United States
Department of
Agriculture

Statistical
Reporting
Service

Statistical
Research
Division

Washington, D.C.

August 1984

Evaluation of the 1982 Corn Objective Yield Validation Survey

Fred B. Warren
Ronald J. Steele

EVALUATION OF THE 1982 CORN OBJECTIVE YIELD VALIDATION SURVEY. By Fred B. Warren and Ronald J. Steele, Statistical Research Division, Statistical Reporting Service, U.S. Department of Agriculture, Washington, D.C. 20250, SRS Staff Report No. AGES840821, August 1984.

ABSTRACT

Objective estimates of yield from 16 plots in each of 48 corn fields in Iowa and Missouri are compared with yields derived from weighed production.

Potential biases in farmer reported acreage and yield are discussed. The effect on estimated yield of a time lapse greater than seven days between final preharvest observations and actual harvest is examined. The relationship between estimated harvest loss and the time elapsed between final preharvest observations and postharvest gleanings is studied.

Some changes in operational objective yield procedures are recommended.

KEY WORDS

Objective yield, corn, validation.

TABLE OF CONTENTS

SUMMARY

Extreme care should be exercised in relating any differences in levels of estimates in this analysis to official estimates of the Crop Reporting Board or estimates from the operational objective yield surveys. In this study, samples were selected from a domain which included only 15 to 25 percent of the corn for grain fields in Iowa and 40 to 50 percent in Missouri.

For the validation study fields, the relative average differences between the objective estimates of yield and the actual yield were 1.1 percent for Iowa and 7.7 percent for Missouri. The observed mean difference of 1.42 bushels per acre (Bu/A) in Iowa is not statistically significant. However, the mean difference of 7.92 bushels per acre in Missouri was so large that the probability that there is no difference is less than one percent. For the two states combined, this mean difference was 4.81 bushels per acre.

The mean differences (objective estimate of yield minus actual yield) between the two states differed by 6.50 bushels per acre. The probability that a difference this large occurred by chance is less than 5 percent. The most plausible explanation is that survey procedures were applied in different manners in different states.

For Iowa, field acreages reported by the farm operator during preharvest interviews averaged 2.5 acres larger than the ASCS digitized acreages. Acreages reported after harvest averaged 0.4 acres larger. The 2.5 acre difference is statistically significant, but should not be interpreted as an indication of bias--it may represent a change in farmers intentions as a function of growing conditions. This analysis was not done for the Missouri samples because of suspected nonsampling errors in the data.

There was a significant difference (4.02 bushels per acre) between farmer reported yields and actual yields for Missouri, even though farm operators had access to weight receipts of actual production. This finding suggests some rewording of the question on the survey instruments may be appropriate. For example, farm operator adjustments for moisture content of grain at harvest were unknown. Although we can not make inferences from this study to the Acreage and Production (A&P) Surveys, this finding suggests that the use of A&P yield indications for official estimates should be the subject of further research.

The effect on yield estimates of a time lapse of more than seven days between final preharvest observations and actual harvest is significant. Using observations obtained within seven days of harvest results in a decrease in the average estimated net yields of 3.54 bushels per acre for Iowa, 2.18 for Missouri, and 2.83 for the two states combined. These differences are statistically significant for Iowa and for the two states combined, and nearly significant for Missouri. Even when all observations were obtained within seven

days of harvest, the objective yield estimate was significantly different ($\alpha = .025$) from the actual yield for the Missouri samples.

One-third of the validation fields were not harvested until after the November 1 cutoff date of the operational objective yield survey. There was a 2 Bu/A (33 percent) difference in harvest loss between fields harvested prior to, and after November 1. Although the mean differences were not significant, the variances were significantly different before and after November 1. These results are likely to fluctuate from year to year as a function of the corn harvest progress on November 1. This result indicates the survey cutoff date should be extended past November 1. It also indicates the inadvisability of imputing either historic or current year average harvest loss estimated solely on the basis of observations obtained prior to November 1.

RECOMMENDATIONS

Additional training and/or quality control feedback is needed to ensure procedures are being applied the same in all states.

Future validation studies should have larger sample sizes if it is deemed desirable to be able to detect differences in the various components of the yield estimates. Furthermore, samples should also be selected from the domain of fields whose operators do not intend to take their grain directly to commercial elevators to ensure that inferences can be made for the entire population.

Research should be undertaken into the wording of farmer reported acreage and production questions. Farm operators should be asked for moisture content of corn harvested during the postharvest interview.

Future validation studies should reexamine the effect of the elapsed time from preharvest observations until actual harvest. If the results from this study are reconfirmed, the operational procedures should be modified to ensure that observations are obtained within seven days of harvest.

Extension of the operational survey period beyond the November 1 cutoff date should be considered in light of the large relative difference in harvest loss before and after November 1.

Evaluation of the 1982 Corn Objective Yield Validation Survey

Fred B. Warren
Ronald J. Steele

INTRODUCTION

A recurring concern of the Statistical Reporting Service (SRS) has been that the procedures used in the objective yield surveys may result in yield estimates which are consistently too high or too low. Therefore, SRS has embarked upon a series of "validation studies" to determine the level of bias in objective yield survey procedures for the major crops.

This study was designed to detect significant differences between yield estimates based on the operational objective yield surveys and the derived yield based on weighed production. Each component of the yield estimate--reported acreage, biological yield, and harvest loss--was examined to determine potential sources of bias. Tests are made, and summary statistics presented for the individual states and for the two states combined.

STUDY DESIGN

This validation study was conducted in Iowa and Missouri, in conjunction with the operational corn objective yield survey in 1982. The validation study differed from the regular objective yield survey in the following respects:

1. The "validation fields" were sampled from that portion of the operational objective yield sample where the farm operators intended to haul their corn directly to a facility for weighing, and were willing to participate in the study. In Iowa, 23 of the 240 sample fields from the operational survey were used for the validation study. In Missouri, 25 of the 150 sample fields were used.

2. In the operational survey, the most accessible corner of the field is used as the starting point for locating one pair of sample plots. In this study, two pairs of sample plots were located from each of four corners of the field, resulting in eight pairs of sample plots rather than the usual one pair.

3. Regular monthly observations were taken on the immature plants in the pair of sample plots located by the operational procedures. No observations were taken in the other seven pairs of plots until the operational survey visit when the corn was found to be physiologically mature, at which time

observations were made on all eight pairs of sample plots. Then, if that part of the field containing each pair of plots was not harvested by the farmer within seven days, repeat observations for each unharvested plot were made on "mirror" plots within seven days of actual harvest.

4. The productive area of each field was digitized from aerial photographs taken by the ASCS. All grain harvested by the farmers was weighed and the percentage of moisture content measured.

5. Postharvest gleanings were taken in sample plots near (5 rows and 5 paces further into the field) each of the 16 preharvest plots.

6. The study was limited to an examination of the final estimates of yield based on observations on the mature crop. There was no attempt to evaluate the forecast procedures based on observations on the immature crop.

An analysis of yield components from the operational corn objective yield survey of 1980 indicated that the smallest differences which could be detected, based on a sample size of 25 fields/state and 16 plots/field, was 1.9 bushels/acre (1.4 bushels/acre for the two states combined). Given the large variability in yield between years, that analysis proved to be optimistic.

This study was not designed for inferences regarding multistate validation. Combined state estimates are included for comparison purposes only. Throughout this report, combined estimates for the two states are derived by weighting individual state values by their respective number of sample fields.

DEFINITIONS

To simplify terminology, the following definitions are used throughout the remainder of this report:

"Regular Yield" - Yield estimates based on validation study data obtained on the regular survey visit when the corn was first observed to be physiologically mature. In the operational objective yield program, ears are picked, weighed, and subsampled for laboratory analysis at this time to estimate final preharvest yield. The same procedures were used for the validation study, so "regular yield" is roughly comparable to final preharvest yield in the operational program.

"Mirror Yield" - Yield estimates based on additional preharvest observations taken after the "regular yield observations" and just before the field was actually harvested. Mirror yields were computed only if the field wasn't harvested within seven days after

the observations used to compute "regular yield" were obtained. Since destructive sampling is used to obtain "regular yield" estimates, it was necessary to locate replacement sample plots in these fields. These replacement plots were located in the same rows as the original plots and used the same starting points. The measurements were reversed so the plot was located towards the point of approach rather than away from it, and hence the term "mirror plots." These procedures were implemented to study the effect of delays from the time of "regular yield" observations to the time the field was actually harvested.

"At-Harvest Yield" - Yield estimates based on observations obtained within seven days of actual harvest of the sample plots. The "mirror yield" is used for those portions of fields which had "mirror plot" observations. The "regular yield" is used for the remaining plots in the study.

"Actual Yield" - Yield derived from the total weighed production and the ASCS digitized net acreage of corn for harvest of grain.

"Farmer Reported Yield" - Yield derived from farm operators reports of production and acres harvested for the validation field. Since the farm operators had access to weight receipts, this estimate is not independent from "actual yield."

The regular yield, mirror yield, and at-harvest yield are adjusted to 15.5 percent moisture content based upon laboratory determinations of moisture content of randomly sampled ears. Actual yield is adjusted to 15.5 percent based upon moisture determinations made at the commercial elevators. Farmer reported yields are used as derived without adjustment for moisture content. Farm operators were asked "How many bushels were or will be harvested from [this field]?"

ANALYSES

The validation data was analyzed to determine if:

1. The regular yield estimates differed significantly from the actual yield, by individual states, or for the two states combined;

2. The farmer reported acres of corn for harvest of grain were significantly different from the ASCS digitized net acres of corn for harvest of grain;

3. The time lag between physiological maturity and actual harvest had a significant effect on the estimated yield;

4. Estimated harvest loss was related to the time of harvest, gross yield or net acreage.

RESULTS

"Regular" and "mirror" average yields, farmer reported yields, and actual yield for each validation field are shown in the Appendix.

Overall Difference: Regular yield and actual yield are compared for each validation field to determine if there is a potential bias in the operational corn objective yield survey procedures. We test the hypothesis H_o: regular yield = actual yield, versus the alternative H_1: regular yield $\neq$ actual yield, with a paired t-test. Summary and test statistics are presented in Table 1.

We conclude that for the two states combined, there is a significant difference between estimates based on the operational corn objective yield procedures and actual yield. This difference is almost exclusively due to the Missouri samples. The bias in Iowa is negligible. The difference in the paired differences between the two states is 6.50 bushels/acre, with a standard error of 3.11 bushels/acre. The probability of obtaining this large a difference between states, if the states are equally biased, is less than 5 percent (t = 2.09, 46 d.f.). In the authors opinion, the most viable explanation for a difference this large between states is that the procedures were applied in different manners in different states.

Table 1.--Comparison of Regular Yield with Actual Yield (Bushels/Acre)

	Iowa n = 23		Missouri n = 25		Two States n = 48	
	Mean	S.E.	Mean	S.E.	Mean	S.E.
Regular Yield	130.2	6.4	110.0	6.2	119.7	4.6
Actual Yield	128.8	6.5	102.1	6.8	114.9	5.1
Paired-Difference	1.42	2.00	7.92	2.35	4.81	1.61
(t-statistic)	(0.71)		(3.37*)		(2.99*)	

* The probability of obtaining a larger t-statistic if there is no difference is less than 0.5 percent.

Farm Operator Reports: In the operational objective yield surveys (and the nonprobability Acreage and Production Survey), the farmer reported acreage is used in computing farmer reported yield. Prior to harvest, the farm operator is asked to report acres intended to be harvested for grain in the sample field (Form A). After harvest, the farm operator is asked to report actual acres harvested for grain in the sample field (Form D). The study was designed to compare the farmer reported acreages with the ASCS digitized acreage. However, the Form D data for Missouri was edited to conform with the ASCS digitized acreage. There is also some concern about the validity of the Form A data for Missouri, so this analysis was performed only for the Iowa data.

For Iowa, farmer reported acreages tended to be slightly larger than the digitized acreage. Form A acreage intended for harvest for grain averaged 2.5 acres (S.E. = 1.586) larger than the digitized acreage. The paired-difference t = 1.58, and the probability of obtaining a larger t-value if there were no difference is less than 7 percent. A single-sided alternative hypothesis was used in this case since Form A acreage is believed to be roughly equivalent to planted acreage, and acres harvested can not be greater than acres planted. The average difference, although significant, does not necessarily indicate a potential bias. We should expect substantial differences in years such as 1982 when crop conditions are less than optimum. Form D acreage actually harvested for grain averaged 0.4 acres (S.E. = 0.238) larger. The paired difference t-value for testing the hypothesis of no bias in postharvest reported acreage was 1.52. The probability of obtaining a larger t-statistic (if there is no bias) is less than 15 percent.

Although farm operators had weight receipts, and thus should have had a fairly accurate idea of actual production, the data indicate otherwise. The differences between farmer reported yield and actual yield are summarized in Table 2.

Table 2.--Differences Between Farmer Reported Yield and Actual Yield

State	Mean Difference	S.E.	d.f.	t
Iowa	-2.42 Bu/A	2.95 Bu/A	22	0.82
Missouri	4.02 Bu/A	0.76 Bu/A	24	5.26*
Combined	0.94 Bu/A	1.53 Bu/A	47	0.61

* The probability of obtaining a larger t-value if there is no difference is less than 0.1 percent.

The relative differences are 1.9 percent for Iowa and 3.9 percent for Missouri. Although the combined differences tend to offset each other for these two states, this should not be construed to imply they do so over all states. Some of this difference may be due to farm operators not adjusting reported yield to a standard 15.5 percent moisture content. Farm operators are asked "How many bushels were or will be harvested from [this field]?" They should also be asked for the moisture content of the harvested grain.

Delayed Harvest Effect: To determine if there is a significant difference in estimated yields between the time of physiological maturity and the time of harvest, the difference between regular yield and mirror yield is analyzed.

First, the hypothesis H_o: regular yield = mirror yield is tested for the restricted sample of fields with mirror unit observations. A two-tailed, paired-difference t-test is performed. Table 3 provides the summary statistics. These statistics are not consistent with data in the Appendix because the yields for fields which were partially (not completely) harvested during this time lag were computed differently for the purposes of this comparison. Only plots with mirror units were used to compute regular and mirror yields for the Table 3 statistics.

We reject the null hypothesis for Iowa and for the two states combined. There is a significant difference in yield estimates obtained when the corn is not harvested within seven days after taking observations at physiological maturity. Although we would not reject the null hypothesis for Missouri at the customary 90 or 95 percent confidence levels, the probability of obtaining a larger t-value is less than 11 percent.

Since some fields are harvested within seven days after the final preharvest observations, the differences shown in Table 3 do not accurately reflect the effect of this difference on the final estimated average yield for 1982. A more accurate estimate of its effect in 1982 can be obtained from the differences between the regular yield and the at-harvest yield. The estimated effect in bushels/acre is 3.54 for Iowa, 2.18 for Missouri, and 2.83 for the two states combined, and the probabilities of observing differences this large are less than 0.02, 0.12, and 0.005, respectively. The effect of this difference on final estimated yield will vary from year to year as a function of the relative progress of the corn harvest.

Table 3.--Differences Between Regular Yield and Mirror Yield Estimates: Summary Statistics

State	Mean Difference	S.E.	d.f.	t
Iowa	7.30 Bu/A	2.22 Bu/A	14	3.29*
Missouri	3.41 Bu/A	2.00 Bu/A	15	1.71
Combined	5.30 Bu/A	1.50 Bu/A	30	3.53*

* The probability of obtaining a larger t-value if there is no difference is less than 1 percent.

The at-harvest yields are compared with actual yields to determine if there is still an overall difference with the effects of this time lag removed. Differences between at-harvest yields and actual yields are summarized in Table 4.

When all observations were obtained within seven days of harvest, the overall difference between estimated yield and actual yield for the two states combined was reduced from 4.81 to 1.98 bushels/acre. This difference is not significantly different from zero. This change resulted from reduced yield estimates in each state. The relative magnitude of these changes were 2.7 percent for Iowa, 2.1 percent for Missouri, and 2.5 percent for the two states combined. Even with the effect of this time lag removed, there is still a significant difference between yield estimates and actual yield for Missouri.

November 1 Cutoff: The original analysis plan called for computing the correlations of harvest loss with size of field, gross yield, and the time lag from the date of preharvest observations to the date of postharvest gleanings. These correlations are shown in Table 5. The only one that is significantly different from zero is the correlation between harvest loss and gross yield for Iowa.

The effects of stopping all operational survey work on November 1 are now analyzed. Summary statistics for harvest loss are presented in Table 6.

Table 4.--Differences Between At-Harvest Yields and Actual Yields: Summary Statistics

State	Mean Difference	S.E.	d.f.	t
Iowa	-2.11	1.75	22	1.21
Missouri	5.74	2.32	24	2.47*
Combined	1.98	1.56	47	1.27

* The probability of obtaining a larger t-statistic if there is no difference is less than 2.5 percent.

Table 5.--Correlations with Estimated Harvest Loss

State--Time	ASCS Net Acreage	Estimated Gross Yield	Time Lag 1/
Iowa-Overall	0.214	-0.418 2/	-0.025
Before November 1	0.073	-0.376	-0.313
After November 1	0.336	-0.337	-0.104
Missouri-Overall	0.029	0.127	0.239
Before November 1	-0.095	0.131	-0.148
After November 1	0.123	0.271	0.355

1/ From date of regular final preharvest observations to date of postharvest gleanings.
2/ The probability of no linear relationship is less than 5 percent.

Table 6.--Harvest Loss Summary Statistics

State--Time	n	Est. Harvest Loss (Bu/Acre)		
		Average	S.E.	(t'-statistic*)
Iowa-Overall	23	8.247	0.9879	
Before November 1	14	7.345	0.9727	(1.03)
After November 1	9	9.649	2.0159	
Missouri-Overall	25	5.418	0.5226	
Before November 1	18	5.120	0.4334	(0.66)
After November 1	7	6.183	1.5486	
Both States-Overall	48	6.773	0.5780	
Before November 1	32	6.094	0.5204	(1.40)
After November 1	16	8.132	1.3549	

* $t' = (\bar{x}_1 - \bar{x}_2)/\sqrt{s_1^2/n_1 + s_2^2/n_2}$. A conservative approximation for the critical value of t' is given by Cochran [Biometrics, 20:191 (1964)].

The usual t-test for the difference between average estimated harvest loss before and after November 1 is inappropriate because the variances are significantly different($\alpha = .10$). Computed values of the folded F test statistic to test for unequal variances were 2.76 for Iowa, 4.97 for Missouri, and 3.39 for the two states combined. The difference in the average harvest loss for fields harvested prior to and after November 1 are not significantly different from zero ($\alpha = .10$).

Based on the previous test, we would conclude there is no significant difference between estimated harvest losses for fields harvested before or after November 1. However, there are substantial differences in both estimated gross yield and ASCS net acreage between fields harvested before and after November 1 (See Tables 7 and 8). Since covariances between these variables may be obscuring some relationship between harvest loss and time of harvest (before or after November 1), an analysis of covariance was performed. When harvest loss is corrected for changes in net acreage and gross yield, the effect of time of harvest, prior to or after November 1, becomes almost significant (p-value = 0.13).

If the average harvest loss for the 32 sample fields harvested prior to November 1 had been imputed for the 16 sample fields harvested after November 1, the average yield estimated for the two states combined would have been about 1 percent higher.

Summary statistics for gross yield ("regular yield" unadjusted for harvest loss) are presented in Table 7. These statistics are computed from observations taken prior to November 1. Although these differences are statistically significant, they are not indicative of a deficiency or bias in the operational survey procedures. One logical explanation for the differences is that they reflect farmers preferences for harvesting higher yielding fields first.

Table 7.--Gross Yield by Date of Harvest: Summary Statistics

State--Time	d.f.	Gross Yield (Bu/Acre) Mean		S.E.
Iowa				
Before November 1	13	149.9		5.64
After November 1	8	120.6		9.94
Difference	21	29.3		10.58
(t-statistic)			(2.77) 2/	
Missouri 1/				
Before November 1	17	121.9		5.99
After November 1	6	98.7		15.27
Difference	23	23.2		13.39
(t'-statistic)			(1.41)	
Both States 1/				
Before November 1	31	134.2		4.81
After November 1	15	111.0		8.85
Difference	46	23.2		9.22
(t'-statistic)			(2.40) 2/	

1/ Variances are significantly different ($\alpha = .10$) before and after November 1.
2/ The probability of obtaining a larger t (or t')-statistic if there were no difference is less than 5 percent.

The differences in the average gross yield of fields harvested prior to and after November 1 are substantial in all cases, and significant for Iowa and for the two states combined. The variances of gross yields before and after November 1 are significantly different for Missouri and for the two states combined.

Average ASCS net acreages of fields harvested prior to and after November 1 are presented in Table 8.

The differences in the means are not significant, but the variances of fields harvested before and after November 1 are significantly different ($\alpha = .05$) for Missouri, and are almost significantly different for the two states combined.

Table 8.--Net Acreage by Date of Harvest: Summary Statistics

State--Time	d.f.	Mean	S.E.
Iowa			
Before November 1	13	46.5	10.04
After November 1	8	50.1	14.04
Difference	21	3.6	16.81
Missouri			
Before November 1	17	31.5	4.16
After November 1	6	31.7	11.04
Difference	23	0.2	9.48
Both States			
Before November 1	31	38.1	5.06
After November 1	15	42.0	9.27
Difference	46	3.9	9.69

APPENDIX A

Net Yields by Sample Field: Iowa

	Sample Field Number	Net Yield (Bu/Acre)			
		Objective Estimate			
		Regular Units	Mirror Units	Farmer Reported	Actual
	53	118.9	--	128.4	124.6
	61	143.2	--	144.0	139.5
	78	147.8	--	128.4	162.8
	82	132.1	--	123.7	153.8
	98	129.0	115.5	133.5	130.9
	108	125.7	--	132.1	128.4
	114	144.9	133.3	129.9	128.2
	116	147.7	138.3	139.2	129.9
	121	106.2	--	106.2	104.8
	128	166.7	156.2	143.4	159.3
	130	161.9	154.0	170.4	156.3
	132	149.9	--	147.0	145.9
	135	168.6	152.4	149.5	153.8
	145	137.7	125.1	131.9	142.6
	159	148.8	151.3	156.0	142.0
	170	101.8	91.9	76.7	94.8
	174	92.3	--	98.1	92.1
	202	75.9	78.3	75.4	74.4
	205	73.4	78.2	63.0	70.9
	219	139.5	146.4	169.8	153.5
	230	165.8	167.8	140.6	167.9
	231	66.6	62.6	64.8	59.3
	235	150.4	146.0	154.3	146.2
n = 23 1/	AVE	130.2	126.7	126.4	128.8
n = 15 2/	AVE	131.9	126.5	126.6	127.3

1/ All sample fields. Regular yield is used for mirror yield if no mirror units were observed.
2/ Fields with mirror units.

APPENDIX A

Net Yields by Sample Field: Missouri

	Sample Field Number	Net Yield (Bu/Acre)			
		Objective Estimate			
		Regular Units	Mirror Units	Farmer Reported	Actual
	9	86.3	83.5	80.9	77.8
	18	129.8	--	120.2	114.9
	39	132.5	136.5	128.2	125.2
	40	51.3	48.8	43.9	42.9
	43	130.5	117.7	125.0	122.6
	51	96.2	--	91.3	85.0
	53	87.4	61.2	79.5	77.0
	54	59.8	60.4	57.5	55.0
	62	109.0	107.1	96.8	95.5
	63	119.2	--	134.2	121.4
	72	134.1	--	153.8	145.7
	73	108.8	117.0	117.3	109.3
	75	60.0	59.4	49.4	47.4
	77	80.9	--	71.1	67.1
	91	136.0	134.1	136.5	140.5
	92	145.7	142.3	140.0	140.1
	97	157.5	153.7	149.4	143.1
	104	76.7	--	51.8	50.2
	119	136.2	--	134.6	124.2
	121	124.2	126.7	147.7	148.8
	124	122.7	--	113.6	104.5
	127	118.2	110.6	96.2	94.3
	132	66.4	--	77.2	72.1
	139	154.3	157.3	156.0	148.9
	142	127.1	117.7	101.2	99.2
n = 25 1/	AVE	110.0	107.8	106.1	102.1
n = 16 2/	AVE	111.8	108.4	106.6	104.2

1/ All sample fields. Regular yield is used for mirror yield if no mirror units were observed.
2/ Fields with mirror units.