyclic process is reduced or eliminated. The SPEER system has also been used for data entry of late arriving forms in the Annual Survey of Manufactures. In this case it is possible to edit data as they are entered.

3.2 Interactive Integration Systems

Computer Assisted Data Collection and post-collection editing apply the same (or at least very similar) routing and edit strictures to a form. Some forms of high speed data entry also dynamically route the data entry clerk through a complicated form. Systems that can perform all these functions offer the possibility of saving resources both in the production of instruments and in the completion of production tasks. The Blaise System in the Netherlands Central Bureau of Statistics (NCBS) and the DC2 system in Statistics Canada were developed from the start with the goal of integrating several different survey tasks including Computer Assisted Data Collection, high-speed data entry, interactive data editing, and survey management as well as other tasks. They maximize the use of networked work stations. Blaise is described in Bethlehem (1987), Keller, Bethlehem, and Metz (1990), Bethlehem and Keller (1991), Pierzchala (1990a, 1990b, and 1991), and in FCSM (1991). Hundepool (Proceedings Chapter ??) describes the use of Blaise in establishment surveys in the NCBS. Woefle (Proceedings Chapter ??) describes the use of DC2 in multi-mode data collection efforts in Statistics Canada.

Interactive Editing

Blaise, SPEER, DC2, PEDRO, ARIES, and ***gred*** are among examples of interactive editing systems. A big advantage of interactive editing is that the editor gets immediate feedback from the results of edit actions. Forms are treated just once, cleaned, and stored. This eliminates the cyclical nature of inappropriate batch-oriented editing where some edit actions may engender other edit failures causing the specialist to revisit the same form days or weeks later. The dynamic computer screen can be used to efficiently display comments, historical information, and other auxiliary information as needed. Another advantage of interactive editing is that it is such a powerful way to edit data that it allows data to be key punched without a previous hand edit. In paper-based batch methods the penalty of not doing a hand edit before capture is too severe. There would be too many errors to clean up in this inefficient manner. Two studies in NASS have confirmed that interactive editing is more productive than batch paper methods. In the December 1989 Agricultural Survey in Wisconsin, there was a 13% time savings with interactive methods (without preliminary hand edit) over batch methods (with preliminary hand edit) (Pierzchala, 1991). In the 1993 June Area Frame Survey in Indiana there was 9% time decrease when the preliminary hand edit was partially removed and a 16% time decrease when the hand edit was totally removed (Eklund, Proceedings Chapter ??). However, the Indiana staff was not comfortable with totally removing the preliminary hand edit due to the complexity of the form. Further work must be done to define a minimal preliminary hand edit for this difficult survey. Where it works, the capture of raw data increases productivity by cutting down on the number of separate reviews, and it also allows a more automated way of evaluating the performance of the survey. If data are hand edited before data capture then much information about the badly performing parts of the survey is lost and must be gathered again through special inspection of paper forms.

In the U.S. Energy Information Administration the use of the interactive electronic self-reporting system PEDRO has reduced error rates by as much as 60 to 70 percent. For all petroleum supply forms, numbers of verification calls have fallen by more than 80 percent (Heath, Proceedings Chapter ??).

Results in the Netherlands Central Bureau of Statistics (NCBS)

In the Netherlands, the change from traditional batch editing procedures to interactive editing in Blaise and from mainframe processing to Local Area Network (LAN) processing was used as an opportunity to change procedures (Bethlehem and Keller, 1991). A goal has been to eliminate steps in the processing of data. As a result specialists now enter data while editing it for most establishment surveys. Data entry is slower than before; however overall processing time is reduced. The NCBS managed to absorb a 20% budget cut over a 4 year period recently without a cut in survey program (Bethlehem, 1992 and 1993) and a further 13% cut with minimal loss of program. Keller, Bethlehem, and Metz (1990) note that the change to decentralized processing has led to a reduction of 25% in the Automation Department of the NCBS as a result of reductions in data entry staff and the program development staff (because standard data processing tools decrease the need for tailor-made programs). Surveys experienced a large reduction in processing time therefore improving timeliness while at the same time improving data quality because of broad use of Computer Assisted Data Collection methods including electronic self-reporting in municipalities, fire brigades, and trading firms in the Netherlands (Hundepool, Proceedings Chapter ??). Automation infra-structure costs go down from year to year. In 1987 the NCBS spent 25 million guilders on automation infra-structure, in 1990 it spent 20 million, while in 1993 it spent 16 million guilders (excluding personnel costs). It is difficult to project the same kinds of savings to other agencies as there is no way of knowing relative efficiencies of each organization and their unique operating conditions and goals. However, these data are very suggestive that aggressive, focused, and well planned technological innovation based on well developed standard software can result in substantial savings while increasing data quality, reducing respondent burden, and improving timeliness.

3.3 Interactive Top-Down Graphical and Tabular Query

Top-Down (macro or statistical) editing has long been part of statistical bureaus' procedures but usually under different names such as data listings, statistical analysis, and so on. These have usually been part of batch paper processes and have included much tedious inspection of charts and tables. There is much value added to the process when the macro-statistical inspection is done interactively on a computer in a top-down method using a graphical or tabular screen display. For example, if an aggregate value is outside historically determined bounds the problem can be traced quickly to the offending report. The analyst can see what the effect is of changing the weight or correcting the datum. Care must be taken so that the process cannot be abused. The data editor is given so much power and vision that it would be very easy to inappropriately alter the data and make them appear consistent and plausible all the way down to the level of the individual report. The BLS has taken several steps to ensure that editor actions do not inappropriately affect important labor statistics. For example, audit trails are kept of editor actions and each editor is allowed to see only a specific part of the overall data file. Two systems, the ARIES system from the Bureau of Labor Statistics, and the ***gred*** system from the New Zealand Department of Statistics, put these concepts into practice. See Granquist Chapter ?? for a discussion of macro-editing.

ARIES from the United States Bureau of Labor Statistics

The Automated Review of Industry Employment Statistics (ARIES) system was developed by the Bureau of Labor Statistics for the macro treatment of data from the Current Employment Statistics (CES) survey. It greatly eases the review of data from about 300,000 reports in about 1600 basic estimating cells and about 1000 aggregate cells in an extremely tight time frame (Esposito, Lin, and Tidemann, Proceedings Chapter ??). The system uses graphical and tabular representations of the data at various levels of aggregation. An anomaly map gives an analyst a visual representation of an industry. The anomaly map is a tree of cells portrayed in a circular form in which the outer cells represent basic estimating cells and the inner circles represent data cells that are progressively more aggregated as they approach the center. The cells are connected by lines radiating outward from the center. Colors are used to flag large changes from prior periods data from the aggregate cells to the basic estimation cells. In this way the analyst can quickly trace movements at the aggregate level to one or several cells at a lower level. A mouse is used to click on nodes, the effect of which is to pop up historical data for that node. By such maneuvers the analyst can trace offending data back to the micro-level. There are also tabular ways to review data in the ARIES system. The use of the system has resulted in savings of 14 million lines of computer printout per year (80% savings) and has increased productivity by 50%, while being well received by the analysts using it. ARIES is a dedicated package meant to operate only on the CES survey; however, it is a very good model for other systems.

gred System from the New Zealand Department of Statistics

gred (short for graphical editor) is a general purpose graphical editing system for business and establishment surveys that is under development in New Zealand (Houston and Bruce, 1992). It displays unit records for a single variable for several periods at one time through the use of box plots (Houston, 1993). It is possible to click on a data point with a mouse and pop up a graph showing estimates for an industry with and without the contribution of a particular form. ***gred*** is not used to change data per se but is used to manually adjust weights for chosen firms. It is used primarily for output editing but can also be used to monitor surveys.

4 CONCLUSION

No automated editing system totally replaces the need for human treatment of forms. People enter data into the system while it is edited as in Computer Assisted Data Collection; or the system is used to enhance human review as in Interactive Editing; or the system is applied only after people have treated at least an important subset of the forms. People are still better than the computer at certain tasks, for example recognizing problems that the computer has not been programmed to find and applying subject matter knowledge. On the other hand computers are fast, consistent, and can accurately compare data across all forms. Using interactive graphical and tabular query methods it is easy to find offending reports from a massive file. A range of imputation procedures can be applied rigorously according to specified priorities. There is not necessarily a conflict between human and machine editing. The latter can and should make the former more satisfactory on methodological, personal, and productivity grounds.

REFERENCES

Bethlehem, J. G. (1993), letter to the author.

Bethlehem, J. G. (1992), letter to the author.

Bethlehem, J. G. (1987), "The Data Editing Research Project of the Netherlands Central Bureau of Statistics," *Proceedings of the Third Annual Research Conference of the Bureau of the Census,* pp. 194-203.

Bethlehem, J. G. and Keller, W. J. (1991), "The Blaise System for Integrated Survey Processing," *Survey Methodology*, Statistics Canada, **17**, pp. 43-56.

Boucher, L., Simard, J.-P., and Gosselin, J.-F., (1993), "Macro-Editing, a Case Study: Selective Editing for the Annual Survey of Manufactures Conducted by Statistics Canada," *Proceedings of the International Conference on Establishment Surveys*, Washington DC, American Statistical Association, pp. ?-?.

Cotton, P. (1988), "A Comparison of Software for Editing Survey and Census Data," *Proceedings of Symposium 88, The Impact of High Technology on Survey Taking,* Statistics Canada, pp. 211-241.

Draper, L., Petkunas, T., and Greenberg, B. (1990), "On-Line Capabilities in SPEER," *Proceedings of Symposium 90, Measurement and Improvement of Data Quality,* Statistics Canada, pp. 235-243.

Eklund, B. (1993),"Computer Assisted Personal Interviewing (CAPI) Costs Versus Paper and Pencil Costs," *Proceedings of the International Conference on Establishment Surveys*, Washington DC, American Statistical Association, pp. ?-?.

Esposito, R., Lin, D., and Tidemann, K. (1993), "The ARIES System in the BLS Current Employment Statistics Program," *Proceedings of the International Conference on Establishment Surveys*, Washington DC, American Statistical Association, pp. ?-?.

Federal Committee on Statistical Methodology (FCSM) (1990), "*Data Editing in Federal Statistical Agencies*, "Statistical Policy Working Paper 18, Washington, DC: U.S. Office of Management and Budget.

Fellegi, I. P. and Holt, D. (1979), "A Systematic Approach to Automatic Edit and Imputation," *Journal of the American Statistical Association,* **71**, pp. 17-35.

Granquist L. (1987), "A Report of the Main Features of a Macro-editing Procedure which is used in Statistics Sweden for Detecting Errors in Individual Observations," paper presented at the Data Editing Joint Group, Stockholm, Sweden: Statistics Sweden.

Heath, Charles C. (1993), "Personal Computer Electronic Data Reporting Option (PEDRO),"

Proceedings of the International Conference on Establishment Surveys, Washington DC, American Statistical Association, pp. ? - ?.

Houston, G. (1993), letter to the author.

Houston, G. and Bruce, A. (1992), "Graphical Editing for Business and Economic Surveys," unpublished report, Wellington, New Zealand: New Zealand Department of Statistics.

Hundepool, Anco (1993), "Automation in Survey Processing," *Proceedings of the International Conference on Establishment Surveys*, Washington DC, American Statistical Association, pp. ? - ?.

Keller, W. J., Bethlehem, J. G., and Metz, K.-J. (1990), "The Impact of Microcomputers on Survey Processing at the Netherlands Central Bureau of Statistics," *Proceedings of the Sixth Annual Research Conference of the Bureau of the Census,* pp. 637-645.

Kovar, J. G. (1993), "Use of Generalized Systems in Editing of Economic Survey Data," to appear in the *Bulletin of the International Statistical Institute: Proceedings of the 49th Session*, Florence, Italy.

Legault, S. and Roumelis, D. (1992), "The Use of the Generalized Edit and Imputation System (GEIS) for the 1991 Census of Agriculture," Working Paper No. BSMD-92-010E, Methodology Branch, Ottawa Canada: Statistics Canada.

National Agricultural Statistics Service (NASS) (1992), Report of the Interactive Survey Software Committee, "Criteria for the Evaluation of Interactive Survey Software".

Pierzchala, M. (1990a), "A Review of the State of the Art in Automated Data Editing and Imputation," *Journal of Official Statistics,* Statistics Sweden, **6**, pp. 355-377.

Pierzchala, M. (1990b), "A Review of Three Editing and Imputation Systems," *Proceedings of the Survey Research Section, American Statistical Association*, pp. 111-120.

Pierzchala, M. (1991), "One Agency's Experience with the Blaise System for Integrated Survey Processing," *Proceedings of the Survey Research Section, American Statistical Association*, pp. 767-772.

Stuart, W. J. (1966), "Computer Editing Survey Data - Five Years of Experience in BLS Manpower Surveys," *Journal of the American Statistical Association*, **61**, pp. 375-383.

Woefle, L. (1993), "Mixed Mode Data Collection and Data Processing for Economic Surveys," *Proceedings of the International Conference on Establishment Surveys*, Washington DC, American Statistical Association, pp. ? - ?.

Table 1: Survey Function versus Type of System

Function	System Type: Interactive-Integration	Fellegi and Holt	Top-Down, (Macro or Statistical)
Pre-survey edit rule analysis		Y	
High speed data entry	Y		
Computer Assisted Collection	Y		
Coding	Y[a]		
Interactive edit	Y	Y[b]	Y
Automated item deletion and replacement		Y	
Model-based imputation	Y	Y	
Donor imputation		Y[b]	
Graphical outlier inspection			Y
Tabular outlier inspection			Y[b]
See editing effects on estimates			Y
Weighting	Y[a]		
Tabulation	Y[a]		
Generalized meta-data export to other software	Y[b]		

[a] Included within the system itself or in a generalized family of software.

[b] Not all systems of this type have this capability.

Index for Mark Pierzchala, Editing Systems and Software

EVALUATION AND CONTROL OF MEASUREMENT ERROR IN ESTABLISHMENT SURVEYS

Paul Biemer
Research Triangle Institute

Ronald S. Fecso
National Agricultural Statistics Service

1 INTRODUCTION

Survey estimates will almost never be identical to the population value they are trying to measure as a result of two types of error which are known widely in the survey literature as sampling error and nonsampling error. *Sampling error* is the difference between the survey estimate and the population value of interest obtained by a census keeping other conditions similar, that is, error due to surveying only a subset of the population rather than conducting a complete census of all establishments in the target population. In a complete census of establishments, the census estimate may still differ considerably from the population value as a result of nonsampling error. *Nonsampling error* is the difference that is attributable to all other sources not including sampling error--errors arising during the planning for and the conduct of establishment surveys, as well as during the processing of the data and the preparation of the final estimates.

As shown in Figure 1, the sources of nonsampling error may be classified as either (1) specification error, (2) frame errors, (3) nonresponse errors, (4) processing errors, and (5) measurement errors. *Specification error* occurs when the survey concepts are unmeasurable or ill-defined; the survey objectives are inadequately specified; or the data collected do not correspond precisely to the specified concepts or target variables. *Frame errors* include erroneous inclusions, omissions, and duplications in the sampling frame or in the sampling process. Also included are errors in listing subunits within establishments in surveys that require subsampling within establishments. *Nonresponse errors* include unit nonresponse, item nonresponse, or incomplete data. *Processing errors* refer to the errors in post-data collection processes such as manual editing, coding, keying, computer editing, weighting, and tabulating the survey data. Finally, *measurement error*, the topic of this chapter, refers to those errors which occur at the time of data collection.

At a recent conference on the subject, Biemer, et al. (1991) defined survey measurement error as:

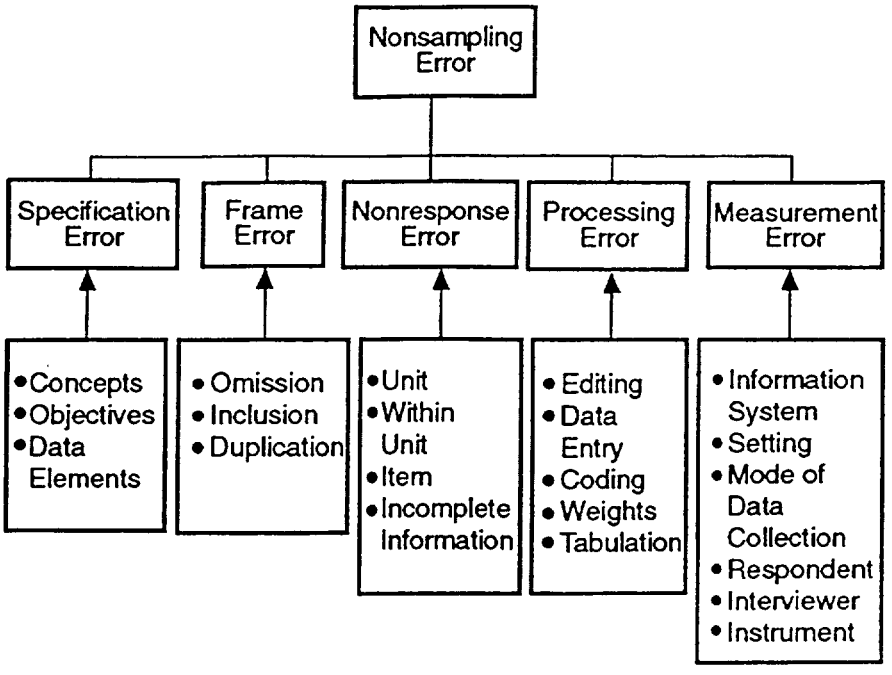
> "... error in survey responses arising from the method of data collection, the respondent, or the questionnaire (or other instrument). It includes the error in a survey response as a result of respondent confusion, ignorance, carelessness,

> or dishonesty; the error attributable to the interviewer, pershaps as a consequence of poor or inadequate training, prior expectations regarding respondents' responses, or deliberate errors; and the error attributable to the wording of the questions in the questionnaire, the order or context in which the questions are presented, and the method used to obtain the responses. At the time survey responses are collected, all of these factors may intervene and interact in such a way as to degrade response accuracy."

There are five general sources of measurement error in establishment surveys that will be discussed in this chapter: (1) the survey instrument, (2) the respondent, (3) the information system, (4) the mode of data collection, and (5) the interviewer, if one is used (see Figure 1). The *survey instrument* refers to the survey form or questionnaire and the instructions to the respondent for supplying the requested information. The *respondent* refers to the person(s) whose task is to supply the requested information, either by accessing the establishment's information system or by relying on either their own knowledge or the knowledge of others. Perhaps the most important feature of establishment surveys and one that distinguishes these surveys from household surveys is a greater reliance on *information or record systems* for the required information. These include the establishment's formal record systems, company or individual files, and informal records. The *mode of data collection* refers to the combination of the communication medium (i.e., telephone, face to face, self-administration, etc.) and the method of data capture (i.e. paper and pencil, computer keyboard, telephone keypad, etc.). Finally, in some establishment surveys, the *interviewer* may play a key role in obtaining the information from the establishment and entering the information into the data collection instrument. For some surveys, the interviewer may simply be a prerecorded voice over the telephone or may be nonexistent for self-administered surveys.

Although all the sources of error in Figure 1 can substantially reduce the accuracy of survey results, our focus in this chapter is on measurement error. Our primary goals are to review the methods for controling and evaluating measurement error and to study the uses of these methods in establishment surveys. Before discussing these methods, two general models for the response process will be briefly reviewed. One model is a cognitive model of the response process proposed by Tourangeau (1984) and adapted for establishment surveys by Edwards and Cantor (1991). The second model is a statistical model for measurement error originally developed by Hansen, Hurwitz, and Bershad (1964). Knowledge of both models is necessary to understand the evaluation and control techniques that will be subsequently described.

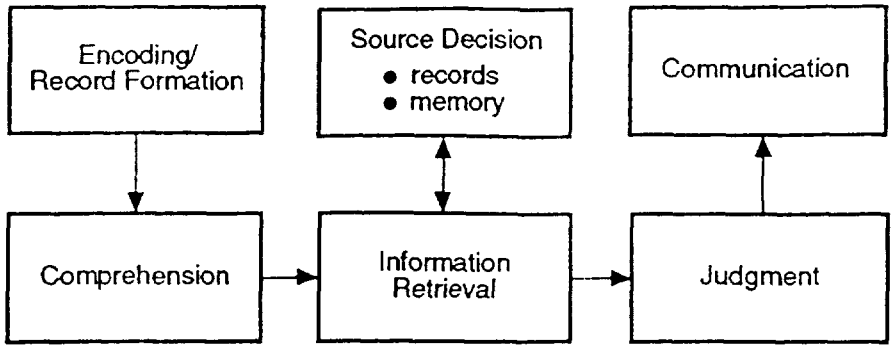
FIGURE 1 *Sources of Nonsampling Error*



2 A COGNITIVE MODEL OF ESTABLISHMENT SURVEY RESPONSE

Tourangeau (1984) and, more recently, Eisenhower, et al (1992) described five stages of the survey response process for household surveys. As depicted in Figure 2, the five stages are: (1) the encoding of information, (2) comprehension of the survey question, (3) retrieval of information, (4) judgment, and (5) communication. These stages reflect a somewhat idealized paradigm for response which may be described as follows: In formulating a response to a survey question, the respondent must first have the knowledge, belief, or attitude required to provide a valid response (*encoding*). There must be a shared meaning among the researcher, the interviewer, and the respondent with respect to each of the words in the question as well as the question as a whole (*comprehension*). To respond to questions concerning events or behaviors occurring in the past, the respondent will attempt to retrieve the required information from memory (*retrieval of information*). Once the information has been retrieved, the respondent decides how to respond appropriately, taking into account risk, benefit, available answer choices, etc. (*judgment*). Finally, the respondent communicates the response to the interviewer (*communication*). For a single question, the respondent may cycle through some or all of the stages repeatedly as the interviewer continues to probe for a codable or otherwise satisfactory response.

FIGURE 2 *Stages of the Response Process*



Adapted from Edwards and Cantor (1991).

As Edwards and Cantor recognized, this response paradigm does not describe a typical response process for an establishment survey. Establishment surveys rely much more extensively on information (record) systems than do household surveys. As an example, when an establishment survey respondent accesses company records to provide a response for a survey item such as "the number of full time employees during the period (date) to (date)", "encoding" and "retrieval of information" take on new meaning. The respondent must have knowledge of where to obtain the required data. The cognitive encoding of this basic *procedural* information is not enough, however, since the required information must have been entered in the company's employee record system in order to be retrieved by the respondent. This stage of the response process, which is analogous to cognitive encoding in the household survey model, is called *record formation* by Edwards and Cantor. The retrieval of survey information may require the respondent to invest considerable effort to locate and access the appropriate record sources. The respondent may have a number of options for responding to the request and the accuracy of the information may vary considerably with the effort required to access each potential source. The traditional household response model does not explicitly account for this *source decision* stage of the process. To account for these differences between the household and establishment surveys' paradigms, Edwards and Cantor proposed a revised response model for establishment surveys which explicitly accounts for the establishment survey respondent's reliance on information systems rather than recall to respond to survey questions (see Figure 2).

An important benefit of response models is that they allow survey methodologists to decompose the response process into smaller steps that may be treated separately in the design and evaluation of surveys. Let us now consider how the response model can be used as an aid in the control and evaluation of measurement errors in establishment surveys.

2.1 Encoding/Record Formation

All establishment survey questions, to some extent, rely on the knowledge, attitudes, and/or beliefs of the respondent. Thus, careless or uncontrolled selection of the establishment respondent can be an important source of measurement error. The survey procedures should clearly specify the minimum qualifications the respondent should posess in order to be confident in the accuracy of the respondent's reports. An important aspect of the interviewer's or the addressee's (for mail surveys) task, then, is to identify a respondent

meeting these minimum qualifications. In some cases, multiple persons within an establishment may serve as respondents for various parts of the questionnaire as appropriate. This may necessitate separate questionnaires for each of the respondents, or alternatively, a single questionnaire may be used with separate sections corresponding to the changes in respondents. For surveys which rely on information contained in the establishment information system, knowledge of the information system as well as the terms and definitions of the data items are essential qualifications for the respondent. Ultimately, however, the choice of respondent is the decision of the establishment's management.

The record formation process is subject to a number of error sources which may undermine the integrity of the data. Some record systems allow input from multiple sources with little or no validation or editing of the input. As an example, university student enrollment records may be updated by administrative assistants in each college. Errors are corrected when they are identified as the system is used but no formal system for error detection may exist. Another common problem in using administrative record systems to obtain survey data is the inconsistency between the record items and the survey items. The record system item may include unwanted data components or exclude components which are part of the survey item definition. The record values may reflect a time frame different than desired for the survey. The burden on the respondent to adjust for the errors may be so great that few are willing to do so.

Incompatibilities between the survey and the establishment record system can sometimes be remedied by revising the survey item definition if respondents are unwilling or unable to revise their record systems. Ponikowski and Meily (1989), in a study of the Current Employment Survey (CES), reported that 59% of the CES establishments did not adhere entirely to the survey employment definition. The most common problem cited was the inclusion of employees on unpaid vacation. Only 56% of the respondents who made this error agreed to adjust their figures to correct the error for future waves of the survey.

2.2 Comprehension and Retrieval of Information

The next two stages in the establishment survey response process are comprehension of the survey question and the retrieval of required information, either from memory or from records. Corby (1984) reports that in a study to evaluate the accuracy of the 1977 Economic Censuses data, these two stages of the process explained most of the error found in the censuses data.

Comprehension

Comprehension of the survey question is particularly problematic in establishment surveys. Often, the questions contain "legalese" and technical jargon which are not well-understood by all respondents. Further, the questions may require that the respondent aggregate data values across time or other units and it is not clear what components of data should be included or excluded. The Economic Censuses study provides many examples of this. For Annual Payroll in the 1977 Census of Manufactures, the amount of erroneous inclusion and exclusions totaled $3.7 billion or about 2% of the census total. Erroneous

exclusion of vacation pay accounted for almost one third of this error. In a study conducted by the National Agricultural Statistics Service (O'Connor, 1993), respondents were asked to define "calf", a common term in agricultural surveys. Over 30 different meanings were recorded. Among them were: not yet weaned; up to 800 lbs; under 600 lbs.; anything under a yearling; under 300 lbs.; everything except stock cows; 8 months or less; and not sure.

Retrieval of Information

Following comprehension, the next stage in the response process is the retrieval of information. Records which contain the required information may be available and accessed by the respondent with some effort. If there are no records or the burden to retrieve the information from records is too great, the respondent may attempt to either recall the information or to "estimate" the correct response. Eisenhower, et al (1992) and Groves (1989) provide excellent reviews of the literature on recall error. Estimation error accounted for 75% of the error in the reported value of employment for the 1977 Census of Retail Trade (Corby, 1984). Krosnick and Alwin (1987) suggest that survey respondents "satisfice"; that is, they exert minimal effort in responding to survey questions often providing answers which are "good enough" rather than precise responses. This theory may operate both for the retrieval of information from memory as well as the retrieval from records.

2.3 Judgment and Communication

Following comprehension and information retrieval, the last two stages of the response process involves the respondent's judgment regarding the appropriate response and communication of the response to the interviewer or directly in the instrument. For a closed-ended question, the respondent must decide which response category best fits the information retrieved in the previous stage. For open-ended responses, the respondent must formulate his/her own response. The respondent may judge a response as posing some economic or social risk or threat and alter the response accordingly.

Response judgment and communication may be influenced by the characteristics of the interviewer and the interview mode as well as by the instrument. An interviewer who probes completely and persistently until an acceptable response is obtained can counteract some of the effects of satisficing. Deliberate misreporting arising from a perceived economic or legal risk may be somewhat abated by emphasizing the neutrality of the survey sponsor. For self-administered surveys, the mechanism by which a respondent records his/her response, such as a laptop computer or telephone keypad, can exacerbate communication problems if it is faulty or inefficient.

3 A STATISTICAL MODEL FOR ESTABLISHMENT SURVEY RESPONSE

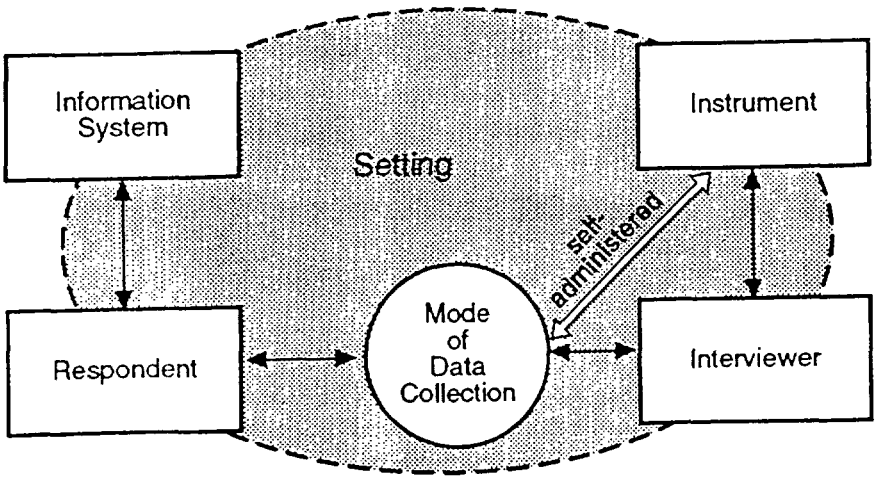
The cognitive model of survey response can be used as an aid for understanding and investigating the causes of response error in surveys. By considering each stage of the response process in the design of a survey, the problems associated with the successful completion of the stage can be identified and mitigated. However, the response model

focuses primarily on the respondent. Other factors that may induce measurement error, such as the interviewer, the mode of data collection, and the instrument, are represented in the model only through their effects on the respondent. In this section, we consider an alternative model which explicitly represents these other error sources. Such a model is useful for measuring how much of the total error in a survey estimate is attributable to the interviewer, mode, etc. without regard to which stages of the response process (Figure 2) are affected.

3.1 Sources of Measurement Error

The model postulates that the error in an estimator which is due to measurement error is attributable to six error sources: (1) the information system, (2) the respondent, (3) the mode of data collection, (4) the interviewer, (5) the survey questionnaire or instrument, and (6) the interview setting. As depicted in Figure 3, the respondent is the link between the interviewer (or instrument, for self-administered surveys) and the establishment information system. Communication through this link may be distorted by the mode of data collection. Each error source may give rise to errors in the reported values. Incorrect data may reside in the information system and these data are duplicated in the survey reports. As we have seen in Section 2, the respondent may give rise to an error at each of the five stages of the response process. Errors may be attributable to the mode of data collection in that the data collected by one mode may be more accurate than that of another. Although the mode does not "commit" errors in the way that respondents and interviewers do, the mode may "cause" an error by interfering in or somehow hampering the response process. Likewise, the instrument or questionnaire may cause errors through the construction and/or position of the questions, the format/layout of the questionnaire, etc. Finally, the interview setting may contribute to measurement error through its affects on the response process. The setting includes such factors as the survey sponsorship, location of the interview, degree of confidentiality and privacy perceived by the respondent, and so on. Focusing on these error sources, as well as on the stages of the response process, allows us to further isolate the causes of measurement error and to, thereby, understand what actions are necessary to eliminate the causes. A number of statistical methods exist which allow the survey methodologist to partition the total error into error components corresponding to each source and to estimate their magnitudes. In the next two sections, we will present a brief introduction to these statistical methods.

FIGURE 3 *Sources of Measurement Error*



3.2 A Brief Introduction to Statistical Modeling

Several recent references are recommended as introductions to the modeling of measurement error: Groves (1989) Chapter 1, Biemer and Stokes (1991), and Lessler and Kalsbeck (1992) Chapter 10. This section provides a brief background in the concepts and notation of measurement error modeling. These ideas provide the foundation for the error evaluation and control techniques that will be discussed in Section 4.

Our approach to measurement error modeling is an adaption of earlier work by Hansen, Hurwitz, and Madow (1953) and Sukhatme and Seth (1952). The simplest form of the model specifies that a single observation, y_j, from a randomly selected respondent j, is the sum of two terms: a true value, μ_j, and an error, ε_j. Mathematically, this may be written as

$$y_j = \mu_j + \varepsilon_j \tag{20.1}$$

where $\varepsilon_j \sim (0,\sigma_j^2)$, $\mu_j \sim (\mu,\sigma_\mu^2)$, and all covariances between the terms on the right are zero. Under this model, the variance of the mean, $\bar{y}$, of a sample of n observations, ignoring the finite population correction factor, is

$$\text{Var}(\bar{y}) = \frac{\sigma_\mu^2}{n} + \frac{\sigma_\varepsilon^2}{n} \tag{20.2}$$

where σ_μ^2 is the finite population variance of the true values and, $\sigma_\varepsilon^2 = E(\sigma_j^2)$, is the finite population mean of the individual response variances, σ_j^2. The term, σ_ε^2, is often referred to as the *simple response variance* (SRV).

Finally, we may define the reliability ratio, R, as

$$R = \frac{\sigma_\mu^2}{\sigma_\mu^2 + \sigma_\varepsilon^2} \tag{20.3}$$

Since $\text{Var}(\bar{y}) = R^{-1}\ \text{Var}(\bar{\mu})$, where $\bar{\mu}$ is the sample mean of the true values. R determines the increase in variance of the sample mean or total due to measurement error. R is widely used as a measure of the stability of the response process. A ratio of $R=1$ indicates total reliability ($\sigma_\varepsilon^2 = 0$) in the measure while a ratio that is near $R=0$ indicates the lack of response stability, i.e., the variability among the true values, μ_j, in the population is small relative to the variability attributable to the response process.

There are a number of extensions of this simple model which serve purposes in the exploration of measurement error. Two extensions which will be considered here are the extension to: (1) accomodate remeasurements of the same sample unit and (2) incorporate multiple sources of measurement error.

Extension for Reinterview Data

Some methods for evaluating measurement error require at least two measurements on some of the sample units. Let $y_{j\alpha}$ denote the observation on unit j on the α-*th* occasion. Extending (1) we have

$$y_{j\alpha} = \mu_j + \varepsilon_{j\alpha} \tag{20.4}$$

where $\varepsilon_{j\alpha} \sim (0, \sigma_{j\alpha}^2)$ and we assume that the covariance involving the terms on the right are zero both between trials and within trials.

A special case of (4) assumes that $\sigma_{j\alpha}^2 = \sigma_j^2$, $\alpha = 1,2$, that is, the second trial produces responses which are distributed identically to the responses of the first trial. This model is usually assumed for reinterview studies where the second interview is intended to be an independent replication of the first interview. Under these assumptions, an estimator of the SRV is

$$\hat{\sigma}_\varepsilon^2 = \sum_j (y_{j1} - y_{j2})^2 / 2n \tag{20.5}$$

and an estimator of R is $(1-\hat{\sigma}_e^2/\hat{\sigma}_y^2)$ where $\hat{\sigma}_y^2$ is the average of

$\hat{\sigma}_{y\alpha}^2 = \sum_j (y_{j\alpha}-\bar{y}_\alpha)^2/(n-1)$, for $\alpha = 1,2$.

There are a number of reasons why the assumptions for (4) may not be satisfied in practice (discussed further in Section 4.4).

Another use of the model in (4) is the estimation of measurement error bias using record check studies or true value reinterviews. For this purpose, the assumptions associated with the error term, $\varepsilon_{j\alpha}$, are altered. Instead, assume $\varepsilon_{j1} \sim (B_j, \sigma_j^2)$ and $\varepsilon_{j2} \equiv 0$. That is, assume that the error associated with the first observation has a non-zero mean (referred to as measurement bias) denoted by B_j and that the second observation has no error. Under this model

$$E(\bar{y}) = B \tag{20.6}$$

where $B = E(B_j)$, the measurement bias in the estimator $\bar{y}$. Thus, the aim of studies which employ (4) under this set of assumptions is to estimate B. Under the model assumptions, the bias can be estimated by

$$\hat{B} = \bar{y}_1 - \bar{y}_2. \tag{20.7}$$

Again, there may be severe difficulties in designing evaluation studies which satisfy the "true value" assumptions (discussed subsequently in Section 4.3).

Extensions to Incorporate Error Sources

The statistical models discussed previously do not explicitly provide for the partitioning of the error variance or bias into separate components corresponding to the five error sources in Figure 3. For example, suppose we wish to investigate the contribution to the error variance due to interviewers. To accomplish this, let y_{ij} denote the response obtained from the *j-th* respondent in the *i-th* interviewer's assignment. Then partition the error as $\varepsilon_{ij} = b_i + e_{ij}$ where b_i is a random error variable associated with interviewer i and e_{ij} is the difference $e_{ij} = \varepsilon_{ij} - b_i$. Substituting this error term into (1) we have

$$y_{ij} = \mu_{ij} + b_i + e_{ij} \tag{20.8}$$

where $b_i \sim (0,\sigma_b^2), e_{ij} \sim (0,\sigma_e^2)$, and as before, all terms on the right are uncorrelated. If we further assume that interviewer assignments are *interpenetrated*, that is, each assignment is a

random subsample of the initial sample, then the model can be rewritten as the traditional ANOVA model; viz.,

$$y_{ij} = \mu + b_i + e'_{ij} \tag{20.9}$$

where $e'_{ij} = e_{ij} + (\mu_{ij}-\mu)$, $e'_{ij} \sim (0,\sigma^2_{e'})$ and μ is a fixed constant. Using ANOVA methods, the interviewer variance, σ^2_b, can be estimated as a function of the between and within interviewers sums of squares (see Groves, 1989, p. 318).

A useful measure of the contribution of interviewers to the error variance is

$$\rho_y = \frac{\sigma^2_b}{\sigma^2_b + \sigma^2_{e'}} \tag{20.10}$$

referred to as the *intra-interviewer correlation.* An estimator of ρ_y is also available from the analysis of variance of interviewers.

The model (9) can be readily extended to include other sources of error that may be viewed as random effects. Examples are coders, editors, and supervisors. Additional uses of this model are discussed in Section 4.2.

When the model effects associated with an error source must be considered not as random variables but as fixed effects, a slightly different model is more appropriate. For example, we may be interested in three particular modes of data collection or two alternative question wordings. Further, we may be interested in the response reliability under the alternative modes or question versions as well as their contributions to response bias. To fix the ideas, we shall consider a model appropriate for investigating mode effects.

Rewrite model (8) as

$$y_{ij} = \mu_{ij} + M_i + e_{ij} \tag{20.11}$$

where now y_{ij} is the response of the *j-th* respondent assigned to the *i-th* mode and M_i is a constant effect associated with the *i-th* mode. Again, by randomly assigning respondents to modes, we may invoke the ANOVA model for analyzing the mode "biases." Under this model, the mode "differential biases," $M_i - M_{i'}, i \neq i'$ are estimable but the biases are not unless one assumes that $M_i = 0$ for some i. As an example, in a study of the effects of face to face interviewing, telephone interviewing, and mail self-administered methods, we may be

willing to assume that the bias associated with face to face interviewing is zero for some characteristics. In this case, the mode biases associated with the other two modes are estimable.

Finally, we may be interested in comparing the reliability ratios, R_i, associated with alternative modes or questionnaire versions. For this objective, we extend the model (3.11) for repeated measurements in analogy to (4) as follows:

$$y_{ij\alpha} = \mu_{ij} + M_i + \varepsilon_{ij\alpha} \tag{20.12}$$

and assume $\varepsilon_{ij\alpha} \sim (0,\sigma^2_{i\alpha})$. The reinterview methods described earlier may be used to obtain estimates of σ^2_i and, thus, R_i.

4 SOME TECHNIQUES FOR ERROR EVALUATION AND CONTROL

In this section, we briefly describe a number of techniques for evaluating and controlling components of the measurement error models described previously and provide examples of their use. The methods discussed are: (1) cognitive laboratory methods, (2) experimental design, (3) observational studies, (4) administrative record check studies, (5) true value reinterview studies,(6) replicated reinterview studies, and (7) external and internal consistency studies.

4.1 Cognitive Laboratory Methods

Cognitive methods encompass a wide array of exploratory techniques for either (1) investigating one or more components of the cognitive response model in Figure 2 or (2) identifying the types of errors that may be introduced during data collection. In cognitive laboratory methods, the objective is usually to obtain information on the *causes* of error and the ways in which alternative survey methods or design features affect survey response. Laboratory studies usually involve small sample sizes (less than 100 subjects) and, because these studies are usually conducted at a small number of sites where survey conditions can be tightly controlled, nonrandom or restricted random samples. Therefore, it is usually not possible to draw inferences regarding the magnitudes of measurement error components from these studies. For inferring the magnitudes of error, field studies, observational studies, or embedded survey experiments, to be discussed subsequently, are much more effective.

Forsyth and Lessler (1991) provide a comprehensive summary of the cognitive laboratory methods which have been used in demographic surveys. However, documented applications of cognitive laboratory methods to establishment survey methodology are relative rare. In Chapter XX, Dippo, Chun, and Sander describe the recent research in the application of some of these methods to establishment surveys.

4.2 Experimental Design and Observational Studies

Whether it be an investigation of alternative modes of interview, a test of alternative question wordings, or the estimation of interview effects, randomized experimental design has played a pivotal role in the development of measurement error theory and methods. In one of the earliest uses of experimental design, Mahalonobis (1946) randomized (*interpenetrated*) interviewer assignments in an agricultural survey in order to obtain evidence of interviewer biases. Hansen, Hurwitz and Bershad (1961); Fellegi (1964); and Bailar and Dalenius (1969) formalized and further developed the method of interpenetration. The discussion in Section 3.2 regarding the use of ANOVA for estimating interviewer variance is a special case of this more general methodology.

The use of randomized experimental designs in full scale surveys is referred to as an *embedded experiment*. Experimental designs have also been used in small scale field tests and in cognitive laboratory experiments. In the laboratory, where the experimental manipulation of survey conditions is easier, designs with very complex treatment structures are possible. For example, O'Reilly, et al. (1993) describes a laboratory study employing an incomplete block, repeated measures design in which three alternative modes of data collection were tested.

In the field, treatments are typically simpler, simultaneously manipulating only one or two experimental conditions. However, the randomization of experimental units may be considerably more complex. As an example, in a test of two types of interviewer training and two modes of data collection, interviewer training types may be randomly assigned to primary sampling units while the mode of interview treatment may be randomly assigned to establishments within the interviewer assignments.

In many situations, the required randomizations and experimental manipulations would be too expensive, too impractical, or simply impossible to perform and "nonexperimental" or observational studies are conducted. As an example, to answer questions about a possible causal relationship between the position of the survey respondent within the establishment and the accuracy of responses to a mail questionnaire, an experimental study would need to control who in the establishment completes the questionnaire for a large number of establishments over a considerable time period. Due to the difficulties of controlling such an experimental assignment of respondents as well as the costs, an experimental study is not feasible. Further, the results of such a study may be affected by the effort exerted by the researchers to control the type of respondent. Although the researcher may instruct the establishment as to whom should complete the instrument, this is ultimately decided within the establishment.

In an observational study, the survey proceeds normally without experimentally manipulating the type of respondent. Rather, when the returns are received, the reseacher observes the position of the respondent in the establishment and relates this variable to response accuracy in much the same way as in an experimental analysis. For example, we may have observed that response accuracy tended to be better for members of the establishment's clerical staff than for members of the management team. However, we may then ask whether this result has as much to do with the establishment's size as it does with the respondent's position. In larger establishments, special clerical positions may be established for the purpose of furnishing accounting information for regulatory purposes, while in smaller establishments where such managerial luxuries are not affordable, the task falls to the busy proprietor or his/her spouse. To examine this possibility and other similar ones, explanatory variables, such as establishment size, are incorportated in the analysis to eliminate these as causes of the observed relationship. Of course, there is always the risk that some important explanatory variable escapes the researcher's attention or is not available to be incorporated into the analysis and conclusions regarding causality will be misleading. The literature on causal modeling (see Stolzenberg and Land, 1983, for a review) provides a set of methods and principles for making causal inferences using observational data.

4.3 Administrative Record Check and True Value Reinterview Studies

While cognitive methods may provide valuable insights into the response process, it is often desirable to estimate the magnitude of measurement bias. To do this requires that the true value of the characteristic be known. With such information, faulty data items can be identified, measurement variance and bias can be estimated, and, in some cases, the sources and causes of error can be determined. As described earlier in Section 3.2, record check and true value reinterview studies are the two most widely used techniques for estimating measurement bias.

Administrative Record Check Studies

Because of the unavailability or inaccessibility of relevant administrative data for the characteristics typically measured in establishment surveys, and the expense of obtaining records from official sources, record check studies occur relatively infrequently. For this

technique, survey responses are compared to administrative records such as federal/state income or sales tax reports, licensing information, or other government data. Privacy limitations and differences in record-keeping and reporting requirements often limit the ability to make suitable administrative matching studies. In one type of record check study, a sample of establishments is drawn from the establishment frame and then the corresponding administrative records are located. Alternatively, we may start with a sample of records and then interview the corresponding establishments. The later is usually more efficient because the precision for rare items can be better controlled. While selecting records from the files of accountants and other financial services can be inexpensive, the coverage of the population can be poor. For example, Willimack (1989) found that only about 11 percent of farmers reported using financial services.

An implicit assumption of the administrative record check method is that the records contain accurate information (i.e. true values) for the survey charateristics of interest. Under this assumption, estimates of B can be obtained as in (7). However, three problems typically plague the method and limit the usefulness of the records data:

- The time periods for the record data and the survey data may not coincide,
- The characteristic(s) being reported in the record system may not exactly agree with the characteristic(s) being measured in the surveys, and
- To save evaluation study costs, the records study may be confined to a very restricted geographic area and inferences beyond this restricted population may be invalid.

The study by the Federal Committee on Statistical Methodology (1983) provides some guidance on methodological requirements to conduct record checks. Further, Ponikowski and Meily (1989) provide an application where data on employment, earnings, and hours worked reported by establishments in the Current Employment Statistics survey were examined.

"True Value" Reinterview

When accessing administrative records is not feasible because of their availability, cost, privacy concerns or other reasons, a reinterview approach which aims to obtain the true values of survey variable is an alternative. In the reinterview, more time may be taken and greater attention given to the reporting task in an effort to obtain highly accurate data and to reconcile originally reported data against a reinterview report. Reinterview may require that respondents' access company records for "book values" whenever they are available. Yet, many small businesses may keep records informally in a "shoebox" file. Willimack (1989) found that 25 percent of U.S. farm operators used such informal records, while only 3 percent used a computer, and 16 percent used a formal workbook.

In household surveys, it is well-known that obtaining a true value during reinterview is very difficult, and, indeed for some variables such as attitudinal data, true values may not

exist. Establishment data, while seeming to be more factually based, still presents challenges for reinterviews. For example, Van Nest (1985) found that when computer records did not keep breakdowns of data such as expenditures for new versus used equipment, true values for the detailed data could not be obtained.

The reinterview designs vary depending on assumptions which are most reasonable for the particular data and survey. Because the reinterview is often combined with interviewer performance evaluation, the reinterviewer is often the supervisor of the original interviewer. Forsman and Schreiner (1991) discuss some of the difficulties associated with this practice. Alternatively, "more experienced" interviewers have been used provided that they do not reinterview respondents whom they previously interviewed. Although reinterviews conducted by telephone are less costly, some data might best be obtained in a personal interview to ensure records are consulted. Cantwell, et al. (1992) list questions which need to be resolved when designing a true value reinterview, including:

- How should the reinterview be introduced to the respondent?
- Should there be prior notification of reinterview?
- How soon after the original response should the reinterview take place?
- Should question wording be identical?
- How much burden should be placed on respondents?
- Should proxy responses be allowed in reinterview?
- Should the respondent or the reinterviewer determine whether the original response or the reinterview response (or neither) is correct?

The Census Bureau used true value reinterviews to study the accuracy of data in economic censuses (Corby, 1984; 1985). For this study, one of the characteristics of interest was operating expenses including components such as cost for purchased advertising, unemployment compensation paid, and so on. The measurement biases were estimated for each component and then these were combined and later used to correct the reported data. An interesting aspect of the study was the classification of the reinterview responses into three categories: (1) book figures or values the respondent obtained directly from company records; (2) estimated values deemed reliable by the reinterviewer, and (3) estimated values deemed unreliable by the reinterviewer. For unreliable items , the respondent was asked to provide a range of values to include the highest and lowest value the true figure could take.

Another example is provided by Fecso and Pafford (1988) from a reinterview study used to measure bias in a National Agricultural Statistics Service (NASS) survey. Because of the detailed nature of acreage, stocks and livestock inventory items, NASS has relied primarily on personal interviews in the past to get the most accurate answers from the farm population. For example, in collecting on-farm grain stocks data, farmers may store grains in multiple bins on property they own and/or rent. In addition, farmers often are involved in multiple operating arrangements involving their own grains, those of landlords, and those where formal and informal partnerships exist. Like many survey organizations faced with cost containment and the desire to publish survey results more quickly, NASS has switched

to more extensive use of telephoning, including CATI to collect these data. Obtaining accurate responses by phone was considered a problem not only because of the detailed nature of these data, but also because the centralized state telephoning crews lack the familiarity with farm terms that a personal interviewer, recruited from the local farm community, might have.

The focus in the study was to estimate the measurement bias by treating the final reconciled response between the CATI and independent personal reinterview response as the "truth." In order to obtain "truth" measures, experienced supervisory field enumerators were used in reinterviewing approximately 1,000 farm operations for the 1986 December Agricultural Survey. This study of corn and soybean stocks in three states indicates that the difference in the CATI and final reconciled responses, the bias, was significant for all but one item (soybean stocks in Indiana). The direction of the bias indicates that the CATI data collection mode tends to underestimate stocks of corn and soybeans.

In the process of reconciliation, the reasons for differences were collected. Table 1 indicates that an overwhelming percent of differences, 41.1%, could be related to definitional problems (bias related discrepancies), and not those of simple response variance (random fluctuation). Examples of these definitional problems are rented bins or grain belonging to someone else not included, confusion with reporting government reserve grains, and bins on son's farm mistakenly included. Definitional discrepancies were the largest contributors to the large bias. In contrast, the differences due to rounding and estimating contributed little to the overall bias.

Table 1 Reason for Differences in CATI and Reinterview Responses for Corn Stocks in Minnesota - December 1987

Reason	Number	Percent of Total
Estimated/Rounding	28	31.1%
"Definitional"	37	41.1%
Other	25	27.8%
Total	90	100.0%

Source: Fecso and Pafford (1988).

The report suggests that the bias in the survey estimate generated from the CATI telephone sample might be reduced through a revised questionnaire design, improved training, or a shift in mode of data collection back to more personal interviews. Considering the constraints of time and budget, the change to additional personal interviews is unlikely. Thus, the alternative suggested was to use reinterview techniques to monitor this bias over time and to determine whether the bias has been reduced through improvement in

questionnaires and/or training. If large discrepancies continued the estimates for grain stocks could be adjusted for bias through a continuing reinterview program.

4.4 Replicated Reinterview Studies for Measuring Reliability

In (3), the reliability ratio, R, was defined as the ratio of the variance of the population of true values to the total variance, including measurement variance. Estimation of R typically involves obtaining replicated measurements on a subsample of units via a reinterview study. In order for the estimator in (5) to be unbiased for σ_ε^2, two assumptions must hold for the reinterview survey: (1) the mean and variance of the response error associated with the reinterview and the original interview are identical; i.e., $\varepsilon_{\alpha i} \sim (\mu_\varepsilon, \sigma_\varepsilon^2)$ for $\alpha = 1,2,$ and (2) the covariance between the errors ε_{1i} and ε_{2i} is zero.

Assumption (1) requires that the reinterview survey create the same general survey conditions that existed in the original interview, i.e., the same survey questions, type of interviewers, mode of interview, respondent rules, and so on. However, since reinterview surveys must necessarily be conducted at a later point in time, the survey questions may need modification to account for the time difference. Further, to save cost and respondent burden, only a subset of the original questionnaire may be used. If the original survey was conducted in-person, the reinterview may not be affordable unless it is conducted by a less expensive mode. Other changes in the reinterview procedures may be required to save costs. The consequences of these changes may affect the response error distribution thus invalidating assumption (1). In household surveys, this assumption has been well-tested and it appears that this condition is achievable in most situations (see, for example, Forsman and Schreiner, 1992, and Groves, 1989). For establishment surveys, reportd research on reinterview assumptions is scant, yet assumption (1) seems plausible for a well-designed establishment reinterview survey.

Assumption (2), however, is much more problematic. Errors between trials may be correlated due to a number of factors. Respondents may recall their original responses and simply repeat them in the reinterview without regard to their accuracy. If the error is recorded in the company's record system, the error will be replicated if the same record system is accessed in reinterview, or they may repeat the same response process to arrive at an answer. Thus, it is likely that the errors in the two responses are of similar magnitude and direction, inducing a positive correlation between the errors. For household surveys, O'Muircheartaigh (1992) reports correlations as high as 57 per cent in the CPS unemployment classification. However, little is known regarding the between trial correlation for surveys of establishments. Bushery, et.al. (1992) discusses the difficulties of implementing a reinterview by mail when the original survey was conducted by mail. In a survey of schools, they found evidence that respondents who kept a copy of their original responses transcribed these responses to the reinterview form.

It can be shown that, for a between trial correlation of τ, the relative bias in the estimate of the SRV in (5), is $-\tau$. Thus, a positive correlation results in a downward bias in

the estimates of reliability. More research is needed to study the magnitude and direction of τ for characteristics measured in establishment reinterview surveys.

4.5 Internal and External Consistency Studies

Two general techniques for assessing measurement error (at times in combination with other sources of NSE) are internal and external consistency checks. Internal consistency studies include analysis of edit output and analysis allowed by the use of rotating panel designs. External consistency studies consists of a comparison of the estimate with comparable estimates from another source, done most effectively when compared as a data series.

External Consistency Studies

The comparison of several series of estimates can provide an indication that measurement error may exist. If sampling error is small enough, then deviations in the profile of the series would indicate a problem with NSE. Similarly, a difference in the level between the series is an indication of NSE. If one series is considered very accurate and another series is parellel but at a different level, the difference is a crude estimate of overall bias. In crop yield surveys conducted in the U.S., time series produced by different measurement methods (counts from fields or various farmer reporting techniques) are often very parallel, but some level differences of 10 percent occur (Fecso, 1991). Unfortunately, determining the contribution to observed NSE due to measurement error is not possible without other sources of information. In an example of an external series comparison, Silverberg (1983) examined several of the many different series of estimates of retail gasoline prices. He found that all the national estimates of retail motor gasoline prices were within 5 percent of the EIA series, but sizable differences existed for wholesale prices in some series (although he speculates that the differences may be due to a difference in volume weighted price estimates versus a price based on a price index). Further, Silverberg found that internal assessments were more informative than external comparisons.

Internal Consistency Studies

Internal consistency studies are often exploratory in nature. Activities during editing may lead to the most familiar forms of internal consistency study. A detailed description of editing processes can be found in the chapter by Granquist (19940. In a study of U.S. federal establishment surveys (Federal Committee on Statistical Methodology, 1988), almost all of the surveys reviewed used practices such as analyst review of the data and edits for reasonableness with followup proceedures in efforts to control response error. About three-quarters of these surveys compute edit failure rates and about half compute interviewer error rates as an inderect technique to measure response error. The information about the relative magnitude of potential measurement error gleaned from analysis of these error rates and from feedback from respondents and interviewers can help target areas in need of more direct measurement using methods previously described.

For most surveys, these edit failure rates relate to procedures tending to be based on logical relationships among data items and experience with range checks to detect outliers. More statistically based procedures to identify unusual data are available. Such techniques can be useful in associating potential causes of measurement error with the errors. Barnett (1983), for example, discusses principles and statistically based methods, including multivariate approaches, for detecting outliers in data sets. Silverberg presents a graphical technique for examining the changes over time in the distribution of company reported prices of gasoline. Temporal comparison of response distributions is useful in detecting unexpected changes to conditions impacting response or, as Silverberg found, verifying expected changes such as greater price variability during the 1979 oil crisis.

Repeated use of the same establishment and data item in surveys conducted periodically provides an opportunity to measure validity and reliability of the data. Saris and Andrews (1991) and Munck (1991) discuss the use of structural modeling as an approach to evaluation of measurement errors. Reiser et. al. (1992) apply such models in a study of measurement error in a corn yield study in the U.S. Variables which were easy for the enumerator to count, such as number of stalks, were found to be highly reliable, but some measurements, such as length of the ear, were not.

In research on the use of historic data, methodology very similar to a reinterview approach, Pafford (1986) studied the effect of using farmer-reported planted acreage from a spring survey during the fall survey interview. Differences in planted acreage between the spring and fall survey were largest in the control treatment where prior responses were not used. When spring reported data was worded directly into the fall question, the differences were smallest. The control group had more instances of unreported acreage in either of the survey periods, while there were few reports of zero acreage when the spring report was nonzero and incorporated into the fall question. While interviewers had no problems using prior data and response variance between surveys decreased, a recommendation to use historic data could not be made because there was no way to determine which treatment had lower response bias. Similar results were found in a later study of grain stocks (Pafford, 1988).

4.6 Quality Standards

The establishment and use of standards are viewed by many as fundamental needs for a process of quality improvement. Standards reflect principles that are commonly held for guiding and assessing the quality of an activity. In a survey environment, standards are the guide to the process that keeps an organization focused on the essential conditions and procedures that could produce unwanted changes in measurements. Morganstein and Hansen (1990) indicate that without clearly stated standards, followed without compromise, increased and unnecessary variability will be a result. Areas where standards are believed to be useful in controlling measurement error include the design of the survey instrument, recruitment of and training of enumerators, supervision of interviews, and editing.

Freedman (1990) notes that standards alone cannot improve data quality. Enforcement, education, and evaluation are also necessary, as is the commitment from

everyone in the organization. Further, the organization must decide where standards are needed. In doing this, EIA used accuracy, consistency, efficiency, and clarity as guiding principles. Some standards apply regardless of the subject matter covered in the survey, for example pretesting all questionnaires to assure clear wording of questions and the respondents ability to understand and answer the question. Others depend upon the circumstances of the survey, such as the need for interviewers not to know the name of the organization sponsoring the survey when the data will be used as evidence in judicial actions (Morgan, 1990). Examples of organizations with very detailed standards for instrument design and collection activities include the Energy Information Administration (1989) and the National Center for Education Statistics (Cooperative Education and Data Collection and Reporting Standards Project Task Force, 1991). Many agencies do not have formally developed standards for data quality. Further, these studies suggest that the development of standards may often be the response by organizations in which data quality problems have escalated to a crisis.

EIA uses a quality audit to check for compliance with standards. The audits focus on determination of the soundness of the data system and its ability to produce data that meets standards (acceptable quality). Reviews and tests of survey processes are made and documentation is reviewed. Checklists are used to ensure coverage of all standards and to note deviations. Offices are held responsible for implementing the recommendations made by the standards review team when deficiencies are found. However, while enforcement may be necessary to ensure that the standards are followed as intended, there is some concern that auditors may be looked upon as the "quality police" of the organization and a resentment could build. For quality standards to work, a cooperative environment built upon trust, voluntary self-improvement, and an emphasis on organizational goals and missions is necessary. Such efforts are the emphasis of the growing movement from inspection sampling of the final product toward quality improvement strategies.

5 SUMMARY AND CONCLUSIONS

In this paper, we have attempted to provide an overview of the most important methods for evaluating measurement errors and a description of how they have been or could be applied to establishment surveys. We presented a number of models for measurement error which are useful for guiding efforts to improve survey data quality. We presented error evaluation methods, such as cognitve laboratory techniques, which may be applied during the design stages of a survey. Other methods presented, such as reinterviews, experimental designs, and observational studies, may be applied during the execution of the survey. However, the results of the latter evaluations are not known until after the survey is completed. Finally, post-survey methods such as internal and external validity studies and administrative record check studies were discussed in the context of establishment surveys. One area not addressed directly in this paper is evaluation methods that may be applied *during* data collection in order to provide timely information on the quality of the data as it is collected. The need for such methods is obvious. In order to be controlled, survey errors must first be measured and measured continuously during the course of the survey. When information on data quality is available on a "real time' basis, survey managers can

determine the degree to which errors are being controlled and can intervene if action is needed reduce the errors. Such methods are called *process quality control methods.*

It is quite unusual that a component of measurement error can be directly estimated and monitored continuously during data collection. What is usually the case is that some *indicator* of data quality - response rate, edit failure rate, item nonresponse rate, etc. - is measured and monitored. A quality indicator may be any quantity that can be routinely observed that is correlated with some component of error. As an example, interviewer edit failure rates may be correlated with interviewer variance or bias. The belief is that keeping the indicators within acceptable limits will also control the corresponding mean squared error component, resulting in data which are of higher quality and value. Unfortunately, there is no assurance of this and studies which directly measure bias and variance components are still needed to determine the true level of data quality. Further, many serious measurement errors can only be detected using direct approaches. In continuing surveys, evaluating the magnitudes of the measurement errors periodically (say once every two or three years) may be adequate to ensure that the routine quality control practices are controlling errors at the desired levels. For one time surveys, direct evaluation studies may be expensive but still necessary for two purposes: (1) to provide data users with an accurate assessment of the data's true information content and (2) to obtain information on survey error that can be used in future surveys. There is a vast literature on methods for quality improvement in industrial or service settings, but only a few articles specifically for survey operations (e.g. Fecso, 1989; Colledge and March, 1992).

While a variety of methods are available for the control of data quality, the use of the methodology varies. The Federal Committee on Statistical Methodology (1988) profiles the use of various measurements and control procedures for measurement error in federal establishment surveys. A majority of the surveys studied reported the use of editing, analyst review, and the production of edit failure rates. Useful techniques such as reinterviews, record check studies, and cognitive studies were reported as rarely used. In general there is a tendency to look at indirect measures rather than direct measures of error components. The lack of standard approaches to incorporating error measurement strategies into surveys may, in part, have stimulated the recent interest in *total quality management* (TQM) approaches in some survey organizations. Federal agencies are further motivated to use TQM approaches because recent administrations have encouraged such thinking government-wide.

A TQM strategy of management emphasizes a focus on continuous improvement in all aspects of an organization. The framework for such a strategy includes the following elements (Fecso, 1993):

- identify customers and their needs,
- move from only measuring quality to improving it,
- base decisions on the analyses of data,
- anticipate and accept change,
- emphasize an inter-organizational team approach to problem solving,

- ensure that staff are all aware and involved in the organization's quality goals, and
- accomplish the above activities with top management leadership.

While the survey research literature has many reports on the use of techniques associated with TQM in various situations or in units of a survey organization, the literature is devoid of studies about the effectiveness of the efforts as an strategy in survey organizations. One might expect this since most survey organizations appear to be in the development stages of implementation during which they focus on training and team formation. In subsequent stages, we would expect open discussions leading to agreement on the most important quality problems in the organization, further use of real-time measurements during the survey process, increasing communication of the results with those involved in the operations, and delegation of responsibility for taking actions to correct problems.

Quality profiles can be very helpful in quality improvement efforts, yet profiles which consist only of indirect measures of data quality can be quite misleading (Bailar, 1983). Use of the models and techniques described in this paper are also necessary for error evaluations and data quality improvement. The Federal Committee's report convincingly demonstrates the need for more frequent use of the evaluation methods describe in this paper with Federally sponsored establishment surveys. In the private sector were studies which directly evaluate the components of measurement error are seldom reported, the need is even greater.

ENSURING QUALITY IN U.S. AGRICULTURAL LIST FRAMES[1]

Cynthia Z.F. Clark, National Agricultural Statistics Service, Elizabeth Ann Vacca, Census Bureau
Cynthia Z.F. Clark, NASS, 14th & Independence Ave., Washington, D.C.

KEY WORDS: Business registers, record linkage, census and survey list frames

ABSTRACT

In the United States, agricultural data are collected by the Bureau of the Census in the Department of Commerce, and the National Agricultural Statistics Service in the Department of Agriculture. Both agencies are mandated by law to collect data on the agricultural economy. Title 13 of the United States Code (USC) requires the Census Bureau to conduct a quinquennial census of agriculture, providing detailed data on agricultural operations for each county. USC Title 7 requires the National Agricultural Statistics Service to collect data on agriculture and its market. The census of agriculture has mandatory data reporting requirements whereas surveys conducted by the NASS have voluntary participation.

Both agencies are required to protect the confidentiality of the individual record data but are subject to different legal requirements. Title 13 restricts access to census data to sworn officials and census employees, prohibiting the release of the census of agriculture list or census data to the NASS, or to any other organization. Title 7 permits the use of data collected by NASS for statistical purposes, and thus enables the Census Bureau to use the NASS list in building the census list. The NASS list, however, does not have complete coverage of the universe of farm operations. Thus, it is necessary for each agency to compile its own list frame to meet its mandated statistical needs.

Both U.S. agencies use the same definition of a farm - a place from which $1,000 or more of agricultural products were sold during a calendar year. A new census list is created every five years by linking agricultural statistical, administrative, and commodity lists of establishments. The NASS list was developed in the late 1970's from similar types of lists and is continually maintained. The overall quality of both census and survey data is dependent upon the quality of the list frames. The paper focuses on five attributes of list frames - scope, accuracy, duplication, coverage, and cost efficiency - relating to quality. List development procedures and quality indicators are discussed and compared.

1. AGRICULTURE CENSUS LIST FRAME

Since 1969 the census of agriculture has been conducted using a mail-out/mail-back data collection procedure in lieu of personal enumeration. Prior to each census, the Census Bureau assembles records of individuals, businesses, and organizations identified as having some association with agriculture. This includes files from the previous census, administrative records of the Internal Revenue Service (IRS) and the Social Security Administration (SSA), and statistical records of NASS. Additionally, lists are obtained for specialized operations (e.g. nurseries and greenhouses, specialty crop farms, poultry farms, fish farms, livestock farms, cattle feedlot operations, grazing permittees) from State and Federal government agencies, trade associations, and similar organizations. Lists of companies having multiple establishments (or locations) producing agricultural products are obtained from previous censuses and updated using the information from the Standard Statistical Establishment List maintained by the Census Bureau.

After the various address lists are acquired, the Census Bureau performs record linkage to remove duplicate addresses, screens for nonfarm records, and prepares mail labels for each address. Five major operations are required for record linkage: format and standardization, business or personal identification number linkage (Employer Identification Number, EIN, or Social Security Number, SSN), geographic coding and ZIP code verification, alphabetic name linkage, and clerical review of potential duplicate record sets. For the past four censuses, two similar phases of address linkage were conducted to permit incorporating more current addresses than would have been available using a single linkage.

The format and standardization operation places each source record into a common format; edits the source records; and assigns name control,

[1]This paper reports the general results of research undertaken by the staff of the Census Bureau and the National Agricultural Statistics Service. The views expressed are attributable to the authors and do not necessarily reflect those of the Census Bureau or the NASS.

processing codes and size codes to each record (Gaulden, 1990). The name control uses the first four characters of the primary name, containing a minimal of three non-numeric characters, not appearing in a special agricultural "skip list" dictionary. This dictionary contains over 1,000 words and abbreviations (such as Farm, Dairy, Bros) which could conceivably appear in the name field but were not likely to be the surname.

Processing codes facilitate the use of the most reliable information in the final list record. Initially, each record is assigned a name and address priority code based on the expected currency of the address information of the record source. During record linkage, the name and address of the record within a linked record set with the highest priority is retained. Each record is also assigned a size code based on the estimated total value of agricultural products (TVP) to be sold in the census year derived from agricultural data on the source record. During record linkage, the size code from each record in linked record sets is transferred from each of the deleted duplicate records to the retained record. This allowed for the derivation of both a "source combination code" indicating all the sources for the final record and a "final size code" based on the reliability of size information for each source.

Records are then linked by computer on EIN and SSN. All remaining unmatched and potential duplicate records have the addresses coded by geographic area, the names identified according to part, and the names coded using SOUNDEX procedures (where vowels and double consonants are deleted from names). The geographic coding system was designed to ensure that all records contained standardized and edited geographic codes, prior to record linkage. The name and address linkage procedure adapted the Fellegi-Sunter probability model relying on the extent of agreement between name, address, and other record identifiers to determine duplication within ZIP code (or area) blocks. An additional linkage of historically identified duplicate records is conducted. Duplicate records are deleted, retaining the record whose source is deemed to have the highest quality address information. Potential duplicate records are reviewed clerically to determine which duplicate records to delete.

2. THE NASS LIST FRAME

The NASS conducts ongoing monthly, quarterly, and annual probability based sample surveys of the agricultural sector. The samples are selected from the list frame and a land-based area frame. The list frame consists of a frame for each state that is maintained and updated by each NASS State Office using procedures developed by a NASS headquarters unit. The NASS list frame is designed to provide good coverage of large and commodity specific operations. Because NASS surveys use dual frame estimates derived from the list frame and a complete area frame, it is not necessary to achieve farm universe coverage with the list frame.

Prior to the implementation of probability surveys, NASS maintained "reporter lists" of farm operators who were willing to provide their agricultural data to the state agricultural statistical office. The quality of these state lists varied considerably. In the mid-70's the agency undertook a large effort to develop a list that would provide the frame for sample surveys with standard address and agricultural data across states for each farm operation record. To build the new NASS list, each state secured lists of potential agricultural operations to use in conjunction with the state reporter list. Lists were selected based on a composite source evaluation that weighted such factors as the degree of coverage, the frequency of updates, the type of update procedure, the list medium, the identifiers on the list, the agricultural data on the list, the use of the list, the number of records, and the cost. Lists came from such sources as the Agricultural Stabilization and Conservation Service (ASCS), the Rural Electrification Administration, and agricultural trade associations. However, regulations prohibited the release of tax records of the IRS and farm employer records of the SSA to NASS.

Similar to the census procedures, the records for each state were formatted and linked removing duplicate records. The linkage system first matched on SSN, EIN, telephone number, and two other identifiers. Surnames were coded using the New York State Identification and Intelligence Service (NYSIIS) procedure, a coding procedure that NASS research found superior to SOUNDEX for blocking (Lynch). Place names were used to assign latitude and longitude to each record for construction of a distance function for matching place names within the NYSIIS blocks. The linkage system used an adaption of the Fellegi-Sunter model that relied on the frequency of occurrence of name, address, and identifier components, and specified types of errors within the file, to determine duplication within NYSIIS blocks for each ownership type. Records determined to be probable duplicates were linked with one record being placed on the frame. An extensive clerical review of the identified potential duplicate records was conducted and the duplicates were removed from the list. Agricultural data (referred to as control data), compiled from existing farm records and

special surveys, was appended to the address record to use for sample stratification.

Each NASS state office has the responsibility for maintaining the frame for its state. Each office determines how it will update name and address information and control data, add records for new farm operators, and identify records of individuals that no longer have farm operations. The headquarters office supports the state offices by developing and maintaining computer programs and procedures to assist in this process. An interactive system has been developed for updating record information in each office. Each state prepares an annual plan for list frame development and maintenance taking into account evaluation of the state list frame and state specific survey needs. The headquarters list frame group reviews, advises, and supports the plan.

When a state wants to add ASCS or other source records to its state list frame, a match is first conducted using SSN and EIN (if available on the records). Matched records are retained with the NASS name and address and control data from all sources. At the option of each state office, nonmatching records can be run through a resolution program. This program replicates the Fellegi-Sunter process used in the original linkage with the exception that records identified as probable matches are retained as separate records on the frame unless clerically deleted. The resulting nonmatching records are added to the state list as inactive records. These inactive records become the source for "criteria surveys" (described in Section 3.3), clerical review, or personal enumeration to determine their farm status. Records for valid farm operators are placed on the active farm frame.

3. SCOPE OF THE LIST FRAMES

Both agencies face the challenge of determining the scope of the list frame by focusing on its size and composition. The size of the frame impacts the cost of data collection and processing, and the respondent burden (directly for the census and indirectly, through the efficiency of the sample design, for NASS). The frame composition affects the resulting quality of the survey data. Records from the same source often have common characteristics impacting record quality. If the frame contains records that are thought to represent farm operations but, in fact, do not have an associated operation, the integrity of the resulting data is compromised. Nonfarm operators often do not respond to a request for agricultural information. The statistical procedures for either a census or survey need to adjust for such nonresponse.

Response to the 1987 Census of Agriculture illustrates the impact of the scope of the frame on the census operation. Of the 83.1 percent responding to the 1987 census, 54 percent were farm operators, 44 percent were not farm operator addresses, and 2 percent were unclassifiable. Of the 13.3 percent not responding to the census an estimated 44 percent were farm operators. A sample survey with its corresponding sampling variability was used to account for the nonresponding farm operations in the final census data. The remaining 3.6 percent had undeliverable addresses, adding to the operation costs. This section discusses how the two organizations determine the scope of their respective frames.

3.1 Size of the List Frames

For the census, the final list becomes the address list for the mail enumeration. All records not retained in the list are placed in inactive files and are excluded from data collection. For the NASS, the final list is the sampling frame for surveys, using agricultural data residing on each record for stratification.

Cost and burden considerations severely limited the size of the 1992 census mailings -- 3.55 million records. The best strategy for list compilation with such constraints appeared to be to limit the number of less reliable input addresses, resulting in 12.4 million source records. After linking records identified as duplicates by name, address, SSN, and EIN, 4.9 million records remained at the end of the second linkage phase. Of these records, approximately 1.1 million came exclusively from nonfarm past census or NASS sources and were deleted from the mail list. Statistical classification analysis was effectively used to identify 230,000 records as least likely to represent farm operations and candidates for potential removal from the list. After minor modifications, the final census list contained 3.55 million records.

The 1992 NASS list frame consisted of 1.74 million records that were active farms or agribusinesses, a three percent increase over 1991. Additionally, there were 1.66 million inactive records. In 1991, there were 763,930 inactive records on the frame, known to be either nonfarms or out-of-business agricultural establishments. Some of these records came from new sources and their farm status had not yet been resolved; others were being retained on the list as nonfarms to aid in future identification of the record farm status during the linkage and resolution process.

3.2 Sources of List Addresses

Because of the census focus, concerted efforts

are made to include in the list compilation all important sources of agricultural record information. A two stage linkage process is used to permit the incorporation of IRS records from the two years prior to the census. However no IRS births are picked up for the year of the census. Two notable changes in list sources and their content were made for the previous two censuses. For these censuses, the Census Bureau used the agricultural farm operation list of the NASS for all 50 states as contrasted with the 31 states available for use in the 1982 census list compilation. The NASS list, provided to the Census Bureau prior to each linkage stage, enables the census list to include new records and updates to existing records as of April of the census year.

In 1992, the Census Bureau did not use all ASCS records directly as had been done in the 1978 and 1982 censuses. This decision was based on the expected inclusion of valid ASCS farms into the NASS lists, the ASCS list size, and the reliability of the ASCS name and addresses. When the ASCS records were used as a census source many ASCS records did not match other source records and required screening (25 percent of the 3.0 million records in the 1982 Farm and Ranch Identification Survey were obtained from ASCS only). Many of the addressees were determined to be landlords rather than farm operators. Of the 1982 census farm respondents, only 1.4 percent came uniquely from the ASCS list.

Following the 1987 census, tabulations of the mail list by address sources provided the contribution of each source to the list. The percent of enumerated census farm records appearing on only one source was 14.9 for IRS, 3.9 for the previous census, and 3.3 for NASS. The percent of unique farms among the respondents was much higher from the IRS list (10.8) and the special lists (8.7) than for the other source lists (3.9 for 82 census list and 4.1 for NASS list).

The NASS primarily uses information from its own surveys and data collections to update record identifiers and data. It obtains specialized lists from trade associations and other agricultural organizations to extend list frame coverage. These are as diverse as lists of farm vehicles, general livestock, cattle, hogs, sheep, equine, field crops, poultry producers, pesticide applicators, fruit growers, vegetable growers, nut tree producers, bee and honey producers, floriculture and nurseries, wool producers, agricultural labor employers, and other specialty commodity producers. Since the agency received funding for the Pesticide Data Program and the Water Quality Program in the early 90's and began surveys of chemical usage on vegetable, fruit, nut, and field crops, NASS has devoted focused efforts in building a list that had much better coverage of fruit, nut, and vegetable operations.

Recently NASS has worked with ASCS to investigate better use of ASCS records for list building purposes. The NASS tested a procedure that identified approximately 43 percent of ASCS crop records not matching NASS records as unlikely farm operations. In the past a large proportion of the ASCS nonmatch records have been out-of-business or deceased farm operators or landlords. Without a method to identify the likely nonfarm records, NASS use of ASCS data files has resulted in large numbers of records whose farm status needed to be resolved with a resulting low yield of farm records. Additionally, NASS has arranged with ASCS to have direct access to the ASCS records using a relational database. This will enable NASS to extract only those records with specified characteristics and only the items on each record that are of use to NASS. Computer processing costs are expected to decrease, and list building procedures are anticipated to be more effective at increasing the coverage and quality of the NASS list frame.

The NASS plans to provide each state with annual listings of ASCS large or specialty operations that do not match NASS records. Operations producing commodities with insufficient present records for targeted sample designs (such as those new in the estimating program) would be selected. General matches of the entire state ASCS data files to the NASS list would be conducted only once every three years, rotating states each year. The screening procedure would be used to identify the nonmatching records that would be extracted. This approach should result in a more effective use of NASS and ASCS resources to produce higher quality NASS list information. This is an important consideration as determination of actual farm status of potential list records is a time consuming and resource intensive effort.

3.3 List Frame Farm Operator Composition

Although both organizations build their lists by acquiring lists of addresses associated with agriculture, this does not ensure that each record represents a farm operation. For both organizations, the identifier and agricultural data available for each list address does not generally provide adequate information to determine whether or not an address represents a farm operation. To use the list frame effectively for either a census or survey, farm status needs to be identified for each list record. Mail or telephone surveys, personal enumeration or follow-up, or statistical modeling have been used to accomplish this objective with input records. The accuracy of these procedures affects the

size of the census mailing and the efficiency of the NASS sample designs.

Prior to both the 1978 and 1982 censuses, the Census Bureau mailed the Farm and Ranch Identification Survey to approximately 3.0 million addresses that had questionable farm status or were potentially duplicate addresses. The report form contained a set of initial questions which, if answered as "no", allowed addressees to skip the remaining questions. The final 1982 census mailing consisted of 3.65 million addresses of which 1.2 million were respondents to the identification survey classified as representing potential farms and .5 million were survey nonrespondents.

In both the 1987 and 1992 censuses, the Census Bureau used statistical classification analysis to aid in the identification and removal of nonfarm addresses from the list. Census classification analysis is a nonparametric method where previous census record characteristics such as the source of the mail list address, number of source lists on which the address appeared, expected value of agricultural sales, geographic location, and past census farm status were used to separate records into groups according to the proportion of expected census farms and build prediction models. The models were then applied to the 1992 list records and provided an estimate of the probability that a current mail list addressee with that group's characteristics operated a farm (Owens, 1989). The groups of addresses least likely to represent farms (farm probability less than 18 percent for the 1992 census) were removed from the mail list.

An evaluation of the 1987 classification tree analysis (Schmehl, 1990), included a sample survey of records dropped from the census list, belonging to model groups whose proportion of farms was expected to be 11.7 percent or less. Approximately 14.6 percent of these records represented farm operations (46.4 percent survey response), or approximately 25,500 of the 175,000 records dropped represented farm operations. From a separate coverage evaluation program, an estimated 242,850 farms were not on the final census list. Using the two estimates, about 10 percent of the farms not on the mail list (25,500 of the 242,850) were on the preliminary list and the remaining 90 percent were either on the list of nonfarm addresses which were dropped or not on the list at all. This evaluation and others indicated that the analysis accomplished the objective of identifying the groups of records with questionable farm status and provided a reasonably good estimate of the impact of reducing the size of the final mail list on census coverage. For the 1992 census, refinements were made to the model, including the use of standardized computer software and the production of unique state (rather than regional) model groups.

Most NASS State Offices use an annual mail or telephone criteria survey to collect data on operation identification and commodity production. The universe, frequency, and timing of this survey varies considerably from state to state. Field work conducted by personal enumerators often supplement or reconcile the mail and telephone data collection. This survey is used for addressees gleaned from lists acquired by the office and for active farm records that do not have current survey information. New farm records are placed on the active state sampling frame with their agricultural (control) data, and information and data are update for existing records. Nonfarm records are retained on the inactive list frame with a status code that distinguishes records of deceased, retired, or out-of-business operators, of landlords, and of those without agricultural activity.

Annually NASS estimates the number of active list records that do not currently represent farms. This estimate is the difference between the number of active farm records and the sample expansion of the enumerated units in the June Agricultural Survey area sample that are also on the list. In 1992 this estimate was 411,924 or 26 percent of the farm records compared with 474,446 or 28 percent in 1989. Additionally, the NASS measures the percent of sampled addresses in the Quarterly Agricultural Survey that were identified as nonfarms - 10.5 percent in 1992 as contrasted with 13 percent in 1989. This is an estimate of nonfarms on the active farm list frame that were identified as having the commodities of interest.

4. ACCURACY OF RECORDS

The accuracy of both the address and data contained on each list record is critical for a quality census or survey data collection. The accuracy of the address information affects the ability of the survey organization to locate the addressee. The accuracy of the agricultural or control data influences the efficiency of the Census and NASS survey designs. If cases are included in samples inappropriately, the cost and quality of the data collection is compromised.

4.1 Procedures Affecting Accuracy

Specific procedures are used in census mail list development to increase the accuracy of record information (Gaulden, 1990). The name control standardizes the format for each name on the list for comparisons and is essential for identifying potential duplicates during the initial linkage on identification number. The processing code facilitates the use of the

most reliable information (both address and data) in the final record. Both source and size codes are important for sample stratification, classification analysis, census processing, and for evaluating the census mail list.

After every survey, NASS state offices make on-line name and address changes and update the record status code used for effective sampling and data collection. Active record status codes identify refusals, records for special handling, operators linked to additional operations, etc. Inactive record status codes identify deceased, retired, or out-of-business operators, landlords, partner of primary operator (with linkage specified), multiple operations, etc. Inactive source codes such as ASCS, list frame, criteria survey, and other data surveys are retained with the agriculture data. The NASS survey data is captured into a file that is held until the annual classification period. At that time all data for a given record is compiled and the "best" value (generally the largest current year value) is selected by a ranking process as control data for the record. States often specifically extract records without control data for major data items for inclusion in criteria surveys.

4.2 Measures of Accuracy

Several measures from the census data collection permit an assessment of the accuracy of address information. One such measure is the number of forms that are undeliverable as addressed (UAAs) - addressee is deceased, name or address has changed, or another situation is indicated. In 1987 there were 148,252 UAAs (3.6 percent of the census mailout). This compared with a lesser number of 82,792 UAAs (2.3 percent) in 1982 where, 446,000 UAAs had been previously removed from the preliminary 1982 list using information derived from the Farm and Ranch Identification Survey.

Another indication of the accuracy of 1987 census address information was obtained from results from an intensive follow-up to the Nonresponse Survey -- a survey conducted prior to completion of the census data collection for each state to estimate the percent of farms among the nonrespondents. The follow-up examined a sample of 1,263 nonrespondents to the survey. In early 1989, certified mailings, telephone and personal enumeration contacts were conducted on these cases to obtain further information about the nonrespondents. A total of 30.2 percent of the sample cases were UAAs. Of the 31.2 percent for which personal follow-up was attempted, over half had address changes before or during the census data collection. Either the corrected address information had not been received from the Post Office as requested or the Census Bureau had not successfully incorporated the address change into its mail list.

The expected sales code is an extremely important data item on both the census and NASS lists as it is used for sample stratification. Tabulations of expected sales code on the census address record by actual total value of agricultural products sold (TVP) in the census provides an indication of the accuracy of this code. In 1987, approximately 34 percent of all list records had the same coded value for TVP on the census as on the mail list; 72 percent had a value within one size code in either direction.

The NASS computes the correlations of control data (list frame data) with actual survey values for all data items used for sample stratification. In 1992, the U.S. correlations ranged from .766 to .917 for these items with hog, cattle, and land in farms data more closely correlated than cropland and storage capacity. The NASS also measured the percentage of these data items updated in each of the previous six years. Between 51.2 and 69.3 percent of these data items on records selected for at least one survey have been updated in the past two years. States are provided with estimates of percent of the records sampled on the active frame for which these data items are more than five years old. State offices are advised to review the timing of list frame updating and their data selecting and capturing procedures if they have correlations below .40 or if their state has a high percentage of data for any of these items that is more than five years old.

5. DUPLICATION OF LIST OPERATIONS

In the census, list duplication leads to the potential for duplicate census enumerations. For the NASS, list duplication results in incorrect sample weights. For both agencies, the primary objective of list computer and clerical linkage rules is to increase the accuracy of the matching procedure. However, there is a delicate balance between eliminating list duplication and maintaining list coverage. The census computer and clerical procedures have been designed to identify almost exact matches and have relied to a large degree on self-identification of duplicate census report forms, thus increasing coverage with the accompanying risk of increased duplication. The NASS assumptions during list building focused on reducing duplication; the list resolution process for updating and maintenance focuses on increasing coverage. Removing list duplication presents operational challenges for both organizations.

Recently, both agencies have faced an additional challenge to have current and nonduplicated addresses on the list frame. Rural route and box addresses are being changed to street addresses to

facilitate finding addresses for emergencies. If no procedure is instituted in a computer matching process to detect this situation, multiple records for an addressee may be retained on the list frame.

5.1 Procedures Affecting Duplication

In the agricultural universe, farm or ranch operators often have multiple operations with an identified operator participating in one or more partnerships as well as an individual operation. The NASS uses a cross-referencing system of list frame identification numbers. The census mail list compilation flags historically identified operations with different names as possible partnerships or corporation (PPC records). A PPC record flag is used to prevent automatic computer deletion of records as duplicates, causing all paired addresses to be clerically reviewed. During the preparation of the 1992 census mail list, a telephone enumeration of selected PPC addresses was conducted to determine linkage and reduce duplication on the mail list. Of the 25,000 record sets contacted (a total of 107,820 records), 45,464 records were deleted as duplicates.

Recently the sampling unit on the NASS list frame was changed from farm operations to farm operators in order to target a unique unit. This change was implemented by first matching the NASS list against itself using the Fellegi-Sunter procedure to identify matches and probable matches. The matches were identified as inactive records with linkage to the active record. Each state office clerically reviewed all probable matches, determining the name and address of the operator. Operations with multiple operators had the primary operator identified on the file. Additionally, multiple operations for one operator were identified and were linked with that operator.

The Census Bureau introduced a probability based linkage system (variant of the Fellegi-Sunter procedure) for name and address matching in 1992. A modification of this system was used to match 1990 Post Enumeration Survey records to the 1990 Decennial Census. This system permits the user to specify the degree of certainty (threshold values) desired for matched records. Eight percent more duplicate addresses were identified during the first linkage phase; one percent more were identified in the second phase. The evaluation is not yet complete that will provide the percent of increase in duplicate identification attributable to the new linkage system.

The NASS had developed a list duplication check program matching records on numeric fields such as SSN, EIN, and telephone number whose primary use was to identify duplicates after samples were selected from the list frame. Then the probability of selection of each sampled unit was adjusted for list duplication. This program is now additionally used prior to selecting list records from those classified for particular samples. Those identified potential duplicates that are sampled are then flagged for review during the survey edit process.

The Census Bureau used several new (or previously used) procedures to help identify duplicates during data review and processing of the 1992 census. To facilitate self-identification of duplicate report forms, the Census Bureau provided instructions to the respondent on the form and information sheet. A statement was added to the 1992 census envelope to remind the respondent to return all duplicate report forms in the envelope. A duplicate search operation was conducted during data review in all states, sorting all records alphabetically within counties and matching on telephone number and important data variables. Although telephone numbers were first available in the 1987 keyed data, telephone number matching was only used during processing in a few states.

5.2 Measures of Duplication

The census coverage evaluation program measures error in report form farm classification and in list duplication as well as farms not on the mail list. Nonfarms classified as farms and duplicate operations contribute to overcounted farms in the census. The total number of 1987 estimated overcounted farms (135,600) was very similar to the 1982 estimated number (113,623). However, the proportion of the overcounted farms that represented duplicate operations changed from 17 percent in 1982 (19,062 farms) to 47 percent in 1987 (63,290 farms). This increase in 1987 was primarily attributed to the lack of a precensus screening survey.

6. UNIVERSE COVERAGE OF LIST FRAME

The objective of a quality list frame for censuses or surveys is to provide as complete coverage as possible of the target universe. This is an extremely difficult goal for both NASS and the Census Bureau because of the impossibility of identifying operations in the target universe for inclusion in the list frame. Because the census of agriculture list is the basis for a mail enumeration, the overall coverage of the frame and the accuracy of the record information are extremely important. For the NASS, complete frame coverage of the universe of farm operators is not as important since a land based area frame is used to estimate for list incompleteness. Neither is the accuracy of address

information an overriding consideration for NASS as many contacts are initially made through a personal or telephone enumerator rather than a mail delivery person. Both organizations have formal means to evaluate the coverage of their list frame.

6.1 Procedures Affecting Coverage

The NASS has allocated more resources in recent years to increase the coverage of its list frame. The primary mechanism for this has been to use criteria surveys with ASCS records not matching the NASS list frame. As previously indicated ASCS list size constraints have limited the effectiveness of this approach without the ability to screen ASCS records. The NASS has not been able to experience very large increases in its list coverage with this approach. New procedures (described in Section 3.1) have been developed that hold promise for increases in coverage by more effective use of NASS resources.

The Census Bureau has conducted formal coverage evaluation programs for each census of agriculture since 1945. The program measures the accuracy and completeness of farm counts and selected data items and seeks to identify situations that lead to coverage error and to reveal data deficiencies and problems associated with census processes. The evaluation is conducted using an independent sample selected from the list to measure classification and list (duplication) error and the NASS June Agricultural Survey (JAS) to estimate the number of address records not on the census list.

6.2 Measures of Coverage

The much more rigid 1987 and 1992 census size constraints and restrictions on precensus screening necessitated that the resulting list be smaller, yet have a higher proportion of farm operations to ensure good coverage. Data from the census coverage evaluation programs indicate that coverage of the census farm universe is not complete. However, coverage of agricultural production has historically been above 95 percent for all censuses. Historical coverage estimates show that net farm coverage of actual farms has ranged from 85.0 to 92.8 percent for all of these censuses except 1978 where using both a list and area frame census achieved a coverage of 96.6 percent. This methodology substantially improved the state and U.S. level coverage of the census, particularly for farms with sales of less than $2,500 where the census enumeration is least complete. In 1978, the percent of these farms not included in the census was 6.5 percent compared with 28.6 percent in 1982 and 32.3 percent in 1987. Budget constraints have not permitted the use of dual frame methodology in subsequent censuses.

The coverage evaluation program for the 1987 census of agriculture was enhanced to provide estimates of farms not on the census mail list at the state level as well as more reliable estimates of incorrectly classified operations and duplicate operations on the census mail list. The percent of estimated farms on the census mail list was approximately the same in 1987 as in 1982 -- 89.2 percent contrasted with 89.4 percent. An estimated 98.6 percent of the land in farms, 92.3 percent of the crop farms, and 87 percent of the livestock farms were on the 1987 census list. Regional estimates of percent of census published farms not on the mail list are also produced, indicating that the list provides the most complete coverage for the Midwest region and the least complete for the Northeast region.

The reduction in size of the total mail list and the lack of a screening survey did not adversely affect the coverage of the mail list. The changes in the source lists for the 1987 census, improvements in the quality of the source records, and the effectiveness of the classification analysis contributed to maintaining the previous level of list coverage despite the drastic reduction in total list records included in the census. The estimated number of duplicate operations on the 1987 list substantially increased over 1982 -- 2.8 percent contrasted with .8 percent. No 1987 census procedures proved as effective in removing list duplication as the precensus Farm and Ranch Identification Survey.

The NASS has used the area sample from its JAS to estimate universe values for number of farms (by types of crop and livestock) and land in farms since 1985. From this sample it is possible to estimate the coverage of the NASS list for important data items. Studies in 1992, estimated 56.3 percent of farms, 77.6 percent of land in farms, 57.8 percent of crop farms, 53.6 percent of livestock farms, and 37.6 percent of specialty farms were covered by the active records on the NASS list frame. Estimates of the percent coverage by the NASS list for 1990, 1991, and 1992 demonstrate a gradual increase in coverage for these data items -- coverage of number of farms increased 2.6 percent, land in farms increased 3.1 percent, crop farms increased 6.1 percent, livestock farms increased 2.2 percent and specialty farms increased 5.1 percent. As with the census list, the NASS list coverage varies by region.

7. FRAME COMPILATION COST

The costs of compiling an agricultural list frame include salaries of professional staff who design

and implement procedures; procurement of source records; computer processing, linkage, and geocoding of those records; cost of screening list addresses to determine farm status; and salaries and travel costs for field or clerical staff reviewing address information to determine status and potential duplicates. Deriving separate costs for list frame building and maintenance from other survey costs is difficult.

A large proportion of the cost is associated with salaries for professional staff. At the Census Bureau a high level of work occurs for the development, implementation, and evaluation of the mail list during three years of the census cycle. Planning and research occur in the remaining two years. Professional staff work on the list requires a minimum of three statisticians during the entire cycle and three computer programmers during the three year development, implementation, and evaluation period.

At the NASS, list building and maintenance is an ongoing program. A staff of eight people located in NASS headquarters is responsible for developing computer procedures for matching lists and selecting list samples, for developing procedures for maintaining and updating record information, for evaluating the completeness and effectiveness of the frame, and for assisting state office users through documentation, training and consultation. One to two staff in each of the 45 State Offices have on-going state list building and maintenance responsibilities, with other staff used as needed. Extensive computer support has been required for design of new list frame systems.

The Census Bureau obtains source records for minimal cost as they are output of other statistical and administrative data file preparations rather than separate data collection costs. For example, in 1987 the Census Bureau paid the NASS $30,000 for its 2.4 million records and IRS approximately $125,000 for its 6.0 million records. The NASS receives the ASCS list and most of its commodity and trade association lists at no cost. Any costs are routinely associated with the cost of file preparation. Often the most significant costs associated with using another list source are additional programming resources required to standardize the formats of different lists. As programming resources are scarce at both agencies, additional list sources are added selectively if the list format does not meet agency requirements.

The number of records used in the census mail list linkage or incorporated into NASS list building efforts affects the overall cost of computer linkage and resulting staff costs for clerical review or field follow-up. The number of records was 1.1 million (8 percent) less for the 1992 census than for the 1987 census, and 5.4 million (30 percent) less in 1987 than in 1982. The arrangements that NASS has recently developed with ASCS will reduce the number of records in the selected files to be matched to NASS records and the resulting nonmatches. Both lists are computer processed on mainframe computers with their associated costs and overheads. The Census Bureau estimates of these costs are not tracked separately but included in the overall cost of census processing. The NASS has separate cost estimates for the headquarters and state office mainframe processing. The headquarters costs for list maintenance are approximately $450,000 per year, but the state costs are not easily separable from other survey processing costs.

The computer record linkage rules for both organizations are designed to avoid computer deletion of potential duplicates unless there is a high degree of certainty that the potential duplicates are matches. Less stringent matching rules could decrease the number of potential duplicate sets provided for census clerical review, thus reducing clerical staff costs. An indication of staff costs for this clerical work is provided by the number of potential duplicate sets of census records prepared for review during the first phase of linkage--573,148 in 1992, 767,448 sets in 1987, and 1,332,000 sets in 1982. The Census Bureau and NASS, in its initial list building, controlled the cost of this review by selectively setting the parameters for designating records as potential duplicates. At NASS clerical staff and enumerators are used to follow-up on potential farm operations resulting from nonmatching ASCS or commodity list records using mailed criteria surveys and telephone or personal contact. Costs are substantially lessened when mail or telephone contacts are used or personal follow-up is employed in conjunction with other data collection efforts.

The census classification analysis with its associated computer and professional salary costs was an inexpensive substitute for the much more costly precensus Farm and Ranch Identification Survey. However, the application was only designed to remove nonfarm addresses from the list. It did not accomplish the other two objectives of the screening survey - to obtain more current address information and to identify duplicate operations. The costs of such an independent data collection are relatively high. Although the costs of any survey is affected by the size of the survey mailing (3.0 million in 1982), the marginal cost of additional survey cases is small with a large scale data collection.

8. SUMMARY

Building a high quality list frame for large scale data collection is a difficult task. It requires an

ongoing program of research, evaluation, and development whether or not the list is maintained and updated periodically as the NASS list is or recreated cyclically as the census of agriculture list. Changes in source records and postal delivery procedures affect the list compilation process, requiring new list techniques. The purpose for which the list is intended is an important factor in determining the quality requirements for the list.

Compiling and maintaining a list of agricultural operations to conduct either censuses or surveys has a unique set of challenges. An agricultural list frame is unlikely to ever have a complete list of farm operations due to the high turnover in agricultural operations. Research following the 1982 census determined that only 71 percent of farms in 1978 were farms in 1982. Maintaining a high level of list coverage requires continual attention to improvements. In the U.S., tax records are essential for achieving a high level of list coverage, with this source uniquely providing approximately 14.9 percent of all identified farm operations in 1987. This is an important source for identifying new operations and those that have gone out-of-business. The NASS uses a number of different sources and procedures to try to accomplish the same objective.

The many different arrangements under which farms and ranches operate will invariably affect duplication in the list. As the Census Bureau discovered in 1987, controls to eliminate duplicate enumeration in the overall census processing system were lacking once the precensus screening survey was eliminated. Several procedures were initiated in 1992 to identify duplicate addresses and reports. This emphasizes the importance of continually assessing changes in methodology intended to increase quality of one aspect of the list in relation to the impact on other quality attributes.

In order to either build or maintain a high quality agricultural list frame, continual evaluation and measurement of the frame characteristics discussed in this paper will be needed. The attributes of the list -- its scope, accuracy of record information, duplication of records, universe coverage of list frame, and cost of list frame compilation -- will need to be reviewed in relation to the program objectives that the list frame serves. Although both census and NASS require a high quality agricultural list frame, the differing program objectives of these frames affect the importance of each of these attributes to the overall quality and functionality of the frame. The primary objectives of the census of agriculture list are farm coverage and uniqueness whereas the NASS objective is to obtain a high level of commodity coverage. These objectives affect the assumptions underlying the list frame procedures described in this paper.

REFERENCES

Anderson, Carter D. (1993), "NASS Long Range Plan for Use of ASCS Data, 1993 - 1997", NASS Internal Staff Report.

Arends, William L. (1977), "Methodology for the Development, Management and Use of a General Purpose List Sampling Frame", NASS Internal Staff Report.

Coulter, Richard and James W. Mergerson (1977), "An Application of a Record Linkage Theory in Constructing a List Sampling Frame", *Proceedings of Computer Science and Statistics: Tenth Annual Symposium on the Interface*, Gaithersburg, MD.

Fellegi, Ivan P. and Alan B. Sunter (1969), "A Theory for Record Linkage", *Journal of American Statistical Association*, pp. 1183-1210, December 1969.

Gaulden, Tommy W. (1990), "Development of the 1987 Census of Agriculture Mail List," U.S. Census Bureau internal report.

Geuder, Jeffrey (1992), "1992 NASS List Frame Evaluation", NASS Survey Management Division Report, eighth of a series.

Lynch, Billy T. and William L. Arends (1977), "Selection of a Surname Coding Procedure for the NASS Record Linkage System", NASS Research Division Report.

Owens, Dedrick, Ruth Ann Killion, Magdalena Ramos, Richard Schmehl (1989), "Classification Tree Methodology for Census Mail List Development," *1989 Proceedings of the Section on Survey Research Methods*, American Statistical Association.

Schmehl, Richard, Magdalena Ramos (1990), "Evaluation of Classification Tree Methodology for Census Mail List Development," *1990 Proceedings of the Section on Survey Research Methods*, American Statistical Association.

1987 Census of Agriculture Coverage Evaluation Report, Volume 2, Part 2.

Scope and Methods of the Statistical Reporting Service (1983), NASS Miscellaneous Publication No. 1308, Washington, D.C.

THE AREA FRAME: A SAMPLING BASE FOR ESTABLISHMENT SURVEYS

Jeffrey Bush, Carol House, National Agricultural Statistics Service
Jeffrey Bush, NASS, Room 301, 3251 Old Lee Highway, Fairfax, Virginia 22030

KEY WORDS: Area frame, stratification, primary sampling unit

Area frame methodology has formed one of the cornerstones of probability sampling for several decades. While area frames are frequently used in urban settings for household surveys and population censuses, those with a rural focus have proved valuable for targeting farm establishments to provide basic statistics on agriculture and ecological resources. Cotter and Nealon outline the advantages and disadvantages of area frame methodology. They state that area frames are highly versatile sampling frames providing statistically sound estimates based on complete coverage of land area. Although costly to build they are generally slow to become outdated. However, area frame sampling is generally less efficient than list sampling for targeting any individual item and is inadequate for estimating rare populations.

This paper describes four different area frame methodologies currently in use as a base for sampling in rural areas. These are: a) the area frame used by the United States Department of Agriculture's (USDA) National Agricultural Statistics Service; b) the area frame used by Statistics Canada for agricultural statistics; c) the hexagonal area frame used by the U.S. Environmental Protection Agency; and d) the area frame used by the USDA's Soil Conservation Service for the National Inventory Survey. The paper's greatest emphasis is on the NASS area frame. For this frame, the authors provide additional detail on area frame construction, sampling, data collection and estimation. The paper provides a profile of costs associated with these activities, as well as procedures to assess quality deterioration in an "aging" frame.

NASS AREA FRAME

The National Agricultural Statistics Service (NASS) is the major data collector for the U.S. Department of Agriculture. As such it has responsibility to provide timely and accurate estimates of crop acreages, livestock inventories, farm expenditures, farm labor and similar agricultural items. NASS also provides statistical and data collection services to other Federal and State agencies. They have used an area sampling frame extensively for over 30 years in the pursuit of these objectives. Area frame samples are used alone and in combination with list samples (multiple frame). NASS contacts approximately 50,000 farm establishments each year through their area frame sampling procedures.

This section updates the work of Cotter and Nealon as it describes the procedures used by NASS to construct area frames and sample from them. It discusses data collection procedures, estimators, and costs associated with these different activities. Finally it discusses methods to objectively assess the "aging" of an area frame.

Area Frame Construction and Sampling

NASS constructs area frames separately by state and maintains one for every state except Alaska. Generally, two new frames are constructed each year to replace out-dated ones. The most recent frame construction was for Oklahoma. It became operational in June 1993. This frame will be used as the "example" throughout this paper.

Frame construction produces a complete listing of parcels of land, averaging six to eight square miles in size, throughout a state. These parcels serve as the primary sampling units (PSUs) in a two stage design, and each contain a varying number of population units or segments. The sampling process selects PSUs, and only those selected PSUs are broken down into segments. This two stage process saves considerable time and money over that required to break the entire land area into segments.

Construction of and sampling from an area frame involves five basic steps: 1) determining specifications for the frame; 2) stratifying the land area and delineating PSUs within each stratum, 3) allocating stratum level optimal sample sizes; 4) creating sub-strata and selecting PSUs; and 5) selecting segments within PSUs. Each step is discussed in detail below.

Frame Specifications

The specifications for building an area frame consist of strata definitions and target sizes for both PSUs and segments within each stratum. Statisticians define these by examining previous survey data, and assessing

urbanization and other trends in the state's agriculture. Table 1 lists the frame specifications for the Oklahoma frame.

Strata are based on general land usage. A typical NASS area frame employs one or more strata for land in intensive agricultural (50 percent or more cultivated), extensive agricultural (15 to 50 percent cultivated), and range land (less than 15 percent cultivated). Less frequently an area frame contains "crop specific" strata. This occurs when a high percentage of the land in a state is dedicated to the production of a specific type of crop, such as citrus in Florida. In addition, each area frame uses an agri-urban and commercial stratum (more than 100 homes per square mile) plus a non-agricultural stratum including such entities as military bases, airports, and wildlife reserves. Finally, large bodies of water are separated into a water stratum.

Boundary points for agricultural strata are generally restricted to a set of standardized breaks: 15, 25, 50 and 75 percent cultivated. To determine the exact breaks for a given state, the percent cultivated for each segment sampled under the old frame is calculated from survey data. The resulting distribution is examined using the cumulative square root of frequency rule proposed by Dalenius and Hodges. The standardized breaks may be collapsed or expanded based on the structure of the distribution.

Two criteria are the most important for determining target sizes for PSUs and segments within strata: availability of good natural boundaries and the expected number of farm establishments. Generally, a lack of good boundaries will prompt the use of larger target sizes, while a large expected number of farm establishments will prompt smaller target sizes.

Stratification and Delineation of PSUs

Once strata definitions are set, the stratification process divides the land area of the state into PSUs and assigns each to the appropriate land use strata. Each PSU must conform to the definition and target size outlined for its particular stratum. PSU boundaries become a permanent part of the area frame and must be identifiable for the life of the frame. Thus the stratifier uses only the most permanent boundaries available when drawing off PSUs. Acceptable boundaries include permanent roads, rivers, and railroads. The final product of the stratification process is a "frame" file which contains a record for every PSU in a state. Specifically, each record includes the PSU number, stratum assignment, county, and size. This frame file is maintained over the life of a frame as the sampling base.

Prior to 1990 the process of stratification used paper maps, aerial photography, satellite imagery, and a considerable quantity of skilled labor. The end product was a frame delineated on paper U.S. Geological Survey 1:100,000 scale maps. In 1990, NASS implemented its Computer Aided Stratification and Sampling System (CASS). CASS automates the stratification steps on a graphical workstation using digital satellite imagery and line graph (road and waterway) data from the U.S. Geological Survey.

The digital satellite imagery employed by the CASS system is currently obtained from the thematic mapper (TM) sensor on the LANDSAT-5 satellite. The TM has a spatial resolution of 30 meters and is made up of 7 spectral bands. TM bands 1-5 and 7 reside in the reflective region of the spectrum while band 6 is located in the thermal infrared region. NASS experience with the imagery shows that bands 2, 3, and 4 highlight cultivated areas of land most accurately.

Table 1: Oklahoma Frame Specifications

Stratum	Definition	Primary Sampling Unit Size			Segment Size
		Minimum	Desired	Maximum	
		(sq. miles)	(sq. miles)	(sq. miles)	(sq. miles)
11	>75% CULTIVATED	1.00	6.0 - 8.0	12.0	1.00
12	51-75% CULTIVATED	1.00	6.0 - 8.0	12.0	1.00
20	15-50% CULTIVATED	1.00	6.0 - 8.0	12.0	1.00
31	AGRI-URBAN:>100 HOME/SQMI	0.25	1.0 - 2.0	3.0	0.25
32	COMMERCIAL:>100 HOME/SQMI	0.10	0.5 - 1.0	1.0	0.10
40	<15% CULTIVATED	3.00	18.0 - 24.0	36.0	3.00
50	NON-AGRICULTURAL	1.00	none	50.0	pps[1]
62	WATER	1.00	none	none	not sampled

1. Segments are selected with probability proportional to size; i.e. PSU's are treated as segments.

While TM data is very useful in providing information with respect to land usage, its large scale (30 meter resolution) renders it practically useless for identifying good PSU boundaries. Therefore the CASS system also uses digital files of U.S. Geological Survey 1:100,000 scale maps, in which feature class codes are assigned to all roads, water, railroads, power lines, and pipelines. The CASS system incorporates the road and waterway data from these files and overlays it on the TM imagery.

Personnel use a mouse and a "drawing" program to delineate boundaries of the PSUs and label them with their appropriate stratum number and sequence. As each PSU is completed, its size is immediately displayed. If the PSU does not fall within the particular target size, the stratifier immediately makes a correction. In addition, once a county has been completely divided into PSUs, the system can check for overlaps or omissions of land. Though the software provides many quality checks which save much time, reviews are still necessary to check the quality of stratification.

Figure 1 displays the stratification of Muskogee, OK which was performed on the CASS system. Notice the differing PSU sizes created with respect to each stratum.

Sample Allocation, Sub-stratification and Sample Selection

The national sample size for the NASS frame is approximately 15,000. The two stage design selects 15,000 PSUs and then one segment per selected PSU. The sampling process is described in more detail below.

Following stratification a multivariate optimal allocation analysis is performed to allocate the first stage sample of PSUs between land use strata. This is a multivariate procedure because the area frame must target a variety of agricultural items. The analysis requires the following inputs: a) population counts of segments per stratum; b) estimated totals of important commodities from previous year's survey; c) standard deviations from previous survey derived by locating old segments in the new frame and strata; and d) target CV's for major commodities. The analysis produces stratum level sample sizes with expected coefficients of variation less than or equal to the target CV's.

At this point in the process, PSUs have been delineated and stratified according to their land use and the optimal number of PSUs to sample in each stratum has been determined. Next, each land use strata is further divided into sub-strata based, in part, on a criteria of agricultural similarity. This process improves the precision of estimates of individual commodities and facilitates sampling by replication. PSUs are grouped by county, and ordered within counties in a serpentine pattern starting in the North East corner. Counties are then ordered based on results of a clustering algorithm that groups counties with similar crop production. Together these steps produce an ordering of all PSUs throughout the state. Figure 2 displays the county ordering for the

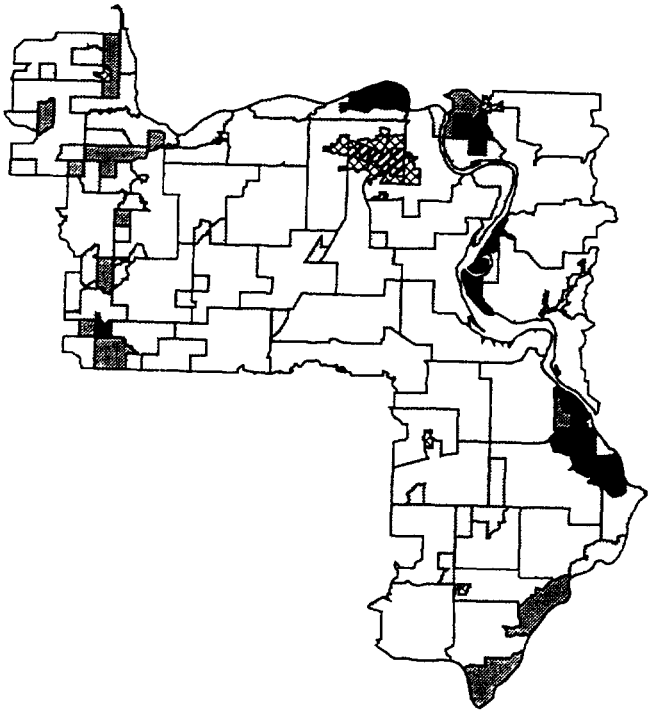


Figure 1: Stratification of Muskogee County, OK

state of Oklahoma. Substrata then equally divide the ordered PSUs within each stratum, where one PSU is selected per replicate per sub-stratum.

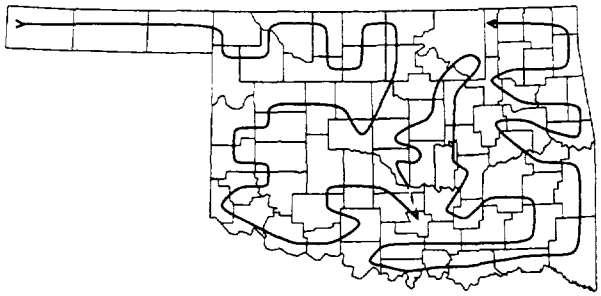


Figure 2: Oklahoma County Ordering

The sampled PSUs in each sub-stratum are randomly selected with probability proportional to the number of segments they contain and are assigned to a replicate. In non-agricultural and some range strata, where the lack of suitable boundaries is a problem, the PSUs themselves also serve as segments.

Replicated sampling has several advantages. First it facilitates sample rotation. Twenty percent of the sample in the NASS frame is rotated each year. Second, it allows estimation of year to year change from the 80 percent of the sample that did not change. Third, it simplifies the process of adjusting sample sizes to improve sampling efficiency.

The PSUs selected in the sample selection program are then located and further broken down into segments. The CASS system is used for this procedure as well. Just as the PSUs were originally delineated, so are the segments within each chosen PSU. Segments must be constructed using permanent boundaries, contain similar amounts of cultivation, and be equally sized. The CASS system randomly chooses one of the segments to sample from each PSU.

For data collection, segment boundaries are transferred to large scale NAPP photography. This process is completed by hand.

Data Collection and Estimation

Data Collection

NASS conducts a major area frame survey once each year in June. It is called the June Agricultural Survey (JAS). Approximately 15,000 segments are enumerated and yield approximately 50,000 farm establishments. These segments account for roughly 0.8 percent of the total land area of the 48 conterminous states. The survey produces estimates of crop acreages, grain stocks, number of and land in farms, livestock inventories, farm labor, and cash receipts at state, regional, and national levels. Importantly, the information collected during this survey provides a database of information about the farm establishments sampled through the area frame. This information is used as a sampling base for follow-on surveys for the remainder of the year. In particular, farm establishments are checked against the NASS list sampling frame to measure the incompleteness of that frame. The follow-on surveys generally use multiple-frame methodology, incorporating list samples with an area sample which account for this incompleteness.

Prior to the JAS data collection period, newly rotated segments along with residential, commercial, and non-agricultural segments are screened for the presence of farm establishments. States implementing a new area frame must screen all segments. Screening usually takes place in late April to early May. A questionnaire is filled out for segments which contain no agriculture.

The data collection period for the June Agricultural Survey begins June 1 and continues for two weeks. Enumerators conduct face-to-face interviews with operators of all farm establishments with land inside a segment, and account for all land within the segment. Enumerators are assigned anywhere from 8 to 15 segments to survey depending on distance between segments and the enumerator's experience level.

Estimation

In the NASS area frame, recall that segments are the population units and the second stage sampling units. The reporting units are the individual farm establishments within the segments. However, depending on the estimator that is used, these reporting units are defined somewhat differently. Sometimes the establishment reports only for its land *contained within the segment.* That part of a segment operated by a single establishment is referred to as a "tract." For other estimators, the establishment reports information for its entire operation. Other times only farm establishments whose operator lives inside the segment report information.

Three different estimators for summarizing area frame data are described below. Each has different advantages and disadvantages. Each may be used alone to estimate agricultural items, or in conjunction with a list frame to estimate for the undercoverage of that list. Variance formulations are not presented here in order to conform