EFFECT OF THE 1978 CENSUS OF AGRICULTURE AREA SAMPLE ON CENSUS ESTIMATES

David D. Chapman, U.S. Bureau of the Census

1. INTRODUCTION

The 1978 Census of Agriculture was the first census of agriculture since 1964 which utilized an area sample. This paper discusses the census of agriculture coverage problem, describes census methodology, shows the improvement the area sample made in census estimates, evaluates county information and summarizes the effect of the area sample. All information in this paper is taken directly from published reports.

The major purpose of the 1978 Census of Agriculture was to make uniform and accurate estimates of agriculture activity and productivity for the United States and each county, state, and region in the United States. The same farm definition and a uniform data collection methodology was used at all geographical levels.

For statistical purposes a farm is defined as any place that had gross agricultural sales of $1,000 or more of agricultural products. This definition does not include as farms small agriculture operations formerly included in the census under the old 1959 definition. The 1959 definition counts as a farm any place with less than 10 acres from which $250 or more of agricultural products were sold, or any place of 10 acres or more from which $50 or more of agricultural products were sold.

2. CENSUS QUALITY

An evaluation of coverage has been conducted for each census of agriculture since 1945. In the 1969 and 1974 Censuses of Agriculture, mail list incompleteness and processing error resulted in a net underenumeration of farms. A small but significant number of farms were missed during these census enumerations. In 1978 a direct enumeration area sample was used to improve estimates at the state level and higher. State and county estimates in 1978 had a different level of quality. Table A gives the magnitude of this net underenumeration. A detailed examination of the quality of coverage of the 1969, 1974, and 1978 Censuses of Agriculture is given in special evaluation reports (U.S. Bureau of the Census, 1969; U.S. Bureau of the Census, 1974; and U.S. Bureau of the Census, 1978).

Table A. Percent Net Underenumeration in 1969, 1974, and 1978 Censuses of Agriculture at State and County Levels using the 1959 and the 1978 Farm Definitions.

Sales	1969 state/ county	1974 state/ county	1978	
			County	State
1959 Farm Definition				
All farms	15.0	14.3	18.3	2.9
Sales of $2,500 and over	3.3	4.7	2.4	2.4
Sales of under $2,500	31.6	32.8	39.2	3.6
1978 Farm Definition				
All farms	NA	10.7	12.0	3.4
Sales of $2,500 and over	NA	4.7	6.1	2.4
$40,000 and over	NA	0.1	1.2	0.5
$10,000 to $39,999	NA	3.9	6.1	3.4
$ 2,500 to $ 9,999	NA	12.0	10.2	3.2
Under $2,500	NA	25.9	29.5	6.2

3. 1978 CENSUS OF AGRICULTURE DATA COLLECTION PROGRAM

The 1978 Census of Agriculture was collected using both a mail list and a direct enumeration area sample. The mail list produced census estimates at the county level. The nonmatched area sample produced estimates of the number and characteristics of farms not on the mail list at the state level. Mail list and direct enumeration area sample data were used together to make state, division, region, and United States estimates.

3.1 1978 Mail List Enumeration

The primary census of agriculture data collection method since 1969 has been self-enumeration using a comprehensive mail list.

Since the final mail list was the result of combining lists of names and addresses, substantial duplication exists. Unduplication of these lists was conducted through a computer comparison and clerical review of records. Census report forms were mailed during the last week of December 1978 to all individuals on the mail list. From February through August 1979, additional mailings were sent primarily to nonrespondents and to new tenants and successors identified during report form processing.

3.2 1978 Census of Agriculture Area Sample

The 1978 Census of Agriculture Area Sample was a stratified one-stage cluster sample. Clusters were selected independently from each stratum with equal probability using a systematic sampling plan. The sampling frame for the 1978 Census of Agriculture Area Sample was all enumeration districts from the 1970 Census of Population and Housing and their characteristics and maps. The area sample was selected only from rural areas. Excluding urban areas from the survey was expected to exclude approximately one percent of the farms from the combined mail list and area sample and resulted in a major savings in costs. In the 1978 Census of Agriculture Area Sample, units are stratified by estimated farm density. A systematic sample of segments was selected from each stratum.

During enumeration, the enumerator located the boundaries of each area segment, located all dwelling units and listed the head of each household, group quarters and vacant structures. A series of preliminary or screening questions was asked of each head of household for every member of a household or group quarters. A complete agriculture census questionnaire was obtained from all individuals identified as a "potential" farm operator.

3.3 Matching

The primary objective of the 1978 Census of Agriculture Area Sample was to estimate the number and characteristics of farms not on the mail list. Operators not matched to the mail list were considered missed from the mail list. Matching was organized into three parts: (a) mail list matching, (b) self-matching, and (c) final matching. All area sample respondents were ultimately declared either matched or not matched to the mail list.

4. RELIABILITY

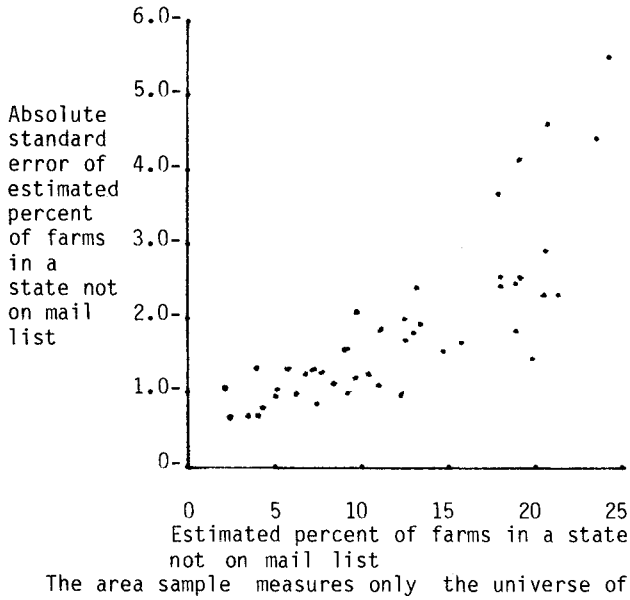
The reliability of the 1978 Census of Agricul-

ture is affected by both sampling and nonsampling error. Statistics in the census of agriculture are estimates based on a census of operators on a mail list and a sample of operators identified in a household area sample survey. Estimates are subject to two types of errors: sampling and nonsampling. Sampling error occurs because observations are made on only a sample and not all farm operations. Nonsampling error occurs for many reasons: failure to include operators on the mail list or to include households into the area sample survey, inability or refusal of farm operators to return questionnairs, inability or refusal of farm operators to provide correct information and other errors of data collection, response, processing, coverage, and estimation. A detailed discussion of errors present in the census and their magnitude is given for each state in the publication: 1978 Census of Agriculture, Volume 1, State and County Data.

An important statistic in measuring the effect of the area sample on the census is the percent of farms in a state not on the mail list or in county estimates. This statistic for a state and major subgroups within a state is used extensively in this paper. The sampling error associated with the estimate of this percent is determined jointly by sampling error associated with mail list and the area sample estimates. The sample error associated with percent of farms not on the mail list for the 48 states is given in Figure 1. The sampling error increases as the size of the percent increases.

The reliability of area sample estimates are influenced by both complete and item nonresponse. Accurate information is not available on the magnitude of either complete or item nonresponse in area sample estimates. Failure to enumerate farm operators who either refuse or are not-at-home could affect the estimates of percent of farms not on the mail list.

Figure 1. Relationship between estimate and the absolute standard error of the estimate for percent of farms in state not on mail list.
(• each dot represents a state)



The area sample measures only the universe of farms located in rural areas not on the mail list. The portion of the universe in an urban area was not included in the census. The area sample also has coverage error due to census processing, classification error, and enumerator error. A post enumeration survey of the 1978 Census of Agriculture Area Sample estimated the number of farms missed during area sample enumeration. This is discussed in detail in 1978 Census of Agriculture Coverage Evaluation (U.S. Bureau of the Census, 1982). The missed rates for the United States and the four regions are given below.

Region	Area sample estimate	Percent missed
U.S.	220,867	10.1
Northeast	17,611	13.8
North Central	52,478	18.9
South	118,219	6.4
West	32,559	7.4

The missed rate varies substantially by region. This nonsampling error affects in an unknown way the estimate of farms not on the mail list at the state level.

5. SIMULTANEOUS COVERAGE IMPROVEMENT AND COVERAGE EVALUATION

5.1 Comparison of Coverage Improvement and Coverage Evaluation

A unique feature of the 1978 Census of Agriculture Area Sample is that it is both a coverage improvement and a coverage evaluation survey. It is a coverage improvement survey for census estimates at the state, division, region, and nation level. It is a coverage evaluation survey for census estimates at the county level. The area sample is not a complete measure of the universe not included in county estimates partly due to design and partly to nonsampling error.

5.2 A Measure of Coverage

A measure of coverage error in the 1978 Census of Agriculture is average percent net underenumeration. This measure is defined below.

$$\text{Average percent net underenumeration} = \frac{\text{Total coverage error}}{\text{Adjusted total}}$$

$$P_\cdot = \sum_i (A_i/A_\cdot)(E_i/A_i)$$

$$= \sum_i W_i P_i$$

where $P_\cdot$ = The average error rate for subgroup levels.

$A_\cdot$ = Adjusted total over all subgroups.

$$A_\cdot = \sum_i A_i$$

E_i = The total coverage error associated with the i^{th} subgroup.

A_i = The adjusted total for the i^{th} subgroup.

W_i = The proportion of the adjusted total in subgroup i.

$$W_i = A_i/A_\cdot$$

P_i = Error rate or adjustment rate for the i^{th} subgroup.

$$P_i = (E_i/A_i)$$

For the 1978 Census of Agriculture Area Sample, the universe for which estimates are made are the nation, region, and state. The geographic level over which error rates are averaged is the county.

The percent of farms not on the mail list in the United States is a weighted estimate of the percent state net underenumeration without the area sample or average percent county net underenumeration. The average reflects the general level of coverage error of states or counties. A specific county or state may differ markedly from the average. While the average over states is 8.9 percent, the average for a particular state varies considerably. The distribution of state estimates is summarized below.

Percentile	Percent farms in state not included on mail list
Minimum	2.0
25.0	6.2
50.0	10.4
75.0	17.0
Maximum	23.8
Weighted average	8.9

The percent of farms not on the mail list in a state varies considerably from the weighted United States average. At the state level this is an indication of the variability of net underenumeration possible at the county level. The net underenumeration of any particular county may be substantially different from the state average county coverage error.

5.3 1978 Census of Agriculture Area Sample: A Coverage Improvement Survey

Only farms on the mail list were used to make county estimates of the number of farms, agricultural activity and agriculture productivity. In 1978, unlike 1969 and 1974, an area sample improved the coverage of farms at the state, division, region, and nation level.

The impact of the area sample in improving state level and higher estimates can be estimated using the results of the 1978 Census of Agriculture Coverage Evaluation Study (U.S. Bureau of the Census, 1978). The adjusted 1978 Census of Agriculture farm counts are compared to the published 1978 national farm counts and what these counts would have been without the area sample. This comparison is made by major subgroups and regions. The percent difference between the sum of county estimates and the adjusted U.S. total is a weighted county average percent coverage error.

Since the area sample was included in state estimates and excluded from county estimates, a comparison of estimates with and without the area sample is a comparison of the total for state estimates and the total for county estimates. The reduction in the average error from county estimates to state estimates is the reduction due to addition of the area sample. The average net underenumeration for states (with the area sample), and counties (without the area sample), are given for major sales and size groupings and for regions in Table B.

Table B. Comparison Average Percent of Net Underenumeration for State and County Estimates

Characteristics	State estimates (mail list plus area sample)	County estimates (mail list only)	Reduction in percent net underenumeration due to area sample	
			absolute	relative
All farms	3.4	12.0	8.6	71.7
Farms by sales (dollars)				
$40,000 or more	.5	1.2	.7	58.3
$10,000 to $39,999	3.4	6.1	2.7	44.3
$ 2,500 to $9,999	3.2	10.2	7.0	68.6
Less than $2,500	6.2	29.5	23.3	79.0
Farms by size (acres)				
500 or more	.7	1.5	.8	53.3
100 to 499	1.8	5.5	3.7	67.3
1 to 99	5.8	21.5	15.7	73.0
Farms by Region				
Northeast	6.2	17.3	11.1	64.2
North Central	3.3	8.2	4.9	59.8
South	2.6	14.0	11.4	81.4
West	5.2	16.2	11.0	67.9

5.3.1 Farm Count--The area sample improved state estimates reducing the average county coverage error from 12.0 percent to an average state coverage error of 3.4 percent. It reduced census coverage error by approximately 71 percent. The reduction in coverage error using the area sample with state estimates varied by region. The minimum it improved regional estimates was 59 percent in the North Central region; the maximum was 81 percent in the South.

5.3.2 Sales--The area sample improved the average coverage error in state and county estimates differently by size of sales group. For both state and county estimates, the size of the coverage error increased as the economic size of the farm decreased. The reduction due to using the area sample increased as the size of farm decreases. The area sample reduced the coverage error of state estimates of farms with sales less than $2,500 by more than 79 percent.

5.3.3 Land in Farms--The area sample had a similar major impact on coverage error by size of farm. The area sample had little effect on reducing the coverage error associated with large farms with 500 or more acres of land. As the size of farm decreased below 500 acres, the area sample reduced the coverage error in proportional larger amounts. The area sample reduced the state level coverage error among farms of less than 100 acres by almost 73 percent.

5.4 1978 Census of Agriculture Area Sample: A County Coverage Evaluation Survey

Estimates of the number of farms not on the

mail list--not published in county estimates--were made separately for each state. These estimates measure, in part, coverage error associated with using a mail list to make county estimates. The area sample is unique as a coverage evaluation survey of county estimates. It contains more detailed information than the typical coverage evaluation survey; it measures minor items and small subgroups. Since the area sample was included as part of all the standard state tables, the same information included in the mail list portion of the census for counties was collected for the area sample portion of the census.

The quality of county data is presented in Tables C and D for two separate items: market value of agriculture products sold and land in farms.

Each table presents the percent of farms not on the mail list and the distribution of the average county coverage error for states. Three statistics are used to describe the distribution of average county coverage error in states: (a) the 25th percentile, (b) the 50th percentile, the median, and (c) the 75th percentile. The 25th percentile gives a value for which 25 percent of the states are smaller and 75 percent of the states are larger. The 50th percentile gives a value for which one half of the states are smaller and one half of the states are larger. The 75th percentile gives a value for which 75 percent of the states are smaller and 25 percent of the states are larger. Since the area sample was not conducted in Alaska and Hawaii, percentiles are based on 48 states.

For example, Table D presents the characteristic land in farms. An estimated 1.5 percent of the land in farms was not on the mail list and not included in county estimates. For the 48 states, 50 percent of the states had an average county coverage error of less than 2.1 percent; 25 percent of the states had an average county coverage error of less than .9 percent; and 75 percent of the states had an average county coverage error of less than 4.0 percent.

The average county coverage error based on the area sample gives a minimum or lower bound for the true average county coverage error in a state. The average county coverage error is based solely on farms enumerated in the area sample. The area sample accounts for a substantial portion of the coverage error in the census of agriculture; it does not account for all the error. Specific reasons for the failure of the area sample to measure all the errors that are (a) the area sample does not cover urban areas, (b) the area sample contains nonsampling error in its estimates, and (c) the area sample does not measure errors associated with processing farms. A detailed discussion of errors is given in the 1978 Census of Agriculture Coverage Evaluation (U.S. Bureau of the Census, 1982).

All estimates of county coverage error are made using information for farms not on the mail list published in Volume 1 of the 1978 Census of Agriculture. A lower bound on the average county coverage error in a state can be made for most census items using Volume 1 tables for a particular state.

5.4.1 Sales--At least 8.9 percent of all farms with sales are not included in county totals. Fifty percent of all states have an average county coverage error of 11.0 percent or less. At least one percent of market value of agriculture sales is not included in county estimates. The counties in 50 percent of the states had an average net underenumeration of less than 1.3 percent. The quality of farm estimates varied by value of sales. The average county error in a state increased as a farm's sales became smaller. Large farms are well covered; small farms have sizeable coverage error at the county level. At least 50 percent of the states had average county error rates of 24 percent or less for farms with sales less than $2,500.

Table C. Total, Percent of Total Not in Counties and Distribution of State Average Coverage Error for Farms, Value, and Size Groups for Market Value of Agriculture Products Sold.

Market value of agriculture products sold	Percent not in counties (not on mail list)	Distribution state average county coverage errors		
		Percentile		
		25	50	75
Number of farms	8.9	6.3	11.0	17.4
Sales ($1,000's)	1.0	.6	1.3	1.8
Farms with				
$100,000 or more	.5	0	0	.8
$40,000 to $99,999	.9	.4	.9	2.3
$20,000 to $39,999	2.3	1.4	2.5	4.0
$10,000 to $19,999	3.4	2.2	3.7	4.9
$ 5,000 to $ 9,999	5.1	3.7	5.8	9.7
$ 2,500 to $ 4,999	3.4	7.2	9.7	13.1
Less than $2,500	24.7	20.7	24.6	35.2
Abnormal	0	0	0	0

5.4.2 Land in Farms--The land in farms is a characteristic of major interest in the census.

Table D. Total, Percent of Total Not in Counties, and Distribution of State Average County Coverage Error for Farms, Value, and Size Groups for Land in Farms.

Land in farms	Percent not in counties (not on mail list)	Distribution state average county coverage errors		
		Percentile		
		25	50	75
Number of farms	8.9	6.3	11.0	17.4
Land (acres)	1.5	0.9	2.1	4.0
Farms with				
1 to 9 acres	26.7	22.2	31.9	37.6
10 to 49 acres	17.6	13.2	20.6	25.2
50 to 69 acres	9.3	5.6	9.0	12.6
70 to 99 acres	7.2	5.1	8.3	13.0
100 to 139 acres	6.9	4.6	8.4	12.4
140 to 179 acres	4.4	2.2	5.3	8.1
180 to 219 acres	4.0	1.8	3.7	7.6
220 to 259 acres	2.6	1.2	3.0	5.1
260 to 499 acres	1.9	.8	2.2	4.2
500 to 999 acres	.9	0	.7	1.7
1,000 to 1,999 acres	.7	0	0	1.0
2,000 or more acres	.5	0	0	0

6. EFFECT OF THE AREA SAMPLE ON STATE ESTIMATES

The 1978 Census of Agriculture Area Sample has

had a major impact on census estimates. The effect of farms not on the mail list for major characteristics is given in Tables E, F, G, H, and I. There is a clear relationship between size of a characteristic and percent of farms not on the mail list. The effect of the area sample is to correct small farms more than for large farms for (a) sales, (b) land in farms, and (c) cropland harvested. The area sample affected specific categories of farms differently.

6.1 Sales--The percent of farms not on the mail list is related to size of the market value of agriculture products sold. As sales decrease, the proportion of farms not on the mail list increases rapidly. The area sample appears to significantly increase the quality of coverage of small farms. Operations which qualified under the "old" 1959 farm definition had a higher proportion of farms not on the mail list than did census farms.

Table E. Number and Percent of Farms Not on Mail List by Value of Products Sold.

Value of agriculture products sold	Farms	Percent of farms not on mail list
All farms	2,478,642	8.9
Farms by sales		
$100,000 and over	222,682	.5
$40,000 to $99,999	363,383	.9
$20,000 to $39,999	306,112	2.3
$10,000 to $19,999	309,594	3.4
$ 5,000 to $ 9,999	331,042	5.1
$ 2,500 to $ 4,999	331,874	9.4
$ 2,000 to $ 2,499	102,116	13.8
$ 1,500 to $ 1,999	119,657	17.5
$ 1,000 to $ 1,495	150,249	22.9
$ 0 to $ 999	239,631	34.1
Operations excluded by current definition but not by 1959 definition	470,013	53.6
Total 1959 definition farms	2,948,865	16.0

6.2 Land Use--The effect of the area sample in improving land use statistics is different for each land use category. The area sample improves the area in cropland used for pasture, total woodland, woodland pastured, and woodland not pastured more than for other land use categories.

Table F. Number of Acres on Farms Not on Mail List by Land Use Type.

Land use	Acres	Percent of acres not on mail list
Land in farms	1,029,694,535	1.5
Total cropland	431,340,542	1.6
Harvested cropland	320,666,222	1.1
Cropland used only for pasture	76,160,352	3.9
Other cropland	64,513,968	1.5
Total Woodland	94,891,726	3.2
Woodland pastured	48,338,533	2.7
Woodland not pastured	46,533,193	3.8
Other land	473,462,267	.9
Pasture land, all types	561,228,118	1.4
Irrigated land	50,837,940	1.0

6.3 Commodities--The number of farms and sales for a commodity were affected differently by the area sample. The area sample contributed 10 percent or more of the farms with vegetables, nursery products, poultry products, hogs and pigs, and other livestock. For commodity sales, the area sample contributes more than 2 percent of the sales of the following commodities: field seed, hay, tobacco, other livestock.

Table G. Number of Farms, Sales, and Percent of Total Not on Mail List by Commodity Type.

Commodity	Total value ($1,000)	Percent of total value not on mail list
Cotton and cottonseed	3,109,057	.3
Other crops	26,930,773	.6
Grains	3,047,142	.4
Fruit, nuts, and berries	4,630,238	.6
Vegetables, sweet corn, and melons	3,262,864	.7
Sheep, lambs, and wool	649,585	.8
Dairy products	11,317,386	.8
All crops	48,617,400	.9
Cattle and calves	298,877,884	.9
All livestock	59,496,118	1.1
Nursery and greenhouse products	2,867,497	1.1
Hogs and pigs	8,165,040	1.1
Poultry and poultry products	8,583,253	1.4
Field seed, hay, forage, and silage	2,322,450	2.0
Tobacco	2,397,380	3.3
Other livestock	892,969	4.7

6.4 Operator Characteristics--The area sample had a major impact on improving the coverage of special groups of operators. The area sample improves the coverage of operators with: (1) a principal occupation other than farming, (2) a residence on the farm operated, (3) female operators, (4) spanish origin, and (5) black operators. The proportion of farms not on the mail list increased with the number of days the operators worked off the farm. The proportion also increased as the operator's age became younger.

Table H. Number of Farms and Percent of Total Not on Mail List by Livestock Inventory Item.

Livestock and poultry inventory	Farms	Percent of farms not on mail list
Sheep and lambs	98,150	6.4
Milk cows	333,620	6.5
Beef cows	1,032,952	7.6
Cattle and calves	1,461,944	7.9
Hogs and pigs	512,292	13.1
Horses and ponies	467,789	14.6
Any poultry	368,181	22.9
Chickens 3 months old or older	315,287	23.6
Turkeys	26,638	28.9
Broilers and other meat-type chickens	81,316	32.9
Goats	28,276	35.7
Other poultry	79,307	38.2

Table I. Percent of Farms Not on Mail List in the United States by Operator Characteristics.

Classification	Farms	Percent of farms not on mail list
Principal occupation		
Farming	1,326,785	4.3
Other	1,151,857	14.2
Place of residence		
On farm operated	1,785,023	11.2
Not on farm opeated	441,594	4.5
Not reported	252,025	.7
Operators age group		
1 to 24	78,975	15.7
25 to 34	323,175	11.7
35 to 44	485,109	10.6
45 to 54	595,532	7.8
55 to 64	588,212	6.1
65 and over	407,639	9.1
Tenure		
Full	1,451,446	10.6
Part Owner	713,548	12.0
Tenant	313,648	11.1
Days work off farm		
None	1,001,416	5.9
1 to 49	188,134	3.5
50 to 99	76,001	6.6
100 to 149	78,447	7.1
150 to 195	117,307	8.0
200 or more	904,946	14.9
Organizations		
Individual	2,175,437	9.6
Partnership	241,209	3.6
Corporation	51,270	2.0
Other	10,645	14.8
Operators by sex		
Male	2,350,472	8.7
Female	128,170	12.0
Operators of Spanish Origin	22,997	19.2
Black Operators	79,916	27.3

6.5 Livestock and Poultry Inventory--Specific livestock and poultry items had a large proportion of farms accounted for by the area sample. Farms with chickens, turkeys, broilers, and goats had over 20 percent of the farms contributed by the area sample. For livestock and poultry inventory, the area sample significantly improved the quality of coverage among farms with horses and ponies.

7. CONCLUSION

The 1978 Census of Agriculture Area Sample was a significant methodological improvement to the census of agriculture. It improved the quality of coverage of farms and their characteristics at the state level and higher. The average coverage error among state level estimates was reduced to 3.4 percent for farms compared to an average coverage error among county level estimates of 12.0 percent. The reduction in coverage error varied by characteristic and size category. The area sample had the greatest impact on small size groups. It affected farm counts more than farm characteristics such as sales, acres, and livestock inventory. The area sample provides a detailed evaluation of county data. The area sample provides a lower bound on county coverage error for data items published.

8. REFERENCES

Chapman, David D. and Charles E. Rogers (1978). 1978 Census of Agriculture Area Sample. Proceedings of Survey Research Section, American Statistical Association.

Chapman, David D. and Charles E. Rogers (1980). Sampling and Estimation in the 1978 Census of Agriculture. Proceedings of the Survey Research Section, American Statistical Association.

U.S. Bureau of the Census (1974). 1969 Census of Agriculture. Volume V. Special Reports. Part 16. Evaluation of Coverage. U.S. Government Printing Office, Washington, D.C.

U.S. Bureau of the Census (1978). 1974 Census of Agriculture. Volume IV. Special Reports Part 3. Coverage Evaluation. U.S. Government Printing Office, Washington, D.C.

U.S. Bureau of the Census (1981). 1978 Census of Agriculture. Volume 1. State and Summary Data. Part 51. United States. U.S. Government Printing Office, Washington, D.C.

U.S. Bureau of the Census (1982). 1978 Census of Agriculture. Volume V. Special Reports, Part 3. Coverage Evaluation.