Research Report

The Feasibility of Obtaining the Social Security Number from Farm

Operators and Its Use in Identifying the "Overlap" Domain in Multiple

Frame Sampling

by

Robert W. Young

Research and Development Branch

Standards and Research Division

Statistical Reporting Service

April 1971

The Feasibility of Obtaining the Social Security Number from Farm Operators and Its Use in Identifying the "Overlap" Domain in Multiple Frame Sampling

SUMMARY

1. Farmers can report their SS number with relative ease. The rates were 87.4 percent from the area frame and 88.5 percent from the list frame. The reporting accuracy was quite high at 94.5 percent. This comparison could only be made on the list respondents. Tables 1 and 2 show these results.

2. About 10 percent (31 reports) of the interviewed respondents either hesitated or showed hostility toward the SS number question. Of the 31 reports, only 3 did not report their number.

3 The SS number match confirmed most of the possible matches made by conventional matching methods based on name and address. In fact, 95.9 percent of the possible matches were confirmed as positive matches.

4. Using the SS number as the only matching criterion resulted in fewer positive matches than conventional methods. But when both methods are used in a complementary manner the total matches exceeds the two separate methods. There would be a gain from both the operational and theoretical standpoint in using the SS number.

5. Enumerators differed in their abilities to obtain SS numbers.

6. The SS matching procedure could be utilized for any multiple frame survey if the list frame has the SS number for each name. This procedure will be used in Tennessee on the 1971 JES on a pilot project basis. Cost analysis for identifying overlap will be obtained for this state and further evaluations will be made.

BACKGROUND

Multiple frame sampling requires that every unit in the population of interest belong to at least one of the frames, and that it be possible to record, for each sampled unit, whether or not it belongs to the other frame. The latter requirement is the underlying reason for this project to test whether the SS number can serve as a unique link between two frames. The current procedure for identifying which units belong in both frames, often referred to as "overlap," consists of an observational check of names and addresses of the sampled units from one frame against the complete list of the other frame. This matching of units is usually done manually, but it can be done on a computer. The current procedures result in a relatively high number of possible matches. Each of the possible matches must be categorized as being in one frame or both. Often a

judgment decision has to be made as to whether the units are the same and in some cases a personal or phone contact is required to make this decision. With ample resources, including time, we can resolve the possible matches. However, it appears from past work that name and address alone does not provide adequate information for determining overlap. Other techniques are more satisfactory but are more expensive to carry out. The study was undertaken in Oklahoma and Tennessee to determine farmer's ability to report their SS number, and to explore the possibilities of using it to assist in identifying overlap in multiple frame estimation.

To use the SS number effectively in the matching process, two conditions must be met. First, one frame, generally a list, must include the SS number for each reporting unit. Second, the SS numbers for respondents must be obtained for members of the other frame, usually an area frame. If these two conditions can be satisfied, the identification of which units are the same (make up the overlap domain) could be achieved by matching the SS numbers. The purpose of the project was to test whether the above conditions can be achieved under survey conditions. Current developments indicate that the farmer's SS number can be obtained for list frames in the future. The reporting unit's number could either be the operator's personal SS number or it could be the employer's identification number in the case of a corporation.

In this study, the second condition, obtaining the SS number in interviews, was studied in depth. Analysis was made of (1) the feasibility of asking SS numbers, (2) the reporting accuracy, and (3) the use of the number in identifying overlap.

THE SURVEY

An area frame and a list frame were used for the frames in two Crop Reporting Districts of Oklahoma and Tennessee. A sample of ten area segments was selected from each Crop Reporting District from the area frame. Within each district, the segments were randomly assigned to two enumerators. All tract operators within each segment were interviewed using a questionnaire similar to that of the June Enumerative Survey (JES). The questionnaire included a question to obtain the farm operator's SS number.

The list frame was obtained from the Agricultural Stabilization and Conservation Service (ASCS) records. Two printouts were obtained for each county with the units arrayed (1) alphabetically on the first four letters of the last name, and (2) ascending SS numbers. The printouts contained the name, address and SS number for each applicant. Questionnaires were mailed to the list frame sample of 536 people in the two states. An interview follow-up was made of non-respondents. Enumerator assignments were random in both frames so that enumerator effects could be examined. A "respondent reaction form" was completed by the enumerator for each respondent immediately following the interview. This form was used to evaluate the respondent's reaction about being asked to report his SS number.

ANALYSIS

The segment identification sheet provided the name and address of each tract operator. This information was used to match the tract operators against the list frame using current methods. The SS numbers obtained during the interview were used independently in determining the overlap portion. The list sample check data (that included the SS number) provided a validity check on the reporting accuracy of their SS number.

In preparing multiple frame estimates, it is assumed that the list frame direct expansions are the same as the area expansions for the overlap domain. Because variation in the list expansion is usually smaller than the area expansion, the list estimate is usually used as the estimate of the overlap domain. Hence, the area frame provides an estimate of the non-overlap domain, i.e., provides an estimate for list incompleteness.

Most of the analysis dealt with attribute data. The proportion of respondents reporting SS numbers and the fraction of the number being reported correct were two of the events studied. Probability statements can be made by using the usual binomial formula:

$$\hat{p} = \sum_{i=1}^{n} a_i/n$$

where a_i is either 1 or 0 for the ith observation and n is the number of observations.

$$\hat{\sigma} = \sqrt{\hat{p}\,\hat{q}\,/\,(n-1)}$$

where $\hat{\sigma}$ is the estimated standard deviation of the proportion $\hat{p}$ and $\hat{q} = 1-\hat{p}$. Multiple frame estimates were not pertinent to this project and not generated. A nested analysis of variance was used to test for variability between enumerators. The model used was:

$$Y_{ijk} = \mu + a_i + b_{j(i)} + \ell_{k(ij)}$$

where Y_{ijk} is the observation from the kth sampling unit by the jth enumerator from the ith area. The following assumptions were made:

$$a_i \sim (0,\ \sigma_a^2),\ b_{j(i)} \sim N\,(0,\ \sigma_b^2),\ \text{and } e_{k(ij)} \sim N\,(0,\ \sigma^2)\ .$$

REPORTING ABILITY AND ACCURACY

Tables 1 and 2 summarize the farmer's ability to accurately report their
SS numbers. Farmers reported their SS number with relative ease. About
88 percent of the respondents reported their SS numbers. The individual
rates were 87.4 percent for the area frame and 88.5 percent for the list
frame. These rates were based on the number of reports with SS numbers
reported, divided by the number of reports from farm operators. The sample
from the ASCS list frame had about 15 percent nonfarm operators. These
returns were handled separately and are noted in Table 1. Table 3 contains
probability statements with associated standard errors about events from
Tables 1 and 2. The lowest sampling errors were associated with reporting
the correct number. This error was 1.6 percent for both states.

The accuracy of the reported SS number was determined by comparing the
number reported on the questionnaire to the number obtained from the ASCS
list for the corresponding person. This check could only be made on the
list sample. The accuracy rate was 94.4 percent in Tennessee and 94.7
percent in Oklahoma. There were 22 reports that were different with seven
having an error in one digit. There are several possible reasons for these
one digit errors: either the ASCS list is wrong, the enumerator wrongly
recorded the number, or the respondent errored in reporting the number.
In one case, the ASCS listing was known to be in error. During the enumerator
training school, an enumerator found his name on one of the ASCS printouts
and noted that his recorded SS number was off by one digit. The inability
to interview the correct respondents accounted for 9 other errors. It was
possible that the farm operator's wife was listed on the ASCS list, but the
interview took place with her husband. The husband considered himself as
the farm operator and reported his SS number.

Table 1.--Reporting ability and accuracy of social security numbers by type of response, state and district, list sample

State, district, or type of response	All reports			Reports including social security number				
	Total	Zero report	Number of farm operators	Total	Percent of all reports	Percent of farm operators	Number correct	Percent correct
	Number	Number	Number	Number	Percent	Percent	Number	Percent
TENNESSEE								
District 2								
Mail	86	12	74	69	80.2	93.2	67	97.1
Interview	48	13	35	28	58.3	80.0	26	92.9
Total	134	25	109	97	72.4	89.0	93	95.9
District 4								
Mail	96	8	88	84	87.5	95.4	79	94.0
Interview	49	11	38	33	67.3	86.8	30	90.9
Total	145	19	126	117	80.7	92.9	109	93.2
District 2 & 4								
Mail	182	20	162	153	84.1	94.4	146	95.4
Interview	97	24	73	61	62.9	83.6	56	91.8
Total	279	44	235	214	76.7	91.1	202	94.4
OKLAHOMA								
District 4								
Mail	81	16	65	55	67.9	84.6	52	94.5
Interview	46	8	38	35	76.1	92.1	33	94.3
Total	127	24	103	90	70.9	87.4	85	94.4
District 6								
Mail	82	14	68	60	73.2	88.2	56	93.3
Interview	48	0	48	38	79.2	79.2	37	97.4
Total	130	14	116	98	75.4	84.5	93	94.9
District 4 & 6								
Mail	163	30	133	115	70.6	86.5	108	93.9
Interview	94	8	86	73	77.7	84.9	70	95.9
Total	257	38	219	188	73.2	85.8	178	94.7
2 State Total	536	82	454	402	75.0	88.5	380	94.5

Table 2.--Reporting social security number from the area frame, by resident and non-resident operator, by district within states

State and area	Resident farm operator			Non-resident farm operator			Total		
	Tracts	Social security number reported		Tracts	Social security number reported		Tracts	Social security number reported	
		Total	Percent		Total	Percent		Total	Percent
	Number	Number	Percent	Number	Number	Percent	Number	Number	Percent
Tennessee									
District 2	12	12	100.0	34	29	85.3	46	41	89.1
District 4	22	19	86.4	24	15	62.5	46	34	73.9
Total	34	31	91.2	58	44	75.9	92	75	81.5
Oklahoma									
District 4	7	6	85.7	23	23	100.0	30	29	96.7
District 6	14	14	100.0	23	21	91.3	37	35	94.6
Total	21	20	95.2	46	44	95.7	67	64	95.5
2 State Total	55	51	92.7	104	88	84.6	159	139	87.4

Table 3.--Probability statements for certain events from Tables 1 and 2, for
Tennessee and Oklahoma

Event	Tennessee		Oklahoma	
	$\hat{P}$	$\hat{\sigma}$	$\hat{P}$	$\hat{\sigma}$
Ability to report social security number				
List mail response	.944	.018	.865	.030
List non-response interview	.836	.044	.849	.039
Area frame interview response	.815	.041	.955	.026
Ability to report correct social security number - list frame	.944	.016	.947	.016

RESPONDENT'S REACTION TO ASKING SOCIAL SECURITY NUMBER

The enumerators completed a "Respondent's Reaction" form (Appendix 1) following
each interview. Tables 4, 5, 6, and 7 show the results of this form. About 83
percent of the respondents expressed no visible reaction when asked to report
their SS number. About 10 percent showed some sort of hostility or hesitated
in reporting their SS number. For this group of 31 reports, only 3 refused to
report their SS number.

Table 4.--Respondent's reaction to the social security number question, list
and area sample interviews, Tennessee

Reactions	List sample		Area sample		Total	
	Frequency	Relative frequency	Frequency	Relative frequency	Frequency	Relative frequency
	Number	Percent	Number	Percent	Number	Percent
Hesitated	6	7.9	9	12.3	15	10.1
Showed hostility	1	1.3	2	2.7	3	2.0
Irrelevant conversation	3	4.0	3	4.1	6	4.0
Laughed, joked	---	---.	2	2.7	2	1.3
Questioned reason for asking	8	10.5	8	11.0	16	10.8
No visible reaction	56	73.7	45	61.7	101	67.8
Other	2	2.6	4	5.5	6	4.0
Total answers	76	100.0	73	100.0	149	100.0

Table 5.—Respondent's reaction to social security number question, list and area sample interviews, Oklahoma

Reactions	List sample		Area sample		Total	
	Frequency	Relative frequency	Frequency	Relative frequency	Frequency	Relative frequency
	Number	Percent	Number	Percent	Number	Percent
Hesitated	9	11.3	1	1.4	10	6.6
Showed hostility	2	2.5	1	1.4	3	2.0
Irrelevant conversation	---	---	---	---	---	---
Laughed, joked	---	---	---	---	---	---
Questioned reason for asking	---	---	4	5.5	4	2.6
No visible reaction	69	86.2	66	91.7	135	88.8
Other	---	---	---	---	---	---
Total answers	80	100.0	72	100.0	152	100.0

Table 6.—Reasons given for not reporting social security number, mail response and interview, Tennessee

Area and type of response	Not reporting social security number												
	Total	Non-farm zero report		Refused to give any data		Refused to give SS No.		Different respondent		Did not know number		Left blank	
	Number	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent
District 2													
Mail response	17	12	70.6	—	—	—	—	1	5.9	—	—	4	23.5
Interview	24	13	54.2	—	—	—	—	6	25.0	4	16.7	1	4.2
Total	41	25	61.0	—	—	—	—	7	17.1	4	9.8	5	12.2
District 4													
Mail response	12	8	66.7	—	—	—	—	—	—	—	—	4	33.3
Interview	27	11	40.7	—	—	2	7.4	9	33.3	2	7.4	3	11.1
Total	39	19	48.7	—	—	2	5.1	9	23.1	2	5.1	7	17.9
District 2 & 4													
Mail response	29	20	69.0	—	—	—	—	1	3.4	—	—	8	27.6
Interview	51	24	47.1	—	—	2	3.9	15	29.4	6	11.8	4	7.8
Total	80	44	55.0	—	—	2	2.5	16	20.0	6	7.5	12	15.0

Table 7.-- Reasons given for not reporting social security number, mail response and interview, Oklahoma

Area and type of response	Total	Non-farm zero report		Refused to give any data		Refused to give SS No.		Different respondent		Did not know number		Left blank	
	Number	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent
District 4													
Mail response	26	16	61.5	--	--	--	--	2	7.7	--	--	8	30.8
Interview	12	8	66.7	1	8.3	2	16.7	--	--	--	--	1	8.3
Total	38	24	63.2	1	2.6	2	5.3	2	5.3	--	--	9	23.7
District 6													
Mail response	22	14	63.6	--	--	--	--	1	4.5	--	--	7	31.8
Interview	11	--	--	2	18.2	1	9.1	2	18.2	4	36.4	2	18.2
Total	33	14	42.4	2	6.1	1	3.0	3	9.1	4	12.1	9	27.3
District 4 & 6													
Mail response	48	30	62.5	--	--	--	--	3	6.3	--	--	15	31.3
Interview	23	8	34.8	3	13.0	3	13.0	2	8.7	4	17.4	3	13.0
Total	71	38	53.5	3	4.2	3	4.2	5	7.0	4	5.6	18	25.4

MATCHING AREA FRAME TRACTS TO THE LIST UNIVERSE

Three independent matches were made between the tract operators in the area frame and the list frame operators. The matches were (1) using the name and address section of the segment identification sheet against the same items from the list, (2) using Section E of the tract questionnaire against the list, and (3) using the reported SS number against the list. The segment identification sheet asked for the tract operator's name and mailing address and four screening questions to determine if he was a farm operator or not. Section E of the questionnaire obtained more detail for farm operator identification. Items asked were the farm operator's name, nickname (if any), mailing address, ranch or farm name (if any), county, and telephone number. The matching criteria for the first two methods were:

(1) Good match - name and address were the same.

(2) Possible match

    (a) Same name but incomplete mailing address on either the list or questionnaire.

    (b) Same name but different mailing address with similar zip code. The respondent lives between two communities and gave the enumerator one city as a mailing address and the ASCS another.

    (c) Same address but different or incomplete name - middle name or initial left out, or Mrs. A. B. Jones and Stella Jones.

    (d) Same address, name incorrectly spelled.

(3) No match - None of the above criteria can be met.

The criteria using the SS number as the matching agent are:

(1) Perfect match - nine digits agree.

(2) No match

    (a) SS number given but not on list.

    (b) SS number not obtained from interview.

Tables 8 thru 11 compare the results of two of the matching methods. Tables 8 and 10 show data for the overlap domain for each state by match status. Tables 9 and 11 show the non-match or non-overlap portion. They are the compliments of Tables 8 and 10. For this particular survey, the gains in using the more detailed information from Section E of the questionnaire were minimal and not summarized separately. The matching shown is for the

name and address match only, SS number match only, and a composite match of both. All of the possible matches using the name and address method were assigned to the overlap domain. There were 49 tracts in Tennessee, and 31 tracts in Oklahoma in this category.

Of the possible matches, 96 percent were actually confirmed as being positive matches using the SS number, therefore, eliminating the judgment decision. Out of the 80 possible matches, 73 gave SS numbers of which 70 were matched from the ASCS list. The possible matches require a judgment decision and, depending on the situation, could have been included or excluded in the overlap domain using current procedures.

Table 8.—Actual and expanded data for the overlap domain from the area frame, by match status, Tennessee

Item		Composite match	Name and address match	Social security match
		Number	Number	Number
Tracts	– Actual	73	72	63
	Expanded	97,647	96,435	83,808
Hogs and pigs	– Actual	419	411	396
	Expanded	564,002	554,306	529,534
Cattle and calves	– Actual	708	707	584
	Expanded	1,007,701	1,006,489	826,460
Hens and pullets of laying age	– Actual	517	502	402
	Expanded	712,871	694,691	544,831
		Acres	Acres	Acres
Corn	– Actual	279	279	269
	Expanded	343,880	343,880	331,760
Sorghum	– Actual	18	18	18
	Expanded	21,816	21,816	21,816
Soybeans	– Actual	489	489	453
	Expanded	597,209	597,209	549,278
Cotton	– Actual	230	230	219
	Expanded	278,862	278,862	266,894
Wheat	– Actual	55	55	5
	Expanded	66,660	66,660	6,060

Table 9.--Actual and expanded data for the <u>non-overlap</u> domain from the area frame, by match status, Tennessee

Item		Composite non-match	Name and address non-match	Social security non-match
		Number	Number	Number
Tracts	– Actual	19	20	29
	Expanded	27,041	28,253	40,880
Hogs and pigs	– Actual	18	26	41
	Expanded	26,115	35,811	60,583
Cattle and calves	– Actual	103	104	227
	Expanded	150,630	151,842	331,871
Hens and pullets of laying age	– Actual	7,022	7,037	7,137
	Expanded	10,232,638	10,545,648	10,695,508
		Acres	Acres	Acres
Corn	– Actual	0	0	10
	Expanded	0	0	12,120
Sorghum	– Actual	0	0	0
	Expanded	0	0	0
Soybeans	– Actual	0	0	36
	Expanded	0	0	47,931
Cotton	– Actual	6	6	17
	Expanded	8,151	8,151	20,119
Wheat	– Actual	0	0	50
	Expanded	0	0	60,600

Table 10.--Actual and expanded data for the overlap domain from the area
frame, by match status, Oklahoma

Item		Composite match	Name and address match	Social security match
		Number	Number	Number
Tracts	– Actual	43	42	39
	Expanded	32,049	31,460	28,901
Hogs and pigs	– Actual	0	0	0
	Expanded	0	0	0
Cattle and calves	– Actual	1,111	1,083	1,067
	Expanded	895,507	879,037	852,140
Hens and pullets of laying age	– Actual	96	96	96
	Expanded	79,517	79,517	79,517
		Acres	Acres	Acres
Corn	– Actual	0	0	0
	Expanded	0	0	0
Sorghum	– Actual	207	207	117
	Expanded	121,757	121,757	68,819
Soybeans	– Actual	0	0	0
	Expanded	0	0	0
Cotton	– Actual	215	215	190
	Expanded	126,522	126,522	111,817
Wheat	– Actual	342	342	342
	Expanded	201,164	201,164	201,164

Table 11.--Actual and expanded data for the non-overlap domain from the area
frame, by match status, Oklahoma

Item		Composite non-match	Name and address non-match	Social security non-match
		Number	Number	Number
Tracts	- Actual	24	25	28
	Expanded	22,064	22,653	25,212
Hogs and pigs	- Actual	10	10	150
	Expanded	8,267	8,267	69,076
Cattle and calves	- Actual	424	452	29
	Expanded	404,780	421,250	17,058
Hens and pullets of laying age	- Actual	249	249	0
	Expanded	220,379	220,379	0
		Acres	Acres	Acres
Corn	- Actual	0	0	0
	Expanded	0	0	0
Sorghum	- Actual	60	60	10
	Expanded	59,136	59,136	8,267
Cotton	- Actual	4	4	468
	Expanded	2,353	2,353	448,147
Wheat	- Actual	0	0	249
	Expanded	0	0	220,379

Table 12.—Matching area tract operators to the list frame, expressed as a percent of total tracts, by matching method, by state

Method	Tennessee	Oklahoma	Total
	Percent	Percent	Percent
A. Name and address matching			
1. Positive matches	25.0	16.4	21.4
2. Possible matches	53.3	46.3	50.3
3. Probability of a positive match given a possible match	95.3	96.7	95.9
4. Line 2 X line 3	50.8	44.8	48.2
5. Total expected matches line 1 + line 4	75.8	61.2	69.6
B. Social Security matching followed by name/address matching			
1. Positive match from SS No.	68.5	58.2	64.2
2. Positive match from name and address (SS No. not given or in error)	2.2	3.0	2.5
3. Possible match (SS No. not given or in error)	8.7	3.0	6.3
4. Probability of a positive match given a possible match	95.3	96.7	95.9
5. Line 3 X line 4	8.3	2.9	6.0
6. Total expected matches line 1 + line 2 + line 5	79.0	64.1	72.7
C. Gains from Social Security number			
Difference between the two methods	3.2	2.9	3.1

Table 12 shows the results, on a percentage of total tracts basis, in matching area tracts. Part A of the table represents the current matching methods. Part B represents the tracts matched by SS numbers first and then using the name and address information for the remaining tracts. This composite matching method results in about a three percent gain from using name and address only. It might be concluded that the three percent gain resulting in the use of SS numbers is nominal; however, from an operational standpoint the matching process would almost be instantaneous. From the survey 64.2 percent of the tracts were categorized as positive matches without any need for further verification. The remaining 35.8 percent of the tracts would be matched using the name and address information.

ENUMERATOR EFFECTS PROVE TO BE SIGNIFICANT

The secondary objective of testing enumerator differences in obtaining SS numbers proved to be significant at the five percent level. The nested analysis of variance is given in Tables 13 and 14.

The corresponding tabular F values are as follows: (1) List frame, 2.40 at the five percent level and 3.40 at the one percent level with (4, 183) degrees of freedom, and (2) area frame, 2.67 at the five percent level and 3.97 at the one percent level with (4, 32) degrees of freedom. We can conclude that the enumerator component of variance is significant at the five percent level in the list sample and is significant at the one percent level in the area frame sample. The difference can be traced back to one enumerator who failed to make contact with the proper respondents. The respondent in many cases was a neighbor or the farm operator's wife. This type of situation could be corrected by placing greater emphasis requiring the enumerator to make contact with the farm operator.

Table 13.—Nested analysis of variance testing the hypothesis of no differences between enumerators, list frame

Source	df	Sum of squares	Mean squares	F
State	1	0.902	0.902	5.638
District/state	2	0.319	0.160	0.292
Enumerator/district	4	2.194	0.548	2.727*
Error	183	36.167	0.198	
Total	190	39.582	0.208	

* Significant at $\sigma = .05$

Table 14.—Nested analysis of variance testing the hypothesis of no differences between enumerators, area frame

Source	df	Sum of squares	Mean squares	F
State	1	0.1273	0.127	2.153
District/state	2	0.1181	0.059	0.670
Enumerator/district	4	0.3501	0.088	6.286**
Error	32	0.4530	0.014	
Total	39	1.0485	0.027	

** Significant at $\sigma = .01$

U. S. Department of Agriculture
Statistical Reporting Service
Data Collection Branch

RECORDING FORM FOR RESPONDENT'S REACTION TO SOCIAL SECURITY NUMBER QUESTION

Segment _____ Tract _____

or List Sample-Number _____

Complete this form as soon as you have left the respondent.

1. Respondent's reaction to social security number question (Check one or more):

( ) Hesitated
( ) Showed hostility
( ) Irrelevant conversation
( ) Laughed, joked
( ) Questioned reason for asking
( ) No visible reactions
( ) Other (describe) _____

_____

2. Respondent's answer to social security number question (Check one or more if appropriate).

( ) Didn't know the number
( ) Didn't have a social security number
( ) Referred to card in billfold
( ) Gave from memory

    ( ) easily
    ( ) with difficulty
    ( ) other description _____

    _____

( ) Referred to other record (describe) _____

_____

( ) Consulted other person
( ) Refused to answer (reasons given) _____

_____

Enumerator _____