IMPACT OF NONSAMPLING ERRORS ON WEIGHTED TRACT SURVEY INDICATIONS

by

George W. Hill

Martha S. Farrar

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

December 1977

CONTENTS

Page

ACKNOWLEDGEMENTS iii

INTRODUCTION 1

PURPOSE OF STUDY 1

SUMMARY 2

RECOMMENDATIONS 3

THE SAMPLE 3

THE SURVEY 4

EXPANSION OF DES AND RECONCILED DATA 4

DIRECT EXPANSIONS 5

WEIGHTED DATA 6

TEST OF HYPOTHESIS 8

- Reported Land, Cattle and Hogs 8
- Weights 9
- Weighted Cattle and Hogs 9

REASONS FOR DIFFERENCES BETWEEN DES AND RECONCILED DATA 10

- Entire Farm Acres 10
- Entire Farm Cattle 12
- Entire Farm Hogs 13
- Tract Hogs 14

APPENDIX A - Supporting Analysis Tables 16

APPENDIX B - Test of Hypothesis 22

LIST OF TABLES

Table — Page

Narrative

1. Reconciled Data Direct Expansion as a Percent of DES Direct Expansion 5
2. Expanded Number of Cattle by Weighting 7
3. Expanded Number of Hogs by Weighting 8
4. Number of Differences in Entire Farm Acres by Reason 10
5. Number of Differences in Entire Farm Acres by Interval 12
6. Number of Differences in Entire Farm Cattle by Reason 12
7. Number of Differences in Entire Farm Cattle by Interval 13
8. Number of Differences in Entire Farm Hogs by Reason 13
9. Number of Differences in Entire Farm Hogs by Interval 14
10. Number of Differences in Tract Hogs by Reason 14
11. Number of Differences in Tract Hogs by Interval 15

Appendix A

1. Number of Tracts Sampled-1976 JES, DES and December Reinterview .. 17
2. Number of Tracts Sampled Including Sub-tracts and Split Tracts-December Reinterview 18
3. Effect of Reconciled Farm Acres on the Weights 19
4. Effect of Reconciled Farm Acres and Cattle on the Weighted Expansion 20
5. Effect of Reconciled Farm Acres and Hogs on the Weighted Expansion 21

ACKNOWLEDEMENTS

Reinterviewing operators as a quality check for the same data they supplied earlier is a "tough job". For this reason the enumerators in Indiana, North Carolina, and Oklahoma deserve a special thanks for their determination and hard work in this endeavor. Thanks goes to each SSO for their responsiveness and support by taking on the extra task during a heavy work load period.

INTRODUCTION

Area frame sampling is the back bone of the data collection efforts of the Statistical Reporting Service (SRS). The major advantage of sampling from an area frame is that the frame is always complete. The sampling unit from the area frame is a segment of land. Within the segment boundaries each unit of land under one operation is a reporting unit. This reporting unit is commonly known as a tract.

Three separate livestock estimates can be obtained from the area frame sample segments. The tract estimate results from recording and expanding the number of livestock physically located inside the segment boundaries. The farm estimate is derived by expanding the number of livestock on all land operated by farmers who reside inside the segment boundaries. The third and lesser known estimate which may be computed from the area frame segments is called the weighted tract estimate. This estimate is computed by prorating all the livestock on each farm into the tract based on the proportion of land inside the segment boundaries relative to the total farm acres. The number of livestock associated with each tract in the sample segment is equal to (tract acres/farm acres)*(farm livestock).

PURPOSE OF STUDY

The validity of the data collected in a survey has an important bearing on the quality of the estimates resulting from the survey. Errors in our estimates caused by inaccurately reported data are known as non-sampling errors. These errors, unlike sampling errors, are not generally measurable and cannot be completely controlled. Therefore, a concerted effort must be made from time to time to identify these errors, determine their cause and minimize their effect on the estimates in future surveys. The purpose of this study is to isolate the nonsampling errors in the December Enumerative Survey (DES) as they affect the weighted tract expansion. The DES was used as the basis for this study because the weighted tract estimate was computed on 16 states for the 1976 survey. The weighted estimates provided smaller sampling errors than the tract estimates, but the magnitude of non-sampling errors could cause the total error of the weighted estimate to be larger than that of the tract estimate. Nonsampling errors in the weighted indications are a combination of errors in tract and farm acres as well as farm livestock. Previous research has shown farm operators have particular problems in reporting farm acres.[1/] The primary objective of this study is to identify nonsampling errors in entire farm acres and livestock and to report their effect upon the weights and weighted livestock expansions.

[1/] a) Bosecker, Raymond R. and Kelly, William F. "Nebraska Survey Concept Study" SRS, USDA, November 1975.

b) Hendricks, and Searls and Horvitz "A Comparison of Three Rules for Associating Farms and Farmland With Sample Area Segments in Agricultural Surveys"

c) Report on 1962-63 Research Projects, R & D Branch, Jan. 1968

SUMMARY

This study was undertaken in three States to test for response biases in the weighted tract estimator. This estimator is derived through the formula (tract acres/farm acres)*(farm livestock). In all three States there was a significant downward bias in number of farm acres reported in the 1976 DES. This understatement of farm acres caused the weights (tract acres/farm acres) to be significantly too large. Therefore, even if the number of livestock were reported perfectly, the weighted livestock indications were subject to an upward bias.

It could be argued that further testing for significance in the weighted livestock indications was not necessary. Unless the weights can be determined accurately, the weighted estimate has a built in source of bias. However, testing was continued to examine the effects of response errors in entire farm livestock as well as acres operated. Farm cattle and farm hogs were found not to be significantly different from the DES numbers. Errors in reported livestock were largely offsetting. However, the weighted indications obtained after correcting both the farm acres and the farm livestock were significantly below the DES level for cattle in Indiana and for hogs in North Carolina. Comparisons of weighted indications after correcting for errors in farm acres, farm livestock or both are provided in Tables 2 and 3 for cattle and hogs respectively.

Any change in the DES data was accompanied by an explanation from the farm operator of why the new data was more correct. Reasons for differences in reported acreage are given in Table 4 and the reasons associated with changes in cattle or hogs are provided in Tables 6 and 8 respectively.

The most prevalent reasons given for undercounting farm acres centered around failure to include separate parcels of land and portions of land not actively used such as woodland, idleland and wasteland. This is not a new revelation since these problems have been reported in several earlier studies dating at least as far back as 1962-63 SRS research projects and a study for the Census Bureau connected with the 1959 Census of Agriculture. What this study demonstrates is that unfortunately these problems still exist even after sketching out parcels of land and intensive training of enumerators.

Based on the results of this study, continued use of the weighted indication should be accompanied by redoubled efforts at measuring entire farm acreage accurately. The problems of determining the correct number of livestock on the tract or farm continue regardless of the estimator used. It is hoped that the sources of errors detected by the reinterview and reconciliation will provide insight to areas where additional emphasis could reduce nonsampling errors.

RECOMMENDATIONS

1. Entire farm acres must be more accurately obtained from the operator if the current method of weighting is to be continued. This can be accomplished in at least two ways. First, more effort should be given to the content of Section E - Acres Operated. Alternative procedures should be prepared to deal with the sources of error presented in this study. Special emphasis must be given to woodland, land not in use and rented land. These problems and others which must be addressed are shown in Table 4.

Second, enumerators should be cautioned not to ask for "best estimates" too quickly. Many of the detected differences were the result of rough estimates in the DES being replaced by more careful counting in the reinterview.

2. Alternative weights other than using entire farm acres should be investigated. Would cropland or land "actively used" during the past year provide more accurate weighted estimates? Statistics Canada uses (tract acres minus woodland acres/entire farm acres minus woodland acres) for their weighted indications. Other alternatives are also available.

THE SAMPLE

The reinterview was done on a subsample of DES tracts in three states: Indiana, North Carolina and Oklahoma. A convenient random subsample was provided from segments newly rotated into the 1976 JES sample and segments to be rotated out before the next June survey. The number of tracts sampled varied by state from 28 to 39 percent of those used for the DES.

After the June Enumerative Survey (JES) each area tract was classified into an "assigned strata". This classification was based on the enumerated data obtained during the JES. The December Enumerative Survey (DES) sample was a subsample of the June Survey tracts taken from the "select strata". Select strata are the same as the assigned strata except for tracts having some special characteristics such as very large operations or nonoverlap tracts. Table 1 in Appendix A identifies the select strata subsampled for this study and gives the number of tracts selected for the follow-up. The reinterview sample was a subsample of the original DES tracts selected. This was necessary to maintain as short a time period as possible between the DES collection date and the reinterview date. The three state total sample size of 631 tracts for the reinterview, therefore, does not include any sub-tracts from the DES. However, as shown in Table 2 of Appendix A, there were an additional 28 sub-tracts associated with the original sample which provided a total of 659 tracts to be reinterviewed. There were 528 useable reports, for 80 percent completion rate. The count does include tracts that should have been split during the DES but were not.

The sample coverage in terms of expanded cattle and hogs is as important as the sample counts. The useable reports for which comparable data is available from both the DES and reinterview accounted for 20 percent of the weighted hog indications and 30 percent of the weighted cattle indication (excluding extreme operators) in the three States.

THE SURVEY

Enumerator training was provided in a six hour period for each state. Enumerator Reinterview Instructions were provided to each person working the survey. Most of the field work was completed before the end of December.

The reinterview questionnaire contained many detailed questions. The main idea on the acreage and livestock sections was to ask for the data in small units or lots. These pieces would then be added together to arrive at the total. The acreage questions were primarily based on land use (cropland, pastureland, wasteland) by unit instead of the current procedure (owned, rented, or managed) by parcel although acres owned, rented from others and to others were also asked. The livestock questions asked for the total number by groupings or lots instead of by the present classification questions.

Once the reinterview data was obtained, the respondent was asked to reconcile the difference between the reinterview data and the information obtained on the original DES interview. The DES data had been entered on the back of the questionnaire prior to the field work. The operator was then asked for the most accurate data based on the reconciliation. The reconciliation resulted in one of the following three conditions: the original DES interview data were correct; the reinterview data were correct; neither of these were correct. This reconciled data is considered "truth" for analysis purposes. This final determination by the operator of reconciled data must be perceived in practical terms as the most correct data possible within the framework of our data collection procedures and survey concepts.

EXPANSION OF DES AND RESPECTIVE RECONCILED DATA

Certain records required special consideration for the joint expansion of the DES and reconciled data. For each record obtained in the sample both parts (DES and reconciled data) must be useable. A useable record cannot have any part observed or estimated. This created particular problems where tracts were split during the reinterview that had not previously split during the DES. This means the DES tract operator was not the correct operator for the entire tract sampled. Data for the entire tract originally selected must be useable. If any piece of the original tract is not accounted for and useable, the entire tract as selected must be excluded for the comparable data expansions.

There were 659 tracts to be reinterviewed for the three states combined. This included 28 sub-tracts that were determined during the DES. There were 15 tracts that should have been split during the DES but were not. About one half of these were useable. These useable split tracts were combined to form one record which contained reconciled data comparable to the DES data for the tracts selected. These combined records were then used for the comparable data expansions.

DIRECT EXPANSIONS

The DES and reconciled tract and farm data were expanded by the JES expansion factor and December interval. The reconciled expansions were calculated as a percent of the DES expansions. These percentages for farm acres, cattle and hogs for each of the three states in the study are shown in Table 1 below. Note, the farm expansions include both resident and non-resident farm operator data. The expanded farm acres include extreme operator acres unlike the livestock expansions which have their respective extreme operator data removed. This is consistent with the way the DES extreme operator data is summarized.

Table 1: Reconciled Data Direct Expansion as a Percent of DES Direct Expansion

Item	Indiana %	North Carolina %	Oklahoma %	Total %
Farm Acres	103	111	105	106
Farm Cattle	102	107	97	99
Farm Hogs	96	99	103	98
Tract Hogs	103	122	102	109

The reconciled entire farm acreage ranged from three to eleven percent above the comparable DES data over the three states. The three states combined show a six percent increase over the DES expansion.

The reconciled entire farm cattle were above the number of head obtained on the DES in two out of the three states. The third state expanded to three percent fewer cattle than reported in the DES. Contributing to the DES over-count were two reports with major differences in cattle numbers. These two reports expanded to a difference of 224,000 head (DES above reconciled data). If these were excluded the reconciled cattle would be two percent above the DES not three percent below. One operator maintained there had been no personal contact for the DES and the data was obtained over the phone with a bad connection. The other operator interpreted the DES questions as wanting the number of head he had raised during the year rather than the number he had on hand. In North Carolina the expanded cattle on one report showed 39,000 head more for the reconciled data and another report showed 26,000 head more. If these reports were deleted the expanded number of reconciled cattle would only be one percent above the DES data instead of seven percent above.

Nonsampling error research is directed at finding the "rare occurence". The items being identified and measured are "few and far between". However, these mistakes do alter the outcome of survey indications. Without the rare large mistakes the number of cattle after reconciliation were rather consistently one or two percent above the DES but with the few large errors the results were considerably different among the three States.

Two out of the three states showed expanded farm hogs after reconciliation below the respective DES data. Indiana, with the largest decrease, had a large problem report. A partnership involved the hogs but not the land on which they were located. If the 98,000 head (expanded) were removed from the DES expansion, the reconciled data would be one percent above the DES. No noticably large differences were identified in the other two states.

It is noteworthy that while the reconciled farm hogs were generally lower than the DES hogs, the reconciled tract hogs were above the DES in all three states. Two of the three states show two and three percent increases over the DES data. North Carolina had reconciled hogs at 22 percent above the reported DES tract hogs. This was due to one report of 453 head which expanded to 61,000 head. It was verified that the hogs should have been recorded on the tract during the DES. The column for hogs on tract acres was left blank on the DES questionnaire through an oversight. If those hogs had been recorded for the DES, the reconciled hogs for the state would have been three percent below the DES data, not 22 percent above.

WEIGHTED DATA

The weighted livestock for each record is calculated from three variables in the following formula:

wtd. lvsk. = (tract acres/farm acres) * (farm livestock).

For example, the formula shows that if farm acres are too low (understated by respondent) the resulting weighted livestock will be too high. Essentially the change in weighted livestock moves in the opposite direction as the change in farm acres and the same direction as the change in tract acres and farm livestock.

It should be noted that an error in farm acres might be offset by an error in the tract acres and/or farm livestock. Thus, two "wrongs" within a questionnaire could be at least partially offsetting. One should not count on this happening since two errors could also compound the total error. Unless the weights (tract acres/farm acres) can be accurately determined, the weighted indication is open to serious bias.

DES data for all three variables in the weighted formula were used to obtain a base expansion. This base was then compared to other weighted expansions. Thus, by using DES data for some variables in the weighted formula and reconciled data for others, the effect of individual and combined changes can be measured. The discussion of cattle and hog expansions are presented separately. Table 2 shows four weighted expansions for cattle in each of the three states in the study. These four were obtained by changing the data source of the three variables. As previously discussed, the base expansion (line 1) uses the DES as the data source for all three variables.

Table 2: Expanded Number of Cattle by Weighting

LINE	Weighted Variables DES Data = X Recon. Data = R			Indiana		North Carolina		Oklahoma		Total	
	acres tract	acres farm	farm cattle	(000)	% of base	(000)	% of base	(000)	% of base	(000)	% of base
1	X	X	X	347	100	333	100	1,171	100	1,851	100
2	X	R	R	332	96	332	100	1,133	97	1,797	97
3	X	X	R	358	103	357	107	1,165	99	1,880	102
4	X	R	X	325	94	316	95	1,139	97	1,780	96

The weighted expansion using DES tract acres but reconciled data for the entire farm acres and farm cattle (line 2) resulted in fewer head than the comparable expansion using all DES data. This is true for all three states in the study. The smaller number of head may be the result of: (1) overstatement of DES cattle or (2) understatement of DES entire farm acres or (3) a combination of the previous two. The first of these conditions may be ruled out of the possibilities in two of the three States.

When only reconciled farm cattle numbers were substituted into the formula (line 3), the new weighted indication was three and seven percent above the DES base for Indiana and North Carolina respectively. However, when cattle were maintained at the DES number and reconciled farm acres were substituted (line 4), the weighted indication declined 6 percent in Indiana and 5 percent in North Carolina. Obviously, a change in either variable has considerable impact on the indication. The combined or net results from using reconciled data for both acreage and cattle (line 2) was 4 percent fewer cattle than the DES in Indiana and only a slight downward adjustment in North Carolina. A combination of both fewer cattle and more acreage determined during reconciliation in Oklahoma resulted in a weighted indication three percent below the original DES weighted expansion.

The data in Table 3 on the next page summarize the weighted hog numbers in a manner consistent with those for cattle.

Table 3: Expanded Number of Hogs by Weighting

LINE	Weighted Variables DES Data = X Recon. Data = R			Indiana		North Carolina		Oklahoma		Total	
	acres tract	acres farm	farm hogs	(000)	% of base	(000)	% of base	(000)	% of base	(000)	% of base
1	X	X	X	420	100	328	100	130	100	878	100
2	X	R	R	376	90	283	86	133	102	792	90
3	X	X	R	393	94	315	96	134	103	842	96
4	X	R	X	405	96	303	92	129	99	837	95

The line 1 weighted expansion is the DES base which was calculated using DES data for all three variables. When reconciled farm acres and farm hogs were substitued (line 2), the weighted hog numbers declined dramatically in Indiana and North Carolina. These declines were the result of more farm acres than originally reported and fewer hogs than the DES data. The two percent increase (reconciled above DES) for Oklahoma came from an increase in hog numbers which more than offset an increase in farm acres.

TEST OF HYPOTHESIS

The previous data showed differences between the DES and reconciled data. The Wilcoxon Signed Rank Test [1/] was then used to test whether there is evidence of significant upward or downward bias. In other words, were those differences skewed in one direction or were the errors largely offsetting. The details of this test are given in Appendix B. Tests were performed on the following differences: reported land, cattle and hogs, the computed weights, and the weighted cattle and hogs.

Reported Land, Cattle and Hogs

The hypothesis of no difference in reported farm acres between DES reported data and reconciled data was rejected in all three states with an α level $< .01$. That is, there were significantly more and/or larger understatements of farm acreage in the DES than overstatements. This was not the case with livestock as neither reported cattle nor reported hogs were significantly "one sided" in the errors made.

[1/] Hollander and Wolfe, Nonparameteric Statsitical Methods, Wiley and Sons, N.Y., 1973 pgs. 26-32.

Weights

The weights for expanding the cattle and hog numbers in each record were (tract acres/farm acres). Weights were computed for each observation using first the DES farm acres and then the reconciled farm acres. Table 3 of Appendix A shows the number of differences between weights categorized by three size criteria (abosolute percent change from DES weight): greater than ten, twenty five, and fifty percent. The data in Table 3 shows that of the 200 weights which changed, 60 differed by more than 25 percent from the original weight (three state data combined). Also, there were 127 of the total 200 weights which decreased in size from the DES to the final reconciliation. Finally, the table shows the lowest alpha level for which the null hpyothesis could be rejected. The results for each of the three states rejects the hypothesis that the median of the errors in weights is zero. Therefore, the difference between DES and reconciled farm acres significantly changed the weights. The median of the DES weights was significantly biased upward.

This result is in direct contrast to the "Nebraska Survey Concept Study" where it was found that mistakes on the weights were offsetting.[1/] It is evidenced from the direct expansions as well as from the indicator variable used in calculating the test statistic that the DES farm acres were smaller than the reconciled farm acres. Therefore, the resulting overstatement in DES weights had the potential for expanding the DES weighted livestock too high; i.e. an upward bias. The high expansion will occur unless: (1) the errors in weights for livestock operators are fewer and smaller than for non-livestock operators and (2) errors in the livestock numbers offset the errors in the weights. Neither of these situations are very comforting to rely on when a consistent estimator is desired. Accurate weights are essential to the weighted estimator.

Weighted Cattle and Hogs

The differences in weighted cattle and hog numbers were tested for significance even though of lesser importance since the weights were found uniformly too large in the DES. The hypotheses and calculations for the test statistics are the same as previously done for the weights. Test statistics are summarized for cattle in Table 4 and for hogs in Table 5 of Appendix A. The weighted DES cattle indication for Indiana was significantly above the reconciled indication. Weighted DES cattle numbers for North Carolina and Oklahoma, due to one or both of the previously named situtations, were not significantly above the reconciled data. The DES weighted hog indication was significantly higher than the reconciled weighted hogs in North Carolina. The weighted hogs using reconciled farm acres and hogs were 86 percent of the DES weighted hogs.

REASONS FOR DIFFERENCES BETWEEN DES AND RECONCILED DATA

Differences between DES and reconciled data was limited to four major areas: (1) entire farm acres, (2) farm cattle, (3) farm hogs, and (4) tract hogs. Reasons for differences were established for each of these areas.

Entire Farm Acres

Forty-four percent of the tract operators reinterviewed reported different entire farm acres. Approximately one out of two were different in North Carolina while one out of three were different in Indiana and Oklahoma. The following data in Table 4 gives the reasons for these differences.

Table 4: Number of Differences (DES Versus Reconciled) in Entire Farm Acres by Reason

Reason	Three States Combined
REPORTED FARM ACRES TOO LOW IN DES	
Acreage was estimated	26
Miscounted acreage, left some out	24
Entire parcel left out - idleland or woodland	19
Failed to report land rented from others	15
Failed to report land not in use	13
Attributed to a differenct respondent	11
Omitted entire farm acres	8
Split tract not picked up in DES	7
Don't know	6
Misunderstood questions	5
Entire parcel left out - pasture	3
Failed to report land in a separate location	2
Left out operated land owned by family members	2
Didn't remember first interview	2
Land was to be sold in the near future	2
REPORTED FARM ACRES TOO HIGH IN DES	
Acreage was estimated	18
Included land rented out	15
Included public land	10
Attributed to a different respondent	8
Split tract not picked up in DES	6
Miscounted acreage, included too much	5
Don't know	5
Included land operated by family members	5
Misunderstood questions	4
Included land in a diff. business arrangement	4
Included entire parcel of non-operated idleland or woodland	3
Didn't remember first interview	2
Miscellaneous	2
TOTAL	232

The largest number of differences (19 percent) in entire farm acres occurred because the operator estimated his acreage rather than taking the time to account for exactly the acres in the operation. This may not be easy to correct given the short time period in which the enumerators must perform and the length of the questionnaire. These differences should not be viewed as differences between the operator's "best estimates". The data were collected independently and then the operator was asked to explain what he thought caused the difference. Operators often commented that for the reinterview they took pencil, paper and time to correctly report the data whereas on the first interview they generally had not done this and were sometimes encouraged by the enumerator to give their "best estimate". Of the operators who gave this reason for the differences, 59 percent under estimated the acreage while 41 percent overestimated the acres operated.

The next largest number of differences (13 percent) involved rented land. Fifteen operators did not report land rented from others and 15 operators incorrectly reported land rented to others.

There were 29 operators (13 percent) who miscounted their acres by the amount that was identifiable or meaningful to them. Twenty-four of these operators left out the identifiable amount of acres while five miscounted by including too much.

Problems in reporting entire separate parcels of idleland and/or woodland occurred for 22 operators or 9 percent of the total differences. Nineteen of these differences were due to leaving out the entire parcel. Some idleland and/or woodland not physically separated from other reported land also caused differences. These differences number 13 (6 percent) and are categorized as "failed to report land not in use". This land was generally not separated by fences (or any physical barrier) and was not in a spearate location. When the two categories of reasons involving woodland and/or idleland are added together they account for 15 percent of the total differences. As evidenced, this is a particular problem for the more heavily wooded areas experienced in the eastern states.

Nineteen of the differnces (8 percent) could only be attributed to a different resondent. The operator contacted in the reinterview had not been the respondent earlier.

There were 13 differences (6 percent) resulting from the DES tract being split. This means it was determined during the reinterview that a part of the DES tract was operated by someone other than the DES tract operator. The portion of the tract under question generally involved woodland and whether the woodland was rented to or from someone.

There were eight differences due to omitting entire farm acres, these operators only accounted for the tract acres. The two miscellaneous reasons were due to misreading the photo boundaries and including land that was to be bought in the near future.

The magnitude of these individual differences shows there were a few large discripancies but the 61 percent were under 50 acres (Table 5). The differences between the DES and reconciled farm acres were summarized by size of difference as shown on the next page.

Table 5: Number of Differences (DES Versus Reconciled) in Entire Farm Acres by Interval

Interval	3-State Total
less than -500	4
-401 to -500	8
-301 to -400	6
-201 to -300	6
-101 to -200	19
-51 to -100	20
-1 to -50	84
1 to 50	58
51 to 100	11
101 to 200	10
201 to 300	3
greater than 300	3
TOTAL	232

Entire Farm Cattle

The number of differences and reasons for differences in farm cattle are prsented in Table 6 below.

Table 6: Number of Differences (DES Versus Reconciled) in Entire Farm Cattle by Reason

Reason	Three States Combined
REPORTED FARM CATTLE TOO LOW IN DES	
Miscounted, left some out	10
Failed to report non-owned lvsk.	7
Split tract not picked up in DES	4
Miscellaneous	4
Number of head were estimated	3
Questions were misunderstood	3
Attributed to a different respondent	3
REPORTED FARM CATTLE TOO HIGH IN DES	
Number of head were estimated	7
Respondent did not know	3
Miscounted, too many	3
Questions were misunderstood	3
Miscellaneous	3
Attributed to a different respondent	2
Duplication occurred in other classes	2
TOTAL	57

The largest number of differences at 13 (23 percent) were due to miscounting while 10 (18 percent) were due to estimating the number of head during the DES. The next largest category of differences (12 percent) were due to failure to report non-owned livestock on the acres operated. The miscellaneous reasons for farm cattle being reported too low for the DES involved one difference for each of the following: failed to report owned cattle on land rented from others; different interpretation of tract boundaries; respondent did not remember giving the information the first time; respondent did not know what caused the difference. The miscellaneous reasons for farm cattle being reported too high on the DES included the following: reported cattle on land not operated; respondent remembered no first interview; split tract not picked up in DES.

The size of these differences and frequency by interval are shown in Table 7 below. The three largest differences were 146, 374 and 433 head.

Table 7: Number of Differences (DES Versus Reconciled) in Entire Farm Cattle by Interval

Interval	3-State Total
less than -20	9
-11 to -20	6
-6 to -10	6
-1 to -5	11
1 to 5	14
6 to 10	3
11 to 20	1
greater than 20	7
TOTAL	57

Entire Farm Hogs

The reasons for differences in entire farm hogs are shown in Table 8 and the size of the differences for certain intervals are given in Table 9.

Table 8: Number of Differences (DES Versus Reconciled) in Entire Farm Hogs by Reason

Reason	Three States Combined
REPORTED FARM HOGS TOO LOW IN DES	
Miscounted, left some out	3
Number of head were estimated	2
Questions were misunderstood	2
Only attributable to diff. respondent	2
Failed to report "pet" pigs	2
Failed to report no. for home use	1
Failed to report non-owned hogs on op.	1
Split tract not picked up in DES	1

Table 8: (Con't)

Reason	Three States Combined
REPORTED FARM HOGS TOO HIGH IN DES	
Number of head were estimated	2
Reported hogs owned but were not on op.	2
Miscounted, counted too many	1
Only attributed to diff. respondent	1
Included fathers hogs not on op.	1
TOTAL	21

Table 9: Number of Differences (DES Versus Reconciled) in Entire Farm Hogs by Interval

Interval	3-State Total
less than -10	5
-6 to -10	3
-1 to - 5	6
1 to 5	3
6 to 10	-
greater than 10	4
TOTAL	21

Tract Hogs

As might be anticipated, there were fewer differences for tract hogs than for farm hogs (Table 10). Differences in tract hogs did not always result in a difference for the entire farm number of hogs.

Table 10: Number of Differences (DES Versus Reconciled) in Tract Hogs by Reason

Reason	Three States Combined
REPORTED TRACT HOGS TOO LOW IN DES	
Reason could not be determined	2
Miscounted number of head	2
Questions were misunderstood	1
Only attributed to diff. respondent	1
Failed to report hogs considered "pets"	1
Different interpretation of tract boundaries	1
Split tract not picked up in DES	1

Table 10: (Con't)

Reason	Three States Combined
REPORTED TRACT HOGS TOO HIGH IN DES	
Confused tract and farm boundaries	2
Miscounted number of head	1
Different interpretation of tract boudaries	1
Included fathers hogs not on op.	1
TOTAL	14

The size and distribution of these differences are "roughly" the same as for the entire farm hogs (Table 11). However, the end points (largest positive and largest negative) of the differences are considerably larger than for the farm hogs. This is not what would be expected. One of the reports that had a difference of 453 head was the result of completely missing the tract hogs and just recording farm hogs. The hogs were definitely on the tract for the DES but the tract hogs column on the questionnaire was completely blank.

Table 11: Number of Differences (DES Versus Reinterview) in Tract Hogs by Interval

Interval	3-State Total
less than -10	4
-6 to -10	1
-1 to -5	5
1 to 5	1
greater than 5	5
TOTAL	16

Identifying the source of these nonsampling errors is only the beginning toward reducing the total error in the survey. The challenge of solving current problems while avoiding the creation of new problems is not easy. Periodic monitoring of response errors should be maintained to continuously update knowledge of the amount, direction and nature of nonsampling errors in the surveys.

APPENDIX A

Supporting Analysis Tables

TABLE 1.--NUMBER OF TRACTS SAMPLED - 1976 JES, DES and December Reinterview

Select Stratum	Indiana			North Carolina			Oklahoma			Total		
	JES	DES	Reint.	JES	DES	Reint.	JES	DES	Reint.	JES	DES	Reint.
1) Wheat, rye and chickens	58	45	9	37	20	4	89	45	11	184	110	24
2) Wheat or rye	369	225	63	231	64	28	562	210	64	1162	499	155
3) Hogs and chickens	30	20	5	109	35	17	52	26	8	191	81	30
4) Chickens	29	16	5	134	50	23	131	65	25	294	131	53
5) Hogs	104	61	17	143	45	33	50	21	8	297	127	58
6) Cattle	126	70	22	337	85	35	495	206	78	958	361	135
8) NOL w/hogs or w/cattle	326	326	92	258	155	37	444	222	47	1028	703	176
TOTAL	1042	763	213	1249	454	177	1823	795	241	4114	2012	631

TABLE 2.--NUMBER OF TRACTS SAMPLED Including Sub-tracts and Split Tracts - December Reinterview[1/]

Select Stratum	Indiana		North Carolina		Oklahoma		Total	
	n	n'	n	n'	n	n'	n	n'
1) Wheat, rye and chickens	9	5	4	6	14	11	27	22
2) Wheat or rye	67	49	29	30	65	50	161	129
3) Hogs and chickens	5	3	19	18	9	8	33	29
4) Chickens	5	1	24	22	28	23	57	46
5) Hogs	17	11	34	33	8	7	59	51
6) Cattle	22	20	35	30	79	62	136	112
8) NOL w/hogs or w/cattle	95	60	37	33	54	46	186	139
TOTAL	220	149	182	172	257	207	659	528

[1/] n = number of tracts in sample including sub-tracts

n' = number of useable tracts including split tracts, therefore it is admissable that n' > n.

TABLE 3. -- Effect Of Reconciled FARM ACRES On The Weights [1/]

State	n	No. changes more than 10	25	50	$\Sigma \Psi_i$	T^+	T^*	α [2/]
		PERCENT						
Indiana	54	30	14	7	37	1,019	2.38	.017
North Carolina	84	53	37	18	53	2,280	2.21	.027
Oklahoma	62	28	9	4	37	1,296	2.24	.025
3-States Combined	200	111	60	29	127	NA	NA	NA

[1/] Notation: Z_i = (DES tract acres/DES farm acres) - (DES tract acres/recon. farm acres);

Ψ_i = the number of positive Z_i; $T^+ = \sum_{i=1}^{n} R_i \Psi_i$ where

R_i = rank of $|Z_i|$; $T^* = \dfrac{T^+ - [n(n+1)/4]}{[n(n+1)(2n+1)/24]^{\frac{1}{2}}}$ = test statistic for large sample approximation.

[2/] α = the lowest level at which H_o could be rejected, where H_o: $\Theta = 0$ and H_1: $\Theta \neq 0$.

Summary: The median of the differences in weights is significantly different from zero in each of the three states at the five percent level of significance. The weights based on DES farm acres were significantly different than the weights using reconciled farm acres. Supporting data indicates the DES weights were too large.

TABLE 4. -- Effect Of Reconciled FARM ACRES AND CATTLE On The Weighted Expansion [1/]

State	n	No. changes more than 10 PERCENT	25 PERCENT	50 PERCENT	$\Sigma \Psi_i$	T^+	T^*	α [2/]
Indiana	45	29	13	6	32	720	2.29	.022
North Carolina	44	28	19	11	25	499	.05	.960
Oklahoma	60	30	15	7	29	1,039	.91	.363
3-States Combined	149	87	47	24	86	NA	NA	NA

[1/] Notation: Z_i = (DES tract acres/DES farm acres)(DES cattle) - (DES tract acres/recon. farm acres)(recon. cattle); Ψ_i = no. of positive Z_i;

$$T^+ = \sum_{i=1}^{n} R_i \Psi_i \text{ where } R_i = \text{rank of } |Z_i|;$$

$$T^* = \frac{T^+ - [n(n+1)/4]}{[n(n+1)(2n+1)/24]^{\frac{1}{2}}} = \text{test statistic for large sample approx.}$$

[2/] α = the lowest level at which H_o could be rejected, where H_o: $\Theta = 0$ and H_1: $\Theta \neq 0$.

Summary: The median of the differences in the weighted expanded cattle is significantly different from zero only in Indiana. The weighted cattle using reconciled farm acres and cattle are not significantly different from the weighted cattle using DES farm acres and cattle in North Carolina and Oklahoma.

TABLE 5. -- Effect Of Reconciled FARM ACRES AND HOGS On The Weighted Expansion [1/]

State	n	No. changes more than 10 PERCENT	25 PERCENT	50 PERCENT	$\Sigma \Psi_i$	T^+	T^*	α [2/]
Indiana	21	12	8	4	15	157	1.44	.150
North Carolina	49	32	26	13	35	920	3.06	.002
Oklahoma	12	7	4	4	4	28	-	.424
3-States Combined	82	51	38	21	54	NA	NA	NA

[1/] Notation: Z_i = (DES tract acres/DES farm acres)(DES Hogs) - (DES tract acres/recon. farm acres)(recon. hogs); Ψ_i = no. of postive Z_i;

$$T^+ = \sum_{i=1}^{n} R_i \Psi_i \text{ where } R_i = \text{rank of } |Z_i|;$$

$$T^* = \frac{T^+ - [n(n+1)/4]}{[n(n+1)(2n+1)/24]^{\frac{1}{2}}} = \text{test statistic for large sample approximation.}$$

[2/] α = the lowest level at which H_o could be rejected, where H_o: $\Theta = 0$ and H_1: $\Theta \neq 0$.

Summary: The median of the differences in the weighted expanded hogs is significantly different from zero only in North Carolina. The weighted hogs using reconciled farm acres and hogs are not significantly different from the weighted hogs using DES farm acres and hogs in Indiana and Oklahoma.

APPENDIX B

Test of Hypothesis

Test of Hypothesis

The Wilcoxon distribution - free signed rank test is used. [1/] This test is particularly fitted to one sample location problems where there exists paired replicate data. The paired data generally refers to pretreatment and posttreatment observations for each record. The test then identifies a shift in location due to the application of a treatment. The location parameter is the median. Remember the median is the middle value if there are an odd number of observations, and the average of the two middle values if there are an even number of observations. The analog is the classical student's t test which uses the mean as the location parameter and assumes the population of differences is normal. The normal theory test on the mean is very sensitive to extreme observations compared to the nonparametric test.

Since each record has two observations the data takes on the following form:

Record i	X_i	Y_i
1	X_1	Y_1
2	X_2	Y_2
⋮	⋮	⋮
n	X_n	Y_n

The assumptions are: 1) $Z_i = Y_i - X_i$ where the Y_i are the DES data and X_i are the reconciled data. The model becomes $Z_i = \Theta + e_i$; $i = 1, \ldots, n$, where the e's are unobservable random variables and Θ is the unknown treatment effect; 2) e's are mutually independent; 3) each e comes from a continuous population that is symmetric about zero. Note, these assumptions do not state that the sample had to come from a normal population.

The test is performed by obtaining the absolute differences $|Z_i|$ and ranking these from smallest to largest (R_i). The indicator variable Ψ_i is defined as:

$$\Psi_i = \begin{cases} 1 \text{ if } Z_i > 0, \\ 0 \text{ if } Z_i < 0. \end{cases}$$

[1/] Hollander and Wolfe, Nonparameteric Statistical Methods, Wiley and Sons, N.Y., 1973 pgs. 26-32.

The test statistic is the sum of the products $R_i \Psi_i$

$$T^+ = \sum_{i=1}^{n} R_i \Psi_i.$$

This is essentially the sum of the postive signed ranks. An adjustment to T^+ is made for the large sample approximations. The null hypothesis is that the treatment effect is zero (H_o: $\Theta = 0$), here $P(X_i > Y_i) = P(Y_i > X_i) = \frac{1}{2}$. The alternative hypothesis is:

$$H_1: \Theta \neq 0$$

The Y_i and X_i observations for this study take on different values for the specific comparison to be tested. However, generally these two variables may be thought of as DES and reconciled data respectively. The treatment effect is really due to non-sampling errors. The null hypothesis and associated probability statement says it is just as likely that the DES data is greater than the reconciled data as it is for the reconciled data to be greater than the DES data. In terms of differences between the X_i and Y_i observations, the positive and negative differences of equal absolute magnitude will have the same probability of occurrence. The relative magnitude of these differences is represented by the ranking.