



**United States  
Department of  
Agriculture**

National  
Agricultural  
Statistics  
Service

Research Division

SRB Research Report  
Number SRB-95-03

July 1995

# **Analysis of List Coverage Problems and the NOL Domain**

Orrin Musser

**ANALYSIS OF LIST COVERAGE PROBLEMS AND THE NOL DOMAIN**, by Orrin Musser, Survey Quality Research Section, Survey Research Branch, U.S. Department of Agriculture, Washington, DC, 20250-2000, July 1995, Report No. SRB-95-03.

## **ABSTRACT**

The National Agricultural Statistics Service(NASS) relies heavily on sample surveys based on a multiple frame sampling design. This design utilizes a list frame of farm operators with control data for efficient sampling, and an area frame of all land in the United States, which provides complete coverage. NASS multiple frame expansions are simply the sum of the list frame sample expansions and the area frame sample expansions, using data from only those area sampling units(tracts) that are not represented on the list frame. These are referred to as non-overlap(NOL) tracts. As part of the overall NASS effort to improve survey quality, there has been some concern with the volatility of the NOL component and its impact on the overall expansions. This volatility seems to be caused by the presence or absence of large operations(outliers) in a particular sample. Now, the question is one of coverage: Why aren't these large operations represented on the list frame?

NASS has devoted substantial resources to list building, and has made steady progress in coverage of the farm population, but coverage remains below target levels. List maintenance is also very important for coverage because information on the list is used to determine whether a record is eligible for sampling. Analysis of data from three states suggest that almost one third of the NOL tracts are represented by records on the list frame which were not eligible for sampling. Many of these records should have been active and classified but were not, due to inaccurate or out-of-date control data. There is also evidence that these NOL operations matching unclassified list frame records tend to be larger than other NOL operations, and thus have a greater potential to cause volatility. Data from the 1991 through 1994 June Agricultural Surveys for all states show that the percentage of overlap tracts rises substantially as tracts move through the five year cycle. Thus early, correct identification of overlap tracts is an important coverage issue.

<p>This paper was prepared for limited distribution to the research community outside the U.S. Department of Agriculture. The views expressed herein are not necessarily those of NASS or USDA.</p>
---

## **KEY WORDS**

Multiple frame; Extreme operator; Outlier; Coverage; List frame; Classified records; List building; List maintenance.

## **ACKNOWLEDGEMENTS**

The author would like to thank Charles Day for his time and effort in performing the record linkage. Thanks to Dale Atkinson for advice and support with this project. NASS research reports by Ron Bosecker detailing the evolution of multiple frame sampling were very helpful.

# TABLE OF CONTENTS

SUMMARY .....	iii
INTRODUCTION .....	1
BACKGROUND .....	2
METHODS AND RESULTS .....	5
Observations on 1993 JAS NOL Tracts that Match Unclassified List Records .....	7
Are NOL Operations that Match Unclassified List Records Different? .....	9
Overlap Percentage Trends Over Five Year Survey Cycle .....	10
CONCLUSIONS AND RECOMMENDATIONS .....	11
BIBLIOGRAPHY .....	13

## SUMMARY

The National Agricultural Statistics Service(NASS) relies heavily on sample surveys based on a multiple frame sampling design. This design utilizes a list frame of farm operators and an area frame of all land in the United States. NASS multiple frame expansions(estimates) have two components which represent these two domains of the target population of farms. The list component is calculated from a list sample, while the area component is calculated from an independent area sample, using data from only those sampling units(tracts) that are not represented on the list frame. These are called non-overlap(NOL) tracts. NASS first implemented multiple frame sampling in the 1960's in an effort to reduce the variance of its estimators caused by the presence or absence of very large livestock operations in its area frame samples.

As part of the current overall NASS effort to improve survey quality, there has been some concern with the volatility of NOL expansions, and their impact on the multiple frame expansions. Like the original problem which led NASS to the multiple frame methodology, this volatility seems to be caused by the presence or absence of extreme operations. Now, the question is one of coverage: Why aren't these extreme operations being accounted for by the list frame? How can we improve the coverage of these operations, short of a massive and very expensive list building program?

Looking at the relatively short history of the list frame, it appears that coverage of large farms is good. However, NASS is well below overall coverage goals as expressed in the early part of this decade. In spite of the fact that the population of farms is largely composed of old operations that have been in business for 15 years or more, a substantial number of these older farms(about 30%) remain in the NOL domain. On average it takes a new farming operation about 3.4 years to get added to the list frame, and in the three states that we focused on, there were large operations that remained in the NOL domain for three to ten years.

In addition to list building activities, list maintenance is also very important for maintaining good coverage because list frame information is used to determine eligibility for sampling. In the three states we examined, about one third of the NOL tracts were represented on the list frame by records which were not classified(eligible for sampling). These list records were not classified for one of two reasons. Either they were active records not classified due to small control data, or they were inactive records erroneously thought to be out-of-business, retired, moved out of state, etc. Our findings suggest that many of these records should have been classified, but were not because of inaccurate or out-of-date control data. There is also evidence that these NOL operations represented by non-classified list records tend to be larger than other NOL's, and thus have potential to contribute heavily to the outlier volatility problems.

Data from the 1991 through 1994 June Agricultural Surveys for all states show that the percentage of overlap tracts increases as they move through the current five year area frame rotation cycle. This is thought to be the result of the lack of time and adequate information to correctly identify overlap tracts in the initial part of this five year cycle. Thus the overlap checking process is also an important coverage concern.

## INTRODUCTION

The Agricultural Survey Program is a yearly cycle of surveys conducted by The National Agricultural Statistics Service(NASS) to provide timely and reliable agricultural statistics, which include estimates for crop production, grain stocks, and hog inventories. NASS relies heavily on sample surveys which use a multiple frame sampling design with a list frame of farms and farm operators, and a supplemental area frame of all land in the U.S. The list frame provides cost efficient sampling, with good coverage of medium and large farms. The area frame provides complete coverage of all farms and also a measure of the completeness of the list frame.

NASS multiple frame survey expansions have two components which represent domains of the U.S. population of farms. These domains are two mutually exclusive and exhaustive categories of farms: those that are represented on the NASS list frame and those that are not. The multiple frame expansion is the sum of the expansions for these two domains. The expansion for the "not on list" domain is based on data from sampled units(tracts) from the area frame that are not represented by records on the list frame that were eligible for sampling, or "classified". During a survey, all area frame tracts are matched against the list frame. Tracts which are represented by records on the list frame that were classified for that survey are categorized as "overlap"(OL) and similarly the remainder are categorized as "non-overlap"(NOL). Data for the OL domain is zeroed out, so that only the NOL tracts contribute to the area frame portion of the multiple frame expansion.

There have been two important concerns raised over the role of the NOL sample in

the overall effort to improve survey quality at NASS. The first is a data collection concern that there is excessive respondent burden for this sampling domain. Under the NASS area frame sample design, one fifth of the area frame sample is replaced yearly, and each sampled unit is included in the Agricultural Survey Program cycle for a five year period. Given a relatively small pool of NOL tracts to choose from, most NOL's are contacted very often during the five year period, resulting in substantial respondent burden.

The second is an estimation concern based on a history of extreme volatility of the NOL domain's contribution to the multiple frame estimates. Due to a very low sampling rate for the area frame, the sampling weights, or expansion factors for the NOL tracts are very large. When a large expansion factor is applied to a large NOL operation, this can result in an extreme outlier in a survey. The presence or absence of these outliers can cause extreme volatility in the NOL domain expansions from survey to survey.

This report brings together data from several sources to provide a better understanding of the NOL domain, and our problems in sampling from that domain. If our list frame were complete, we wouldn't need an area frame. If we had a complete list of all large farms, we wouldn't have the outlier problem. So a primary concern of this report is list frame coverage.

First, we will look at NASS' goals and results in improving list coverage. It is also helpful to look at a recent "snapshot" of the age of farm operations in the U.S., which is interesting in itself but also provides some information on how long farms remain in the NOL domain. We will look at a study which matched NOL operations in three

states against the list frame, and determined why the matching list records were not classified for the survey. We will also examine the tendency for the overlap percentage to increase as tracts move through the five year survey cycle. The overall goal is to understand why the large operations that we would expect to be accounted for by the list frame continue to "surface" in the NOL, causing volatility in survey estimates.

## BACKGROUND

A brief history of the list and area frames is helpful in understanding the evolution of the OL and NOL domains. Since 1863 when the U.S. Department of Agriculture issued its first crop report, lists of farms and farm operators have been used to gather agricultural statistics. The use of an area frame for probability sampling in agriculture began in 1943 with the development of the Master Sample of Agriculture under a cooperative agreement with Iowa State College(now Iowa State University), the Department of Agriculture and the Bureau of the Census. The first recurring agricultural statistical program using probability sampling from the Master Sample Frame began in 1954 with the June Enumerative Survey. By 1965, the June and December enumerative surveys were operational in all 48 conterminous states.

One of the early concerns with area frame sampling was the impact of the large livestock operations on estimates(and this is still a major concern for the NOL, as outlined in the Introduction). The presence or absence of these "extreme" operations could distort survey results and was thought to be a source of variance that might be greatly reduced with alternative techniques. In 1963, a list of "extreme operators"(EOs) was developed which would allow list

sampling of large operations, and the removal of the volatile extreme operator data from the area sample. Research and evaluation of these multiple frame methods began in 1965 and pilot studies were conducted in several states in the late 1960's. In 1969, four states began a continuous series of multiple frame surveys using more extensive lists which were stratified by size. These provided significantly lower sampling errors than surveys using the small lists of extreme operators. During the 1970's, lists were compiled and maintained in the State Statistical Offices(SSOs), and in 1979 work began on a comprehensive national list frame system. By 1981 a national list frame was consolidated on a central computer system, and sampling for national multiple frame surveys was conducted from headquarters.

Since the initial list building in the late 1970's, a continuous maintenance program has been in place, with goals of keeping list frame data accurate and current. Control data providing various measures of size are used to provide more efficient sampling, and play a very important role in classification and stratification. Farms can change in size and type of operation very quickly, or go in and out of business. Many active records on the list frame are not classified for a particular survey because their control data do not meet the requirements of any of the strata. If this occurs due to inaccurate control data, then these operations will be erroneously accounted for by the NOL. Similarly, if a large operation is erroneously thought to be out of business, its list record will be inactive and the operation will become a potential NOL outlier. Thus maintenance of accurate list frame information is essential to good coverage.

NASS has periodically published goals for

list coverage(percentage of farms accounted for by the list frame), and there have been several periods of list building to work towards meeting these goals. In 1992, the goals were 95% coverage for large farms and 80% coverage of all farms by 1995. Budget constraints have and may continue to

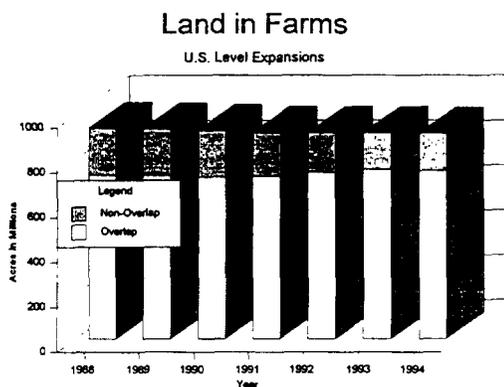
slow the progress towards these goals. The area frame sample is used to measure the coverage of the list frame, and estimates of coverage are published yearly. Coverage for 1994, categorized by farm size and type of farm are given in the following table.

List Coverage of 1994 Ag Tracts by Reported Sales					
Percentage of Overlap Tracts for All States					
Farm Size	All Farms	Crop Farms	Livestock Farms	Specialty Farms	Land in Farms
\$1000-9,999	39	41	39	30	51
\$10,000-99,999	69	74	71	49	76
\$100,000 plus	87	90	89	68	91
\$500,000 plus	89	94	91	76	94
Overall	58	66	57	44	82

Red book coverage figures for the last 10 years have shown a slow but steadily increasing trend, from 54.1 % in 1988 to 57.7% in 1994 for number of farms, and from 77.7% to 81.7% for overall farm acreage. All categories of farms have shown coverage improvements and there is clear evidence that list building activities in 1990-1991 resulted in significant improvements in coverage. However, current coverage of large farms (\$100,000 plus) is estimated at 87% which is considerably less than the goal of 95%. Given the current budget constraints, is likely to remain relatively constant in the near future.

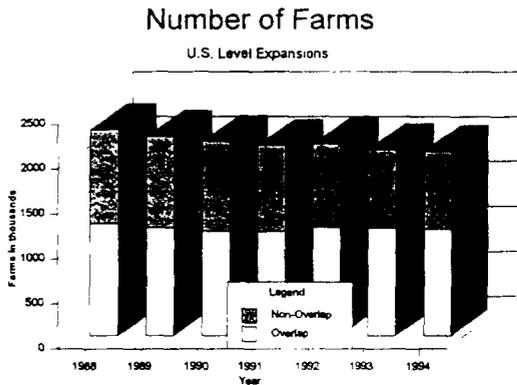
acreage has remained relatively constant, especially for the last four years, while overlap acreage has steadily increased. The expansions for overlap farm acres show an increasing trend in coverage with the NOL domain falling from 210 million acres in 1988 to 167 million acres in 1994, while the

The "Land in Farms" graph shows trends in total farm acreage and in list coverage of that acreage for the past 7 years, based on expansions from the June area sample. Land in farms expansions indicate that U.S. farm



OL domain has risen from an estimated 730 million acres in 1988 to about 750 million acres in 1994.

The "Number of Farms" graph shows the steady downward trend in the number of U.S. farms, while the expansions for OL farms remained pretty constant, resulting in an increase in the coverage percentage.

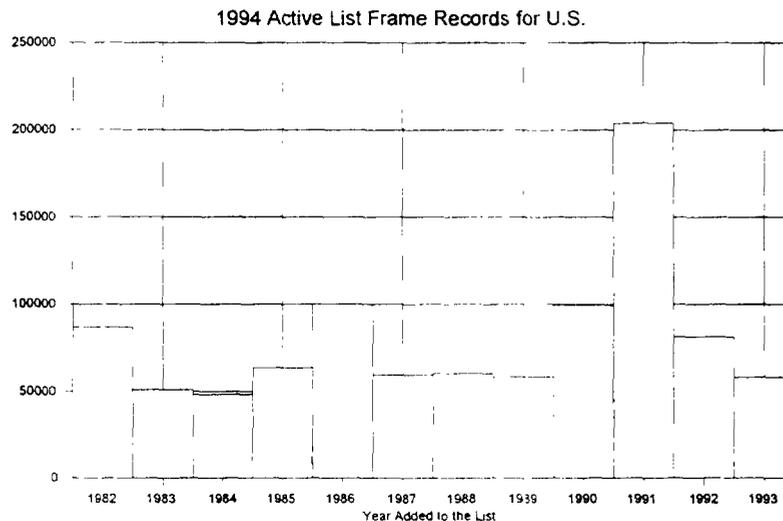


The "List Building Activity" graph below

shows the distribution of the year in which active(1994) list frame records were added to the list(original list construction years 1977-81 omitted). This suggests that list building has been pretty uniform over the history of the frame, with a very clear peak in 1990-1991, when there was a concerted effort to increase coverage.

We should keep in mind that there are also active records which become inactive each year. We often discover that active records represent out-of-business operations when those records are sampled and the interview takes place. So a portion of the list frame is actually becoming "deadwood" each year but the records don't become inactive until they are sampled or are contacted in list maintenance work. New operations which replace these may be discovered through such interviews and contacts and added to the list, but most new farming operations will only be added to the list through list building activities. So ongoing list building activity is essential in maintaining good coverage.

### List Building Activity

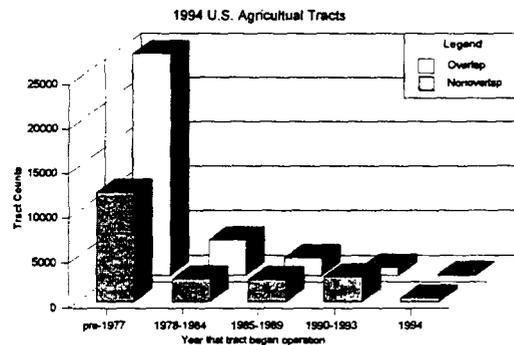


## METHODS AND RESULTS

There was a new question on 1994 June Agricultural Survey (JAS) area questionnaires in all states: "In what year did [the name on label] begin operating in this State? [If before 1978, then record as: 1977]." The collected data provide an interesting "snapshot" suggesting that over 70% of all farms began operating before 1978. It is very clear that the overlap percentage declines for "younger" operations. Of the "older" pre-1978 operations, about 67% are now OL, compared to 44% of 1985-89 operations, and only 24% of the 1990-93 operations.

It is interesting to break down these results by some measure of farm size. For this, we used the survey reported estimate of gross farm sales for the previous year. NASS list building efforts concentrate on getting the

Year that Tract Began Operation



larger operations on the list, and letting the area frame provide coverage of the smaller operations. Also, there are many active list frame records which are not classified for list selection in the JAS and this increases the % NOL for the smaller operations. So we would expect the percentage of OL farms to increase with age and size of farm. The following table clearly shows this tendency.

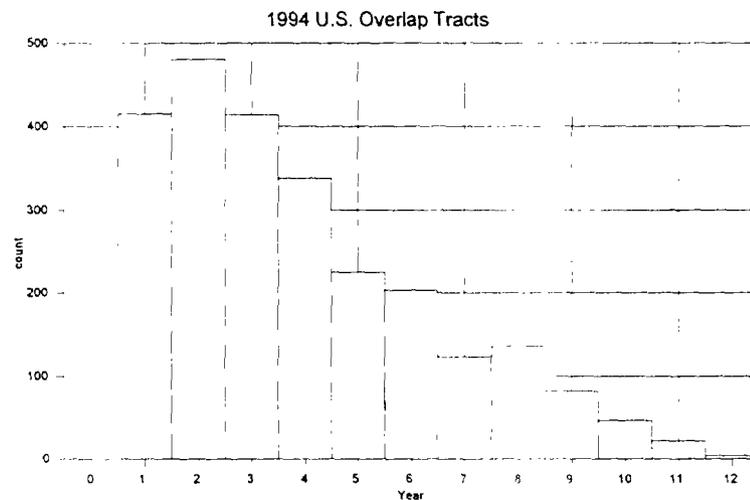
Age of Agricultural Tracts By Reported Sales					
Percentage and (% of those OL) of 1994 U.S. Tracts that Began Operation In ...					
Gross Sales	pre-1978	78-84	85-89	90-93	94
< \$1000	50.6 (9.6)	12.2 (8.3)	13.7 (5.4)	19.0 (2.9)	4.6 (1.0)
1000-20,000	63.0 (31.8)	11.8 (24.3)	11.4 (17.3)	12.2 (8.5)	1.5 (4.2)
20-100,000	73.3 (69.1)	12.1 (62.8)	8.0 (49.2)	5.8 (27.9)	.9 (17.2)
> \$100,000	76.4 (87.2)	12.8 (84.5)	6.6 (76.2)	3.7 (62.0)	.5 (47.1)
overall	71.1 (67.2)	12.3 (60.8)	8.5 (44.4)	7.0 (24.1)	1.1 (15.4)

We can also use the survey reported year in which an overlap tract began operating and the year in which that tract's list frame record was added to the list to get a measure of how long farms remain NOL before being added to the list. The graph below shows 1994 June area tracts that are OL, and how long it took from beginning operation until being added to the list. In this graph we only considered farms that began operating after the list frame construction period was over, 1982 and later. The average time was about three years and did not seem to depend on farm size, as measured by 1994 reported gross farm sales. More than one third of the OL tracts with 1994 reported sales >\$100,000

had spent 4 or more years in the NOL domain before being added to the list (gross sales may have been smaller in previous years). Clearly, new operations are an important cause of coverage problems, and emphasis should be placed on finding sources for names of new operations.

Farms also may be less stable today, and very large chicken and hog operations can go in and out of business or move from state to state in very short time periods. Several states reported that such operations are a major source of NOL outliers, and that list building activity should also focus on finding this type of operation.

## Years from Beginning Operation to OL



**OBSERVATIONS ON 1993 JAS NOL  
TRACTS THAT MATCH  
UNCLASSIFIED LIST RECORDS**

The goal of this phase of the project was to determine if a significant number of the JAS NOL tracts were represented on the list frame. We were primarily interested in NOL records that matched list frame records which were not eligible for list sampling ("classified"). These could be active records which were not classified for JAS due to small control data (which did not meet the minimum strata definitions for their state), or they could be inactive records, thought to be out of business, retired, deceased, etc. A record linkage program was used to match 1993 JAS NOL records against the entire list frame, including

inactive records, for three states: New York, Virginia, and Wyoming. Counts of matches categorized into "reasons why not eligible for sampling" using the record status variable, will be presented for these three states combined, with a more thorough analysis of all matches for one state.

The 1993 June Agricultural Survey area frame samples for New York, Virginia, and Wyoming consist of a total of 983 segments, containing 8719 tracts, of which 2519 were agricultural tracts. Of these, 1731 were overlap for the survey, leaving 788 "non-overlap" or NOL tracts. Of these, 270 (about 34%) matched a list frame record of some kind. These matches for the three states are categorized by record status below:

1993 JAS NOL Tracts Which Were On the List Totals for New York, Virginia and Wyoming Reasons Why Not Eligible for Sampling		
Record Status	Definition	Count
00	Active records not classified for the JAS.	101
02	Retired	17
03	Out of business, sold, foreclosed, etc	22
04	Rent or lease farm	28
05	Moved out of state	3
07	Non Farm by NASS definition	10
40	New Adds or ASCS tryouts	27
other	Assorted.	62

Thus, for these three states it appears that about one third of the NOL domain is represented on the list. With a few exceptions these list records which matched NOL operations were not classified for the JAS. These fall into two major categories.

The first would be active records not classified for JAS due to "small" control data that did not meet their state's strata definitions. This is just part of the multiple frame methodology in which cost efficiency is gained by sampling more heavily from "larger operation" strata and letting the NOL account for the very small operations. This works well when the control data used for stratification are accurate and up-to-date. However, if the control data are inaccurate, there may be large operations which are not classified. Such operations are potential NOL outliers.

The second category is comprised of records that were inactive for reasons like "retired", "out of business", "rented", or "non-farm by NASS definition". In most of the latter cases, the fact that the NOL tracts were considered agricultural tracts and met the NASS farm criteria would suggest that these records are inactive due to erroneous or out-of-date list frame information. The operator of the tract matched the operator on the list frame, and was not retired, out of business, or renting his land.

To carefully examine these potential sampling problems, a very thorough analysis was performed on every match in one state. The largest number of matches were with active records that were not classified for list sampling in the JAS. Under 1993 JAS strata definitions, positive data in the cropland list frame field would cause a list frame record to be classified, but cattle control data alone would not. Thus, a large cattle operation with no cropland might not

be classified for the JAS. However, of these matching active records that were not classified, about 43% reported positive cropland in June, and thus should have been classified. There were some rather large values for cropland reported by these operations, including 10 reports with over 400 acres of cropland.

Therefore, inaccurate or out-of-date control data on the list frame can cause a significant number of operations which should have been classified, and thus accounted for by the list, to become potential NOL outliers.

Of NOL tracts matching list records which were inactive and thought to represent either operators who were retired or who were renting out or leasing all their land, over 50% reported positive cropland, and thus should have been active, classified records. Several of these had large reported values for cropland.

Another large group of matches were those with a record status indicating "out of business, sold, foreclosed". Several of these were fairly large operations, with three reporting over 500 acres of cropland. This could also be labeled as inaccurate or out-of-date list frame information, causing an operation which should have been accounted for by the list to show up as NOL. Given that farms can go in and out of business from one year to the next, it might be advisable to make annual contacts with operations that have recently gone out of business to keep the list frame up-to-date and improve coverage. It is also possible that some operations are incorrectly recorded as out of business due to a respondent error, or attempt to avoid participation in the survey. It might make sense in certain circumstances to attempt to substantiate the out of business status.

There were four "ASCS tryouts", records added to the list from ASCS(now the Consolidated Farm Services Agency) and given an inactive record status until criteria work could be done. All of these records reported positive cropland and thus would have been classified if they had been active records with accurate control data.

If the list frame records were all up-to-date and accurate, many larger NOL's would have been accounted for by the list. Is there a way to target a type of record which may be more likely to change or require updating of control data? We could do a periodic check of out-of-business and retired operators. Active records with control data too small for classification could also be checked periodically.

**ARE NOL OPERATIONS THAT MATCH UNCLASSIFIED LIST RECORDS DIFFERENT?**

We might expect that most of the NOL operations that matched inactive or unclassified list records would be more like

the OL operations than other NOL's, since most if not all were active list records at one time. The following cross tabulation of reported gross farm sales from the 1993 JAS for these three states, by overlap classification, gives some evidence that this is true. NOL operations that are on the list(inactive or unclassified) are more like the OL operations in gross farm sales. While 74% of the NOL's that are not on the list frame report gross sales less than \$20,000, only 48% of the on-list NOL's, and 25% of the OL's fall into this low sales category. Similarly only 12% of the not-on-list NOL's report gross sales of over \$100,000, compared to 28% for the on-list NOL's. We should keep in mind that some of these on-list NOL operations are active records on the list, with accurate small control data values and thus were not classified for the JAS for valid reasons. However, those with inaccurate control data values(or incorrect record statuses) that should have been classified tend to be larger operations than other NOL's and thus have the potential to cause outlier problems.

Overlap Status by Reported Farm Sales For 1993 Ag Tracts in Virginia, New York, and Wyoming Counts and Row Percentages				
	< \$20,000	\$20,000 to 100,000	\$100,000 +	All
OL, Matched Classified	435 25.13	460 26.57	836 48.30	1731 100.00
NOL, Matched Non-Classified	92 47.92	47 24.48	53 27.60	89 100.00
NOL, Did Not Match	439 73.66	88 14.77	69 11.58	596 100.00
All	966 38.35	595 23.62	958 38.03	2519 100.00

**OVERLAP PERCENTAGE TRENDS  
OVER THE FIVE YEAR SURVEY  
CYCLE**

Under the NASS area frame sample design, one fifth of the area sample is replaced yearly, and each sampled unit is included in the Agricultural Survey Program cycle for a five year period. Thus in a given year there are five "rotation groups", with roughly one

fifth of the sample in each. We will label the current year sample, or rotation group, as group 1, last year's rotation group as group 2, etc. So rotation group 5 was first sampled four years ago, and this is the fifth year that the sampled units have been in the survey. The following table gives the overlap percentage for all agricultural tracts by rotation group for the last four years, and also for the four years combined.

Overlap Percentage: Survey Year By Rotation Group						
1991-1994 JAS For All States						
	1	2	3	4	5	overall
1991	56.57	62.63	60.93	62.43	60.93	60.55
1992	54.61	58.03	61.30	58.97	60.88	58.57
1993	58.88	59.30	60.35	63.45	62.35	60.74
1994	53.91	62.64	62.75	62.62	65.12	60.88
Combined	55.93	60.64	61.3	61.88	62.46	60.16

If the five rotation groups represented five independent samples from the same population, and the overlap percentages could be determined exactly, we would expect the numbers to be very close, given the large sample size. If the actual overlap percentage was 60, the standard error for the estimated percentage would be about .5, and thus we would expect the cells to vary by less than 2 percentage points. Instead, the overlap percentage shows a clear increasing trend as the number of years in the survey increases.

Looking at the combined results, there is a very large jump of 4.7 percent from the first to second year tracts. and a difference of

6.5 between first and fifth year tracts. Following the diagonal from the upper left cell to the lower right shows the changes in OL percentage for the 1991 tracts for years 1991 through 1994. Again there is a consistent rising trend. This is true for 1992 and 1993 tracts as well. This is important because it indicates that the differences are not due to differences in the samples, because here we are following the same samples across time.

In 1993, the initial OL percentage was very high, while in 1994 the initial percentage was very low, at 53.91%. With list coverage increasing yearly, we would expect to see an increasing trend in first year

overlap percentage, as we do see for rotation group 5. Instead we see a lot of variation in the first year percentages. In fact, the OL percentage for 1994 first year tracts was the lowest of the last four years, and was almost 5 points less than the 1993 OL percentage.

The increasing trend in the overlap percentage is well known at NASS, and is thought to be a result of the overlap determination process, which is ongoing during the entire five year cycle. The initial overlap determination for new tracts occurs during the June Agricultural Survey, and is subject to strict time constraints and is often based on limited information. The initial interview may not have been with the operator, and crucial information for matching a list frame record, such as Social Security Number or phone number, may not be available. The overlap determination is also based on the "operating arrangement" or type of operation, such as managed or partnership, and this is often not determined correctly initially. Further contacts in the survey cycle give the opportunity to gather more information and thus help identify OL tracts. The process of gathering information and clarifying the nature of an operation is ongoing over the five year cycle, and therefore the probability of correctly identifying all overlap tracts increases during this cycle. System changes that could improve the early phases of this process and speed up identification of OL tracts, would be helpful in controlling the NOL coverage problems.

For estimation reasons, it is very important to maintain the independence of the list and area frames. NASS survey procedures and guidelines stress that list building should never involve adding farm operations from the area sample to the list frame and we would like to rule out the possibility that this

occurs. However, it is possible that this could happen inadvertently during heavy list building activity simply as a result of staff familiarity with NOL operations that had been sampled in the recent past.

## **CONCLUSIONS AND RECOMMENDATIONS**

As part of the overall NASS effort to improve survey quality, there has been some concern with the volatility of NOL expansions, and their impact on overall expansions. This volatility is thought to be caused by the presence or absence of extreme operations in a particular sample. Multiple frame methodology attempts to control such volatility by using a list frame to sample large and specialty operations. This should remove the extreme operation data from the area sample where expansion factors may be large. This approach is effective to the extent that the coverage of large operations is high and control data on the list frame are accurate. Therefore, our recommendations focus on the issues of list coverage and maintenance.

### ***1. CONTINUE LIST BUILDING EFFORTS, WITH EMPHASIS ON FINDING LIVESTOCK OPERATIONS THAT GO IN AND OUT OF BUSINESS AND/OR MOVE FROM STATE TO STATE.***

The short history of the NASS list frame shows that while coverage has steadily improved, it remains below published goals. It is important that list building remains a high priority for NASS, with emphasis on finding large operations. Modern livestock operations can go in and out of business over a short time frame, and often move from one state to another. It is important to target such operations.

*2. PERFORM PERIODIC CHECKS OF INACTIVE LIST FRAME RECORDS AND CONTINUE EFFORTS TO KEEP CONTROL DATA UP-TO-DATE.*

List maintenance is also very important for coverage of large operations because list frame information is used to determine eligibility for sampling. In this study, about one third of the 1993 JAS NOL agricultural tracts in three states matched list records that were ineligible for sampling. Many of these list records were inactive records representing farm operators or operations erroneously thought to be out of business, retired, rented out, etc. This clearly suggests that many inactive list frame records represent operators with a potential to become active again, who should be contacted periodically. Many of the inactive records still have control data which could be used to target potential outliers for more frequent contact.

The data also suggest that many operations which should have been classified were not, due to inaccurate control data. Periodic contacts with such operations are important to keep the list frame current and to maintain coverage. This would also result in more efficient sampling due to more accurate classification. This is particularly important because operations that should have been active and classified tend to be larger than other NOL operations, and thus have a higher probability of being outliers.

*3. CONSIDER SAMPLING FROM INACTIVE RECORDS AND ACTIVE UNCLASSIFIED RECORDS "WITH POTENTIAL".*

Another possible strategy might be to rethink the classification process, recognizing the fact that the list frame

control data are not perfect. One approach, in light of the high percentage of matches of NOL tracts with list frame records not eligible for sampling, would be to classify inactive list frame records that have some probability of becoming active again (i.e. retired, out of business, but perhaps not deceased). It might also be advisable to rethink the minimum control data requirements for classification, with the idea that control data are always in the process of becoming outdated, and that sampling from a low control data stratum is one way to periodically update the control data. These uses of list frame sampling to capture new and/or growing operations would be similar to the practice of area frame sub-sampling from non-agricultural tracts with potential for agriculture.

Unfortunately, unless sample sizes were increased, these measures would have the effect of increasing the expansion factors for the lower list frame strata, and thus increasing outlier problems for the list expansions.

*4. IMPROVE THE OVERLAP CHECKING PROCESS IN THE EARLY PART OF THE SURVEY CYCLE.*

Data from the 1991-1994 JAS indicate that on the average the overlap percentage for first year tracts is more than 5% lower than that for tracts that have been in the survey cycle for 2-5 years. If this difference is in fact caused by inaccurate or incomplete overlap determination for first year tracts, efforts should be made to improve the overlap determination process, including data collection and overlap checking. During training for data collection, emphasis should be placed on the importance of obtaining key information needed for matching against the list frame, and for correctly identifying operation types such as managed or

partnership. Improvements in overlap checking, such as the use of a very efficient record linkage program, might also be helpful in improving this process which takes place under the strict time constraints of the June survey.

Further research is recommended to determine the extent to which first year tracts contribute to the volatility in the JAS NOL expansions. If a disproportionate share of the volatility is due to first year tracts, it might make sense to consider excluding data for these tracts from NOL domain expansions. While this may seem an extreme recommendation, it was part of the early NASS multiple frame survey procedure. Overlap checking for first year tracts was performed after the June survey, and the NOL expansion for the second through fifth year tracts(80% expansion) was adjusted up to 100% in the summary.

## BIBLIOGRAPHY

- Allen, J.D., "1994 List Frame Evaluation," USDA, National Agricultural Statistics Service, Survey Management Division, NASS Staff Report SSB-94-01.
- Bosecker, R. "Integrated Agricultural Surveys." USDA, National Agricultural Research Service, Research Division, NASS SRB Research Report Number SSB-89-05.
- Bosecker, R., Vogel, F. A., Tortora, R. D., and Hunuschak, G. A. "The History of Survey Methods in Agriculture(1863-1989)." USDA, National Agricultural Research Service, 1989.
- Bosecker, R. "The Quarterly Agricultural Surveys," USDA, National Agricultural Research Service, 1987.
- Hill, G.W. and Steiner, M.A. "List Frame Evaluation From an Area Perspective," USDA, Statistical Reporting Service, 1985.
- Geuder, J. "1992 List Frame Evaluation", USDA, National Agricultural Statistics Service, Survey Management Division, NASS Staff Report SSB-92-02.
- Holland, Theresa, "1988 List Frame Evaluation", USDA, National Agricultural Statistics Service, Research and Applications Division, NASS Staff Report SSB-88-11.
- Patterson, Rex, "1993 List Frame Evaluation", USDA, National Agricultural Statistics Service, Survey Management Division, NASS Staff Report SSB-93-02.
- Scott, Carl, "1990 List Frame Evaluation", USDA, National Agricultural Statistics Service, Research and Applications Division, NASS Staff Report SSB-91-01.