THE EFFECT ON SAMPLING ERRORS RESULTING FROM

USING A LIVESTOCK LIST WITH AN AREA FRAME SAMPLE

Dwight A. Rockwell

Sample Survey Research Branch
Research Division
Statistical Reporting Service
U.S. Department of Agriculture
Washington, D.C.

August, 1976

CONTENTS

THE EFFECT ON SAMPLING ERRORS RESULTING FROM

USING A LIVESTOCK LIST WITH AN AREA FRAME SAMPLE

INTRODUCTION

The June Enumerative Survey (JES) and December Enumerative Survey (DES) conducted by the Statistical Reporting Service are commonly referred to as area frame sample surveys. This description is correct as far as most crops are concerned. However, with respect to cattle and hogs, the JES and DES have become multiple frame surveys.

A list of names representing large hog and cattle operations, known as the extreme operator (E.O.) list, has been maintained and used in conjunction with the area frame sample. The livestock data associated with any area frame sample tract operated by a person whose name is on the E.O. list are excluded from the area frame estimate. An estimate of the livestock on land operated by persons on the E.O. list is provided by a sample drawn from the list. A JES or DES estimate of total livestock is computed by adding the area frame estimate (excluding E.O.'s) to the E.O. list estimate. This estimate is a special type of multiple frame estimate known as the "screening estimator." This multiple frame procedure for hogs and cattle was implemented several years prior to the introduction of what we call a multiple frame livestock survey.

This JES and DES sampling procedure was implemented to improve the precision of the livestock estimates. In many states, large livestock operations are a rare item. In general the area frame sample cannot estimate efficiently for a rate item. By excluding each large livestock operation from the area frame, placing a name for each operation on a stratified E.O. list and sampling this list at a high rate, the sampling error for each total livestock estimate can be reduced. However, some additional handling procedures are required.

In spite of this effort to improve the precision of livestock estimates, it was later decided more reliable state estimates were needed, particularly for the major hog and cattle producing states. Beginning in 1969 a procedure called multiple frame (MF) sampling was introduced in four states. This procedure consists of using a large list of farm operators which includes the E.O.'s used for the JES and DES. Since that time the MF sampling procedure has been expanded to include 14 states for hogs and 28 states for cattle.

For each MF state we now have a relatively precise MF estimate which is given primary emphasis in arriving at a Board State estimate. Also, for each MF state we have a JES or DES estimate to combine with those from other states (both MF and non-MF) to arrive at regional and national estimates. Seemingly we now have a nearly optimum sampling procedure: (1) a JES or DES survey using a small E.O. livestock list sample to provide estimates with acceptable precision at the regional and national level and (2) a MF survey to provide state estimates with acceptable precision for the major producing states.

Initially, E.O. lists were small and included only the very largest operations. Over time these lists have grown in size for two reasons. First, the number of large operations has increased in many states. Secondly, for some states the minimum size (i.e., cut-off) for E.O. classification was lowered. The E.O. cut-off was lowered substantially for cattle in six non-MF states and for hogs in nine non-MF states. These are now called modified MF states. This step was taken in states where a more precise estimate was needed, but where a regular MF survey, in addition to the JES and DES, was deemed unjustifiable. The E.O. cut-off has also been lowered for some states that have regular MF surveys. For these states it seems we are trying to get the two survey estimates (JES or DES and MF) to agree where before there have often been substantial differences. This seems contrary to the purpose of having both JES or DES and MF surveys in certain states. With larger E.O. lists, overlap determination between the list and the area frame has become more time-consuming and as experience has shown, less accurate. Although the use of larger E.O. lists has probably resulted in smaller sampling errors in the JES and DES estimates, we may have introduced some nonsampling errors.

The purpose of this study is to examine the reduction in sampling error realized in the JES hog and cattle estimates resulting from using an E.O. list. Data from four MF states are used. Before presenting the analysis, let us examine some of the potential nonsampling errors associated with using an E.O. list in conjunction with the JES.

NONSAMPLING ERRORS

There are several possible sources of nonsampling errors associated with list sampling or multiple frame sampling that are less prominent with area sampling. Some of the more important sources will be discussed.

Sampling Unit vs. Reporting Unit.

With list frame sampling, the sampling unit is a name. The reporting unit is the land operated by the person whose name has been selected. Conveying this concept to a respondent, particularly by mail questionnaire, is difficult and the data are therefore, susceptible to considerable error.[1] Also, a special set of rules for handling partnership operations reported by individuals must be used. It is very difficult to obtain the necessary data and perform the editing to consistently meet the requirements of this set of rules. A certain amount of error is certainly inescapable.[2]

With area frame sampling, the sample unit is a segment of land. One reporting unit is the tract, which is all land inside the segment under one operation (closed segment approach). The other reporting unit is all land, both inside and outside the segment, operated as one operation by a person residing inside the segment boundaries (open segment approach). Since these concepts are presented to a respondent during the course of a personal interview, the open segment approach should be at least as accurate as list frame sampling, and the closed segment approach should be superior. The enumerator can provide immediate assistance to a respondent in understanding the definition of a reporting unit. This is more easily explained if the closed segment approach is used. Unclear segment boundaries can pose a problem, but earlier studies have shown survey procedures to be adequate to overcome them.[3]

Occasionally a name and address on a list will be a very common first and last name, or just an initial and very common last name, in combination with a general rural address. Under these conditions there is a chance the name selected is not that of the person who receives and completes the questionnaire.

Overlap Determination

The use of E.O. lists requires that each area frame sample tract be classified as overlap or nonoverlap with a list. This requirement cannot be met without error. The larger the list, the larger the task and the greater the chance of error. Occassionally the correct operator name is not obtained for a tract or the name is not spelled correctly. Also, names are sometimes spelled incorrectly on the list. In either case a check for match with the list may result in an incorrect overlap determination.

In past surveys once overlap was determined, E.O. tract data were manually edited to zero on the questionnaire. With the use of code boxes indicating overlap, this task is now handled by the computer edit program and the occurrance of error has been reduced. Also, presurvey overlap classification and coding for nonrotated segments has helped. Only those tracts with new operators or with name and address corrections must be reclassified during the survey.

A set of rules consistent with that used for handling partnership operations reported by an individual whose name is selected from the list is used to determine overlap status for partnership tracts. This strict requirement introduces yet another source of nonsampling error.[4/]

Refusals and Inaccessibles

Each time data are not obtained for a sample unit, either because the respondent refuses to cooperate or is inaccessible, a nonsampling error usually occurs for both list and area frame sampling. In either case an estimate must be made for each missing report. In this respect, area frame sampling, using the closed segment approach, has some advantage over list frame sampling. Through observation an enumerator can usually provide a good estimate of the number of livestock located on a tract of land operated by a person who does not provide information.

In the case of a list frame sample unit, estimation by observation is more difficult, since entire farm data are always required. If the missing report is encountered by mail or telephone interview, observation is usually impractical. Where observational data are not available, the present method is to use the average of other reports in the same stratum.

ANALYSIS OF SURVEY DATA

To make an evaluation of the net gain resulting from the use of an extreme operator list in conjunction with the JES or DES, one would need to measure both the decrease in sampling error and the increase in nonsampling error. Unfortunately this is not possible. We can estimate the reduction in sampling error, but sample survey results do not provide an estimate of the increase in nonsampling error.

Very little research has been conducted to measure the nonsampling error associated with using an E.O. list sample. However, during an analysis of alternative methods of overlap determination, editing errors made in carrying out the rules for handling partnership operations were recorded.[5/] This is only a portion of the possible total nonsampling error. For Ohio, the only state in both the earlier study and the present analysis, the June 1975 Multiple Frame Hog Survey, (MFHS) estimate had a 1.0 percent nonsampling error attributable to errors in editing E.O. reports. For the Multiple Frame Cattle Survey (MFCS), the error was less than 0.1 percent. We simply do not have a reliable indication of the total nonsampling error associated with E.O. sampling.

To obtain a measure of the reduction in sampling error attributable to using an E.O. list, data from four states were analyzed. During the 1975 JES, Ohio, Minnesota, Kansas and Illinois collected tract and entire farm cattle and hog data for all tract operators, including E.O.'s. Each of these states was using a new land use stratified area frame sample for the first time. Each tract could potentially have been operated by a person whose name was not on the MF list and thus be used for the MF nonoverlap estimate. This is a weighted segment estimate which requires entire farm data. Each tract questionnaire was coded to indicate whether or not it was for an E.O. In this manner the E.O. data were edited to zero for the area frame estimate as required for the JES screening estimator. Following the June survey, an area frame estimate was computed leaving the E.O. data in the area frame. This we refer to as a "complete" area frame estimate. It was not supplemented with a list frame sample estimate for the E.O.'s.

The results of this analysis are shown in detail in Tables 1-8 of the Appendix, and are summarized for the four states combined in Tables A, B and C. The complete area frame estimates were larger than the JES screening estimator for cattle in Ohio and Minnesota. The cattle estimates were smaller in Kansas and Illinois and the hog estimates were smaller for all states using the complete estimate. The four state total is slightly smaller for cattle but substantially smaller for hogs. The difference for each state as well as for the four state totals was within one sampling error.

The reduction in sampling error by using an E.O. list for cattle appears to be minimal for Ohio, Minnesota or Illinois. The reduction for Kansas is considerably larger. The reason the use of an E.O. list in Kansas is somewhat more successful than in the other three states is probably due to the larger number of large, concentrated cattle operations. The increase in sampling error for the weighted estimate for Kansas occurred because one nonresident tract operator in the sample operated a feedlot located outside the segment.

Table A--Comparison of Ohio, Minnesota, Kansas and Illinois Combined Cattle
and Calf estimates from the June Enumerative Survey Screening Estimator
and from the Complete Area Frame, June 1975.

Procedure	Domain	Direct Expansion	Sampling Error	Relative Sampling Error
		(000)	(000)	(%)
		TRACT		
Screening	Area[1]/	16,074.3	526.7	3.3
Estimator	E.O.	1,468.6	23.4	1.6
	Total	17,542.9	527.2	3.0
Complete Area Frame	Total	17,398.5	666.2	3.8
		ENTIRE FARM		
Screening	Area[1]/	15,364.6	869.4	5.7
Estimator	E.O.	1,468.6	23.4	1.6
	Total	16,833.2	869.6	5.2
Complete Area Frame	Total	16,337.8	957.1	5.9
		WEIGHTED		
Screening	Area[1]/	16,487.9	424.4	2.6
Estimator	E.O.	1,468.6	23.4	1.6
	Total	17,956.5	425.0	2.4
Complete Area Frame	Total	17,637.4	663.9	3.8

1/ Excluding overlap extreme operators

Table B--Comparison of Ohio, Minnesota, Kansas and Illinois, Combined Hog and
Pig estimates from the June Enumerative Survey Screening Estimator and
from the Complete Area Frame, June 1975.

Procedure	Domain	Direct Expansion	Sampling Error	Relative Sampling Error
		(000)	(000)	(%)
		TRACT		
Screening	Area[1]/	10,469.1	857.6	8.2
Estimator	E.O.	1,509.2	67.1	4.4
	Total	11,978.3	860.3	7.2
Complete Area Frame	Total	10,772.5	886.9	8.2
		ENTIRE FARM		
Screening	Area[1]/	10,168.6	852.7	8.4
Estimator	E.O.	1,509.2	67.1	4.4
	Total	11,677.8	855.3	7.3
Complete Area Frame	Total	10,543.2	886.5	8.4
		WEIGHTED		
Screening	Area[1]/	10,512.5	618.6	5.9
Estimator	E.O.	1,509.2	67.1	4.4
	Total	12,021.7	622.3	5.2
Complete Area Frame	Total	11,257.5	652.8	5.8

1/ Excluding overlap extreme operators

Table C--Complete Area Frame Cattle and Hog estimates as percent of June
Enumerative Survey screening estimates, Ohio, Minnesota, Kansas and
Illinois Combined, June 1975,

Survey Estimate	Ohio	Minn.	Kansas	Ill.	4-STATE TOTAL
	(%)	(%)	(%)	(%)	(%)
CATTLE					
Tract	102.3	103.3	95.1	99.0	99.2
Entire farm	103.4	101.4	90.5	98.4	97.1
Weighted	103.0	100.1	94.5	99.3	98.2
HOGS					
Tract	87.9	90.3	82.0	92.0	89.9
Entire Farm	86.7	91.4	89.6	90.7	90.3
Weighted	96.0	94.3	95.1	92.3	93.6

On July 1, 1975, about 14% of the total Kansas cattle inventory was cattle on feed. This compares with 6% for Minnesota, 7% for Ohio, and 12% for Illinois.[6/] In addition, Kansas feedlots tend to be larger in capacity. During the year 1975 Kansas had 131 feedlots each with a capacity in excess of 1,000 head. This number is more than the other three states combined. Also, Kansas had 26 lots each with a capacity of 16.000-31,999 head and 7 lots each with a capacity of 32,000 head and over. None of the other three states had feedlots with capacity of 16,000 head or more.[7/]

These factors relating to concentration of the item of interest suggest the complete area frame estimate can be expected to have a larger sampling error in Kansas relative to the other three states for estimating cattle. In other words, the potential gain from sampling from an E.O. cattle list are greater for Kansas than for Ohio, Minnesota or Illinois. The main benefit is probably attributable to having just the names of large feedlot operators on a list, not the names of all cattle E.O.'s.

The complete area frame estimate is not as precise for hogs as for cattle in any of the four states. This is probably due to there being substantially fewer hog farms than cattle farms and fewer total hogs than total cattle in each of these states. In general, hog operations tend to be more concentrated than cattle operations, especially cow-calf operations. Regardless of the type of hog operation (farrowing or feeding), a very large number of hogs can be concentrated on one tract. In the case of a cow-calf operation, the number of cattle located on one tract will usually be relatively stable, regardless of the size of the total operation. Therefore, the complete area frame tract estimate can be expected to perform better for cattle than hogs.

Estimates of the contribution of large concentrated hog operations to total hog inventories are not directly available for these states. The reduction in sampling error with the use of an E.O. hog list appears to be largest for Kansas and Illinois.

It is of concern that the complete area frame estimate of total hogs is consistently below that from the current procedure. If this relationship were found to exist for many states, one might conclude that we are not excluding all the E.O. hog data from the area expansion.

SUMMARY

This analysis of data from four states demonstrates that a more precise JES hog and cattle estimate is possible by using an E.O. list. However, the level of precision desired at the regional and national level is probably obtainable with a smaller E.O. list in some MF states. This smaller list would include only the names of the very largest livestock operators.

The use of an E.O. list increases nonsampling errors and creates additional expense. The more names there on a list, the more significant these factors become.

Based on this analysis we recommend the following:

1. More information be gathered to indicate the nonsampling error associated with using an E.O. list in conjunction with the JES and DES, and
2. Examine the effect of E.O. lists on sampling error on a state-by-state and specie-by-specie basis.

REFERENCES

1/ Raymond R. Bosecker and William F. Kelly, "Summary of Results from Nebraska Survey Concept Study," Sampling Studies Section, Sample Survey Research Branch, Research Division, Statistical Reporting Service, U.S. Department of Agriculture, 1975.

2/ Frederic A. Vogel, Raymond R. Bosecker and Dwight A. Rockwell, "Multiple Frame Livestock Surveys, An Evaluation of Alternative Nonoverlap Procedures," Sampling Studies Section, Sample Survey Research Branch, Research Division, Statistical Reporting Service, U.S. Department of Agriculture, 1976.

3/ Raymond R. Bosecker and William F. Kelly, op. cit., 1975.

4/ Frederic A. Vogel, Raymond R. Bosecker and Dwight A. Rockwell, op. cit., 1976.

5/ Frederic A. Vogel, Raymond R. Bosecker and Dwight A. Rockwell, op. cit. 1976.

6/ July 1, 1975 Cattle Inventory and Cattle on Feed Reports, Statistical Reporting Service, Crop Reporting Board, U.S. Department of Agriculture.

7/ January 1, 1976 Cattle on Feed Report, Statistical Reporting Service, Crop Reporting Board, U.S. Department of Agriculture.

THE EFFECT ON SAMPLING ERRORS RESULTING FROM

USING A LIVESTOCK LIST WITH AN AREA FRAME SAMPLE

APPENDIX

TABLES 1 - 8

Table 1--Comparison of Ohio Cattle and Calf estimates from the June Enumerative Survey Screening Estimator and from the Complete Area Frame, June 1975.

Procedure	Domain	Direct Expansion	Sampling Error	Relative Sampling Error
		(000)	(000)	(%)
		TRACT		
Screening Estimator	Area$^{1/}$	2,307.9	163.1	7.1
	E.O.	79.5	4.3	5.4
	Total	2,387.4	163.2	6.8
Complete Area Frame	Total	2,441.4	179.2	7.3
		ENTIRE FARM		
Screening Estimator	Area$^{1/}$	2,306.9	223.9	9.7
	E.O.	79.5	4.3	5.4
	Total	2,386.4	223.9	9.4
Complete Area Frame	Total	2,468.7	269.6	10.9
		WEIGHTED		
Screening Estimator	Area$^{1/}$	2,454.1	138.3	5.6
	E.O.	79.5	4.3	5.4
	Total	2,533.6	138.4	5.5
Complete Area Frame	Total	2,608.6	153.8	5.9

1/ Excluding overlap extreme operators

Table 2--Comparison of Minnesota <u>Cattle and Calf</u> estimates from the June
Enumerative Survey Screening Estimator and from the Complete
Area Frame, June 1975.

Procedure	Domain	Direct Expansion	Sampling Error	Relative Sampling Error
		(000)	(000)	(%)
		TRACT		
Screening Estimator	Area[1]/	4,763.4	295.9	6.2
	E.O.	180.2	8.4	4.7
	Total	4,943.6	296.0	6.0
Complete Area Frame	Total	5,105.5	338.4	6.6
		ENTIRE FARM		
Screening Estimator	Area[1]/	4,588.4	296.3	6.5
	E.O.	180.2	8.4	4.7
	Total	4,768.6	296.4	6.2
Complete Area Frame	Total	4,836.0	337.3	7.0
		WEIGHTED		
Screening Estimator	Area[1]/	4,645.2	237.2	5.1
	E.O.	180.2	8.4	4.7
	Total	4,825.4	237.3	4.9
Complete Area Frame	Total	4,832.4	248.2	5.1

[1]/ Excluding overlap extreme operators

Table 3--Comparison of Kansas Cattle and Calf estimates from the June Enumerative Survey Screening Estimator and from the Complete Area Frame, June 1975.

Procedure	Domain	Direct Expansion	Sampling Error	Relative Sampling Error
		(000)	(000)	(%)
		TRACT		
Screening Estimator	Area$\frac{1}{}$	5,394.7	298.7	5.5
	E.O.	1,151.8	21.0	1.8
	Total	6,546.5	299.4	4.6
Complete Area Frame	Total	6,224.8	471.9	7.6
		ENTIRE FARM		
Screening Estimator	Area$\frac{1}{}$	5,047.9	725.2	14.4
	E.O.	1,151.8	21.0	1.8
	Total	6,199.7	725.5	11.7
Complete Area Frame	Total	5,611.7	798.5	14.2
		WEIGHTED		
Screening Estimator	Area$\frac{1}{}$	5,619.3	252.5	4.5
	E.O.	1,151.8	21.0	1.8
	Total	6,771.1	253.4	3.7
Complete Area Frame	Total	6,398.0	560.4	8.8

1/ Excluding overlap extreme operators

Table 4--Comparison of Illinois <u>Cattle and Calf</u> estimates from the June
Enumerative Survey Screening Estimator and from the Complete Area
Frame, June 1975.

Procedure	Domain	Direct Expansion	Sampling Error	Relative Sampling Error
		(000)	(000)	(%)
		TRACT		
Screening	Area[1]/	3,608.3	272.1	7.5
Estimator	E.O.	57.1	4.3	7.5
	Total	3,665.4	272.1	7.4
Complete Area Frame	Total	3,626.8	273.0	7.5
		ENTIRE FARM		
Screening	Area[1]/	3,421.4	303.2	8.9
Estimator	E.O.	57.1	4.3	7.5
	Total	3,478.5	303.2	8.7
Complete Area Frame	Total	3,421.4[2]/	303.2	8.9
		WEIGHTED		
Screening	Area[1]/	3,769.3	202.4	5.4
Estimator	E.O.	57.1	4.3	7.5
	Total	3,826.4	202.4	5.3
Complete Area Frame	Total	3,798.4	203.5	5.4

1/ Excluding overlap extreme operators

2/ No E.O. resident tract operators were in the area frame sample.

Table 5--Comparison of Ohio Hog and Pig estimates from the June Enumerative Survey Screening Estimator and from the Complete Area Frame, June 1975.

Procedure	Domain	Direct Expansion	Sampling Error	Relative Sampling Error
		(000)	(000)	(%)
		TRACT		
Screening Estimator	Area[1/]	1,392.6	321.0	23.1
	E.O.	191.2	19.7	10.3
	Total	1,583.8	321.6	20.3
Complete Area Frame	Total	1,392.6[2/]	321.0	23.1
		ENTIRE FARM		
Screening Estimator	Area[1/]	1,251.5	286.2	22.9
	E.O.	191.2	19.7	10.3
	Total	1,442.7	286.9	20.0
Complete Area Frame	Total	1,251.5[2/]	286.2	22.9
		WEIGHTED		
Screening Estimator	Area[1/]	1,533.7	225.5	14.7
	E.O.	191.2	19.7	10.3
	Total	1,724.9	226.4	13.1
Complete Area Frame	Total	1,655.3	232.1	14.0

1/ Excluding overlap extreme operators

2/ No E.O. resident tract operators were in the area frame sample.

Table 6--Comparison of Minnesota <u>hog and pig</u> estimates from the June Enumerative Survey Screening Estimator and from the Complete Area Frame, June 1975.

Procedure	Domain	Direct Expansion	Sampling Error	Relative Sampling Error
		(000)	(000)	(%)
		TRACT		
Screening Estimator	Area[1/]	2,546.0	314.5	12.4
	E.O.	272.1	18.7	6.9
	Total	2,818.1	315.1	11.2
Complete Area Frame	Total	2,546.0[2/]	314.5	12.4
		ENTIRE FARM		
Screening Estimator	Area[1/]	2,891.3	363.4	12.6
	E O.	272.1	18.7	6.9
	Total	3,163.4	363.9	11.5
Complete Area Frame	Total	2,891.3[2/]	363.4	12.6
		WEIGHTED		
Screening Estimator	Area[1/]	2,908.5	258.1	8.9
	E.O.	272.1	18.7	6.9
	Total	3,180.6	258.8	8.1
Complete Area Frame	Total	2,997.9	265.9	8.9

[1/] Excluding overlap extreme operators

[2/] No E.O. resident tract operators were in the area frame sample.

Table 7--Comparison of Kansas Hog and Pig estimates from the June Enumerative
Survey Screening Estimator and from the Complete Area Frame, June 1975.

Procedure	Domain	Direct Expansion	Sampling Error	Relative Sampling Error
		(000)	(000)	(%)
		TRACT		
Screening	Area[1/]	1,055.4	278.0	26.3
Estimator	E.O.	294.5	21.0	7.1
	Total	1,349.9	278.8	20.6
Complete Area Frame	Total	1,107.6	282.9	25.5
		ENTIRE FARM		
Screening	Area[1/]	993.0	227.3	22.9
Estimator	E.O.	294.5	21.0	7.1
	Total	1,287.5	228.3	17.7
Complete Area Frame	Total	1,153.8	255.9	22.2
		WEIGHTED		
Screening	Area[1/]	1,091.6	242.8	22.2
Estimator	E.O.	294.5	21.0	7.1
	Total	1,386.1	243.7	17.6
Complete Area Frame	Total	1,317.6	260.1	19.7

1/ Excluding overlap extreme operators

Table 8--Comparison of Illinois Hog and Pig estimates from the June Enumerative Survey Screening Estimator and from the Complete Area Frame, June 1975.

Procedure	Domain	Direct Expansion	Sampling Error	Relative Sampling Error
		(000)	(000)	(%)
		TRACT		
Screening Estimator	Area[1]/	5,475.1	675.5	12.3
	E.O.	751.4	57.7	7.7
	Total	6,226.5	678.0	10.9
Complete Area Frame	Total	5,726.3	710.3	12.4
		ENTIRE FARM		
Screening Estimator	Area[1]/	5,032.8	679.3	13.5
	E.O.	751.4	57.7	7.7
	Total	5,784.2	681.7	11.8
Complete Area Frame	Total	5,246.6	711.6	13.6
		WEIGHTED		
Screening Estimator	Area[1]/	4,978.7	454.2	9.1
	E.O.	751.4	57.7	7.7
	Total	5,730.1	457.9	8.0
Complete Area Frame	Total	5,286.7	483.7	9.2

[1]/ Excluding overlap extreme operators