United States
Department of
Agriculture

Statistical
Reporting
Service

**Statistical
Research
Division**

**SRS Staff Report
Number AGES 820225**

SF 82-02

Estimating Dry Bean Acreage in Michigan

Ron Fecso
Jeff Geuder
Bob Hale
Steve Pavlasek

ESTIMATING DRY BEAN ACREAGE IN MICHIGAN, by Ron Fecso, Jeff Geuder, Bob Hale, and Steve Pavlasek, Research Division; Statistical Reporting Service; U.S. Department of Agriculture. Staff Report No. AGES820225.

ABSTRACT

The importance of dry beans as an export crop has increased in recent years. One third of all dry beans in the United States and 90 percent of the Nation's navy beans are grown in Michigan. In an effort to obtain more precise area frame estimates of the acreage of dry beans in Michigan, a specialized area sampling frame was constructed in a 16-county area. New techniques were used in a research effort, and three estimates were obtained. The precision of the estimates was considerably better than the direct expansion estimate from the JES. This paper documents the techniques used in frame construction and estimation. It also points out areas where alterations to the survey design can improve the precision of the estimates in subsequent years.

Keywords: Area frame, dry bean estimates, regression estimates, stratification.

* *
* This paper was prepared for limited distri- *
* bution to the research community outside *
* the U.S. Department of Agriculture *
* *

ACKNOWLEDGMENTS

The Michigan Dry Bean Project was initiated in November 1980 with the goal of an improved dry bean acreage estimate available for the August 1981 crop report. The success of the project, especially in meeting its tight schedules, must be attributed to the outstanding cooperation exhibited by the various groups involved.

Contributions to the project include:

Project initiation - Don Fedewa, Statistician in Charge, Michigan State Statistical Office (SSO).

Fiscal support - The Michigan Bean Commission, the Michigan SSO and the Sampling Frame Development Section.

Management - Don Fedewa, Michigan SSO; Wayne Gardner, Head, Sampling Frame Development Section (SFDS); Robert D. Tortora, Chief, Sampling Frames and Survey Research Branch.

Survey design and Project Leader - Ron Fecso, Mathematical Statistician, SFDS.

Questionnaire, survey instructions and enumeration - Michigan SSO.

Frame construction - Steve Pavlasek, Survey Statistician, SFDS.

Regression estimates and auxiliary data - Jeff Geuder and Ron Fecso, Mathematical Statisticians, SFDS.

Survey summary programming - Bob Hale, Mathematical Statistician, SFDS.

Typing - Phyllis Piper.

CONTENTS

SUMMARY

A Dry Bean Survey (DBS) was conducted in Michigan in July 1981 to obtain more precise acreage estimates than were available from the June Enumerative Survey (JES). Although a multiple frame survey (using the JES as the area portion) has yielded estimates of acceptable precision, there were some concerns about the area frame portion of the estimate. The area frame used in the JES does not estimate efficiently a specialized crop such as dry beans. Also, the JES is conducted in late May, before much of the dry bean crop is planted in Michigan, so there is an inherent downward bias in the JES estimate.

The area frame used in the DBS was constructed specifically for estimating acreage in a 16 county area around Lake Huron, an area which produced over 90 percent of the dry beans in Michigan. The direct expansion estimate (closed segment approach) from the DBS had a coefficient of variation (C.V.) of 8.21 percent. This compares with a C.V. of 12.99 percent from the 1981 JES. There were 205 segments in the 16 counties in the DBS, compared to 99 in the JES in this area. The gains in precision, were due to the increased sample size and the improved design.

A weighted estimate was obtained from the DBS. This estimate was significantly different from the closed estimate, a fact which further substantiates previous work showing a weighted estimate using total land for the weight is biased.

A regression estimate was also obtained from the DBS. Additional work was involved in assigning an auxiliary variable to the count units. The resulting estimate had only a slightly smaller standard error than the closed estimate because the lack of previous survey data resulted in the choice of some regression coefficients which were far from optimal. Improvements in stratification and the regression estimator are expected to lead to increased precision in subsequent years.

Estimating Dry Bean Acreage in Michigan

Ron Fecso

Jeff Geuder

Bob Hale

Steve Pavlasek

INTRODUCTION

The importance of dry beans as a major export crop of the United States has been enhanced since the industry recently signed a 70 million dollar agreement with Mexico which increases dry bean exports. In the U.S., one-third of all dry beans and 90 percent of the navy beans are grown in Michigan. Within the state, over 90 percent of the dry beans are grown in 16 counties in the "Bay-Thumb" area around Lake Huron (see Figure 1).

The economic impact of this region makes it important to be able to estimate dry bean acreage with a high degree of precision. The Michigan State Statistical Office (SSO) has used a multiple frame survey to estimate dry bean acreage. The June Enumerative Survey (JES), conducted in late May, makes up the area frame. A list frame survey was then conducted in July.

There were some concerns about the estimates from the multiple frame design. First, the timing of the area frame survey is not optimal. In Michigan, much of the dry bean crop is planted after the JES is conducted. Therefore, the estimate made from the survey data is based on planted acres and "intentions" to plant dry beans. Second, the precision of the multiple frame estimate has been decreasing (see Table 1). In 1980, after a major effort to update the list of producers, the precision of both the list frame and area frame estimates decreased considerably. One of the purposes of this project was to obtain a more precise area frame estimate.

The area frame used in the June Enumerative Survey (JES) is stratified by land use. Within each land use stratum, substrata called paper strata are created containing counties with similar agriculture. 1/ In 1980, 85 percent of the JES estimate of dry bean acres was accounted for by the two intensive agriculture strata (11 and 12). Within these strata,

---

1/ For further explanations of paper strata and area frame surveys see references (2) and (5).

Table 1--Michigan Dry Bean Acreage Estimates (000 Acres)

YEAR	List Frame		NOL Area		Combined	
	Estimate	C.V.	Estimate	C.V.	Estimate	C.V.
1976	568	12.1	-	-	568	12.1
1977	533	6.6	-	-	533	6.6
1978	305	4.4	225	11.25	530	5.4
1979	287	4.6	209	13.30	496	6.2
1980	379.5	6.6	138.5	24.30	518	7.4

Figure 1--County outline map of Michigan with 16-county area shaded.

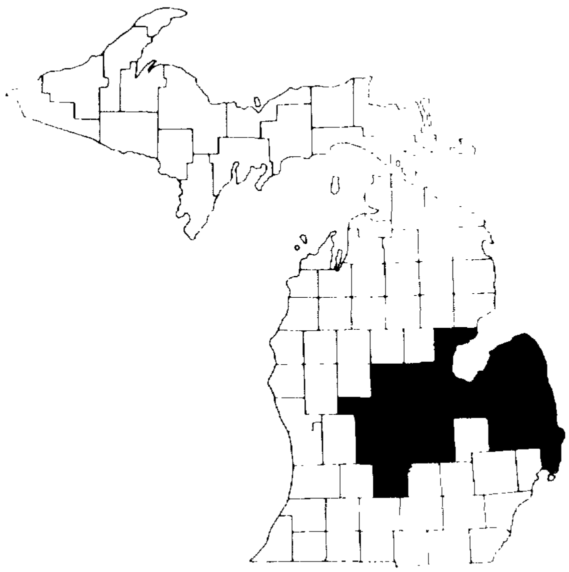

Table 2--Harvested acreage estimates for the 16 county area and for the state, 1979.

Crop	State	16 Counties
	(1,000 acres)	
Dry Beans	490	468
Corn	2500	972
Soybeans	970	482
Wheat	785	408
Oats	270	113
Sugarbeets	88	85

40 percent of the total estimate came from five paper strata located in the Bay-Thumb area.

The 16 county area is also well suited to other crops (see Table 2). As a result of the high concentration of agriculture, nearly one-third of the 343 JES segments were located in this area.

The SFDS found that by doubling the sample size in each paper stratum in the 16 county area, the CV for the JES dry bean estimate would drop from 13 to about 9.4 percent. Even an optimal allocstion of 200 segments to these paper strata would not reduce the CV below the desired 8 percent level. Therefore the development of a specialized frame was recommended.

SURVEY DESIGN

The area frame currently used by the Statistical Reporting Service (SRS) is constructed by delineating large, contiguous blocks of land and classifying them into one of several predetermined land use strata based on percent cultivated. These blocks are subdivided into primary sampling units, called count units. Sampled count units are further subdivided into sampling units, called segments.

One result of this method of frame construction is that segments within a count unit sometimes vary considerably with respect to agricultural content. In some cases, a segment in an intensive agriculture stratum contains very little cultivated land. In cases like this, sampling error is increased and the acreage estimate made from the frame loses precision. Since a new frame was to be constructed for the Dry Bean Survey (DBS), we had the opportunity to use and evaluate frame construction techniques and stratification procedures which were expected to yield improvements over those used in the operational program.

Construction of Count Units

The technique of most interest involved the construction of count units before defining strata. The stratifiers were instructed to draw off count units (on aerial photography) using a different approach from that used in the operational program. In contrast to the usual area frame methodology, strata definitions were not predetermined. All land was subdivided into count units such that segments within a count unit were homogeneous with respect to soil type, amount of cultivation, etc. The count units could range in size from one square mile (in which case the count unit was equivalent to the segment) to ten square miles.

As each count unit was constructed, certain auxiliary information was assigned to be used in stratification. The auxiliary information amounted to estimating (in terms of percentages) how much of the count unit was cultivated and how much was in woods, pasture, houses, water, and waste. The instructions for this process are shown in Appendix B.

Count unit boundaries were transferred to county highway maps and digitized. 2/ A data file was created for each county and edited to verify that count units were between one and ten square miles, and that no count unit numbers were missing or duplicated. The total digitized county area was also compared to the estimated Census area, allowing a two percent variation. The auxiliary information was also edited to ensure that the percentages totaled one hundred and that no count unit had missing data.

Stratification

The method of constructing count units for the dry bean frame allowed us to use multivariate clustering procedures to form strata. In addition to the various stratification variables assigned earlier, each count unit was assigned a soil classification code (which is a relative measure of the yield potential of the soil) and a subjective measure of the likelihood of finding dry beans in the count unit. This measure was based on previous years' survey data for segments in the area, climatological data, soil types and county acreage estimates. Cross tabulations of the variables were produced to determine the characteristics which were common to groupings of count units. The groupings were then used to form strata.

We were liberal in the number of strata created so we could evaluate the auxiliary information and its use in stratification and regression estimation. This was justified in that after the first year, strata could easily be collapsed if the data showed that such a design change would be more efficient. The following table describes the strata used in the DBS.

2/ A digitizer electronically measures the area of a polygon.

Table 3--Stratum Definitions for the Dry Bean Survey, 1981.

Stratum	Definition		
	Percent Cultivated	Soil Code	Possibility of Dry Beans
10	85-100	4	75-100
12	85-100	2-3	75-100
14	45- 84	4	75-100
16	45- 84	2-3	75-100
29	5- 45	2-4	75-100
30	55-100	4	55- 74
32	55-100	2-3	55- 74
36	15- 54	2-4	45- 74
39	5- 54	1	55- 64
40	55-100	3-4	25- 44
44	55-100	1-2	25- 44
48	5- 54	1-4	25- 44
50	75-100	3-4	0- 24
52	25- 74	3-4	5- 24
55	45-100	2	0- 24
56	5- 24	3-4	5- 24
57	15- 45	2	0- 24
58	5- 49	1	5- 24
59	5- 49	1	5- 24
60	0- 4	1-4	0-100

Sample Design

In the analysis of the California area frame (2), it was found that within a paper stratum, the standard error of the direct expansion estimate was approximately equal to the range of the individual sample expansions. 3/ Hence, ranges were estimated and used as an estimate of the standard errors necessary for the optimum allocation procedure. Based on cost estimates, 205 segments were selected with the final sample allocation as shown in Table 4.

Questionnaire Design and Data Collection

The questionnaire used in the DBS was similar to that used in the JES (See Appendix A). It was designed to collect data on all varieties of dry beans and selected other crops. No data was collected on livestock or economic items. The questionnaire contained sections for collecting both tract data (inside the segment only) and entire farm data, in order to compute a weighted estimate (described in the next section of this paper).

Twenty-one enumerators were trained during a one day training school held in Saginaw on July 6. The enumeration period was from July 7 through July 22. Whenever possible, data was

3/ Underscored numbers in parenthesis refer to literature cited in Reference.

Table 4--Sample allocation for the Dry Bean Survey, 1981

Stratum	N	Paper Strata	Repli- cations	n	N/n
10	1170	3	14	42	27.86
12	112	2	2	4	28.00
14	158	2	2	4	39.50
16	41	2	2	4	10.25
24	42	2	2	4	10.50
30	644	3	6	18	35.78
32	183	2	3	6	30.50
36	221	4	2	8	27.63
39	72	1	2	2	36.00
40	641	2	14	28	22.89
44	468	2	6	12	39.00
48	145	2	2	4	36.25
50	3069	5	7	35	87.69
52	1139	3	4	12	94.92
55	836	2	3	6	139.33
56	126	1	2	2	63.00
57	317	2	2	4	79.25
58	408	2	2	4	102.00
59	372	2	2	4	93.00
60	224	1	2	2	112.00

collected by personal interview with farm operators. Data collection by observation was used as a last resort.

The actual enumeration costs for the 205 one square mile segments averaged about $100 with a total cost breakdown 4/ as follows:

$11,238	Regular Salary
722	Overtime
4,262	Fringe Benefits
3,823	Mileage (18.5¢/mile)
500	Payroll Costs
108	Telephone
$20,653	Total Cost

ESTIMATION

The three estimators normally used in SRS area frame surveys are open, closed, and weighted. In the DBS, only the closed and weighted were used, along with a regression estimator. The open segment estimator was not considered because it has consistently been found to be least efficient for crop

4/ Cost computed by Bob Battaglia, Michigan SSO.

estimates from area frame surveys. There was also a problem with collecting total farm data which may be from outside the 16 county area used in the DBS. The closed, weighted and regression estimators are described below. The formulas for the estimators and the variances are shown in Appendix C.

Closed Segment Estimator

The closed segment estimator is the simple direct expansion estimate of the land inside the segment boundaries. Data is collected for every tract of land in the segment, and the estimate for the item in question is the product of the reported data and the expansion factor.

The closed segment approach is attractive because the estimate is unbiased, the concept is easy to understand, and the computations are not complex.

Weighted Segment Estimator

In order to compute the weighted segment estimator, we must collect entire farm data for every farm which is partially or entirely inside the segment boundaries. The reported farm data for the item being estimated is then weighted (or prorated) to the segment level. The weight used in the DBS is the ratio of tract acres (acres inside the segment boundaries) to total farm acres.

The weighted segment estimator usually has a smaller standard error than does the closed segment estimator. However, the data collection costs are higher due to the need to obtain entire farm data. Also, the weight used in the estimate has been shown to be biased due to a tendency by the respondent to underreport total farm acreage (4).

Regression Estimator

A regression estimator can often increase precision by using an auxiliary variable which is correlated with the item being estimated. In most SRS area frame surveys, no such auxiliary variable exists. For the DBS, a special effort was made to assign an auxiliary variable to every sample unit in the population. The variable and the procedure used to assign it are described next.

Auxiliary Variable for the Regression Estimator

For every count unit, (j), in the population we had the following information:

C_{jk} = assigned estimate of percentage of land area in cultivation in count unit j, county k

A_{jk} = digitized area (in acres) in count unit j, county k

n_{jk} = number of segments in count unit j, county k

Working within each county and using various information sources for the county, we assigned a value to each count unit which was an estimate of the percentage, p_{jk}, of the cultivated acreage in dry beans. The auxiliary variable assigned to segment i of count unit j was computed as follows:

$$x_{ijh} = \frac{1}{n_{jk}} (A_{jk} \cdot p_{jk} \cdot C_{jk}) \qquad (1)$$

The value of p_{jk} was assigned based on the soil type which predominated the count unit, the relative amount of dry beans in the county, and previous years' JES data. For example, if a count unit was located in an area with a soil type well suited to dry beans and the county, as a whole, had 35 percent of its cultivated land planted in dry beans the previous year, we would assign a value of p_{jk} about .35 or .40. If the previous years' survey data showed very large amounts of dry beans in the vicinity, we might increase p_{jk} to, say, .50. The values of p_{jk} were, admittedly, very subjective.

In order to remove some of the subjectivity, we adjusted the values of p_{jk} based on the 1980 county estimates of dry bean acreage made by the Michigan SSO. The formula for the adjustment was:

$$x^*_{ijk} = D_k \cdot \frac{x_{ijk}}{\sum_{j=1}^{N_k} \sum_{i=1}^{M_{jk}} x_{ijk}} \qquad (2)$$

where

x_{ijk} is the value of expected dry bean acres obtained in equation (1),

D_k is the estimated dry bean acreage in county k,

N_k is the number of count units in county k,

M_{jk} is the number of segments in count unit j in county k.

This value, x^*_{ijk}, then became the auxiliary variable used in the regression estimator.

Regression Coefficients

In most applications, the regression coefficient, b, is estimated from the results of the sample. However, there are instances when it is necessary to choose the value of b in advance of the survey. In simple random sampling, when b is pre-assigned, the regression estimate $x'_r = x' + b\ (Y - y')$ is unbiased. However, when b is estimated from the sample data, the regression estimate has a bias of order 1/n (1). In the DBS, there were less than 5 segments in some paper strata, and, hence, the potential for bias was extreme.

In order to obtain an unbiased regression estimate, we assigned the regression coefficients in advance of the survey. A regression coefficient is an estimate of the slope of the line plotting the reported survey data versus the auxiliary item. Hence, these slopes had to be estimated for each stratum. To estimate the slopes, variances of the item being estimated and the auxiliary variable were needed, as was an estimate of the correlation between the two variables. These estimates were not available from previous survey data. Thus, they had to be subjectively determined using all available information from the frame itself and from stratification materials.

The coefficients were also estimated from the sample data, both for segments in each paper stratum and for all segments in a land use stratum. These coefficients are shown in Appendix D.

State Estimate of Dry Bean Acreage

Since the estimates obtained from the DBS pertained only to the 16 county area, they had to be combined with an estimate for all other counties in the state. The 1980 JES sampling frame was altered to obtain appropriate expansion factors for segments outside the 16 county area covered by the Dry Bean frame. An estimate was then obtained for that area. The entire state estimator was then calculated as:

$$\hat{x} = x_d' + x_j'$$

where x_d' is the closed estimate from the DBS and x_j' is the closed estimate from the JES in the other counties. Since the two frames were independent the variance was computed as:

$$s^2(\hat{x}) = s^2(x_d') + s^2(x_j') \quad .$$

RESULTS

The Michigan SSO edited the survey data using the SRS Generalized Edit System. The edited data tape was then sent to the Sampling Frame Development Section to be summarized using the Area Frame Analysis Package (3). The results of the survey are discussed below for each of the three estimates computed.

Closed Estimate

Closed estimates were computed for all varieties (shown in Table 5) in the 16 county area. However, a state estimate was available only for total dry beans, since variety estimates were not made in the JES. As expected, the CV's for the varietal estimates were large for the more rare acreages.

Table 5--Closed and weighted estimates of dry bean acreage, by variety, for the 16 county area and for the state.

Variety	Closed		Weighted	
	Estimate	CV	Estimate	CV
	(acres)	(percent)	(acres)	(percent)
Navy	427,114	10.20	483,425	8.31
Dark Kidney	8,744	44.20	15,301	33.38
Light Kidney	8,874	52.67	11,728	32.00
Cranberry	19,370	34.00	23,119	30.02
Yellow Eye	1,713	67.45	1,896	60.57
Pinto	16,385	28.70	29,912	18.87
Black Turtle	122,205	15.23	126,992	9.87
Other	1,750	56.63	12,605	56.19
All but Navy	184,486	11.56	221,556	8.61
Total	611,600	8.21	704,981	6.67
JES 1/	16,823	52.56		
State Total	628,423	8.11		

1/ JES estimate from other counties

The DBS estimate of dry bean acreage for the state (628,423 acres) had a coefficient of variation of 8.11 percent. This compares with the JES estimate of 536,012 acres, with a C.V. of 13 percent. The increase of almost 100,000 acres is related to the later survey date, when the crop was almost entirely planted. The gain in precision is due to three factors: increased sample size, more precise stratification, and a different sample design.

Although increased sample size accounts for an initial drop in the CV of the dry bean estimate, post survey analysis shows that the specialty frame outperforms the JES stratification. By reallocating only 31 segments in the dry bean frame (Neyman allocation using post survey variance estimates) the CV can be expected to drop to 6.4 percent, almost 2 percent below the best CV attainable with the JES frame. Note that a reallocation of 31 segments is less than our usual 20 percent rotation. Thus, our first rotation can be used to achieve the optimum allocation. Further, by utilizing the multi-strata research design, strata can be collapsed making an additional 44 segments available for reallocation. The strata collapse is expected to reduce the CV to near 5 percent which, with improvements in the regression estimator or with ratio estimation, show the possibility of a 3 percent CV.

Weighted Estimate

As mentioned earlier, a weighted estimator using the ratio of tract acres to total farm acres has been shown to be biased. The DBS gave us another chance to compare the direct expansion estimate with a weighted estimate to evaluate the bias. The weighted estimate of 704,981 acres was greater than the closed segment estimate at the 90 percent confidence level.

The potential for bias exists in two places in the weighted estimator used in the DBS. There may be a tendency on the part of the respondent to underreport total acres in farm, as found by Hill and Farrar (4). Also, the questionnaire used in the DBS allowed respondents to account for wasteland in each field in the tract. It did not, however, include a line for waste in entire farm acres of dry beans. This value, then, could be overstated, which would result in an estimate with upward bias.

There was not a weighted estimate available from the JES to combined with that of the DBS to arrive at a state estimate. Therefore, the weighted estimate from the DBS was used only as supplementary information by the Michigan SSO.

Regression Estimate

Regression estimates were not available by variety because the auxiliary variable used considered all dry beans regardless of variety. The problems encountered in assigning the values of the regression coefficients were described earlier. Table 6 contains the regression estimates of dry bean acres in the 16 counties for three cases: (1) with the pre-assigned coefficients, (2) with coefficients computed from the sample data within each paper stratum, and (3) with coefficients computed from the sample within each stratum.

The estimate made using preassigned coefficients is unbiased, but the coefficients themselves were assigned using very subjective measures. The two estimates which used coefficients computed from the sample are both subject to bias. However, both are within one standard deviation of the closed estimate.

The small gains in precision from the regression estimates were probably offset by the additional cost involved in assigning the auxiliary variable and the possible biased low estimates of variance resulting from small sample sizes in the strata. However, the use of regression estimates in this survey was more a research effort than an operational procedure. In this light, the results are more encouraging. It is possible that improved estimates of the regression coefficients and more accurate auxiliary data (i.e. more highly correlated with reported acres of dry beans) can be developed for future surveys. It is not unreasonable to expect the regression estimates to be more precise in the future.

Table [illegible]--Regression estimates and coefficients of variation

	Pre-assigned Coefficients		Coefficients by Paper Stratum		Coefficients by Stratum	
	Estimate (acres)	CV (%)	Estimate (acres)	CV (%)	Estimate (acres)	CV (%)
16 county area	[illegible]96,454	8.24	571,[illegible]	5.47	579,995	7.60
State	[illegible]13,277	8.14	588,4[illegible]	5.52	596.818	7.53

CONCLUSIONS

The Dry Bean Survey was conducted to estimate the planted acreage of dry beans in the major producing area of the state. Using different stratification procedures, it was hoped to obtain an estimate of higher precision than is possible with the June Enumerative Survey. This was achieved, with the area frame estimate from the DBS having a coefficient of variation of 8.[illegible]1 percent and the area frame estimate from the JES having a C.V. of 13 percent. The sample size of the DBS was "larger" than that of the JES in that there were 205 segments in the 16 counties compared to 99 JES segments. The dry bean frame added to the precision of the estimate beyond the level caused by additional segments and reallocation can achieve a CV which is well below the maximum level desired (8%) in the research effort (the JES stratification could not).

The use of a regression estimator resulted in a very slight increase in precision over the [illegible]osed estimator. However, the regression estimator should be studied again next year, since previous survey data is now available for use in assigning the regression coefficients and improving the auxiliary variable. It is not unreasonable to expect a considerable increase in the precision of this estimate next year.

The following additional conclusions are made from the DBS:

(1) This year's survey design had twenty strata. This design was a research tool to determine the optimum number of strata and use of auxiliary information. Based on the survey results, we can identify areas where the design can be improved by collapsing and restratifying the ineffective strata.

(2) The use of auxiliary information was new and effective, although problems with nonhomogeneous count units still existed. The most striking example of the problem was a count unit which contained mostly land along a creek bed. The soil along the creek was not at all suited to dry beans, while the soil further from the creek was suitable.

Based on the amount of "poor" soil, and the other auxiliary data, the count unit was placed in stratum 50 (little dry beans anticipated). The count unit was selected for the sample, and the segment in the count unit fell in the area which had "good" soil and also was planted in dry beans. Because of the low sampling rate in Stratum 50, this segment contributed significantly to the estimated variance of dry bean estimate in the region.

In general, count units containing a mixture of land such as this, should be placed into the stratum with the higher sampling rate. Specifically, this count unit should have been given a soil rating corresponding to the "good" soil. By following this procedure, we may include segments with little or no acres of dry beans in the strata in which dry beans are anticipated. However, this would increase the variance much less than would the reverse case. This conclusion is, at this point, empirical, and should be studied further.

(4) Because the DBS was designed using replications, we can use a rotation scheme to allow a portion of the sampled segments to remain in the survey from one year to the next. This is a cost-saving procedure (new segments do not have to be defined and photography does not have to be purchased), and it allows us to make year-to-year comparisons on the portion of the sample which overlaps the surveys.

In addition, the segments selected for this year can be used in other ways, much the same as rotated JES segments are used in special surveys. Since the segments are already delineated on aerial photographs, it is very efficient to use them whenever there is a need.

REFERENCES

(1) Cochran, William G., Sampling Techniques. 3rd. ed. John Wiley and Sons, Inc., 1977.

(2) Fecso, Ron and Van Johnson, The New California Area Frame: A Statistical Study, U.S. Department of Agriculture, Statistical Reporting Service, No. 22, 1981.

(3) Fecso, Ron, Van Johnson and Jeff Geuder, Using SAS to Evaluate an Area Sampling Frame for Agricultural Surveys, Proceedings of the Sixth Annual SAS Users Group International Conference, 363-368, Orlando, Fl., Feb. 1981.

(4) Hill, George W., and Martha S. Farrar, Impact of Nonsampling Errors on Weighted Tract Survey Indications, U.S. Department of Agriculture, Statistical Reporting Service, 1977.

(5) Houseman, Earl, Area Frame Sampling in Agriculture, U.S. Department of Agriculture, Statistical Reporting Service No. 21, 1975.

APPENDIX A

Michigan Dry Bean Survey Questionnaire

Crop Reporting Board
Economics and Statistics Service
U.S. Department of Agriculture

Form Approved
O.M.B. Number 535-0043

JULY 1981
MICHIGAN
DRY BEAN ACREAGE
Enumerative Survey

Segment: ______________

Tract: ______________

County: ______________

STATE	DISTRICT	SEGMENT	TRACT
__ __	__ __ __ __	00000 __ __ __	__ __

Response to this survey is voluntary and not required by law. However, cooperation is very important in order to establish actual dry bean acreage planted. Facts about your farm will be kept confidential and only used in combination with similar reports from other producers.

START TIME	OFFICE USE
	006

1. I need to make sure that we have your *(the operator's)* name and address complete and correct.

   Name of Farm, Ranch or Operator: ______________

   Name of Operator: ______________
   (Last) *(First)* *(Middle)*

   Address: ______________
   (Route or Street)

   ____
   (City) *(State)* *(Zip)*

   Telephone Number: (___) ______________
   (Area Code)

2. Is the operation named above: Individually operated ·1 ☐ Partnership or Joint ·2 ☐ Manager Land ·3 ☐ .... Enter Code [845]

3. Are there any other persons living in this household who operate a farm inside the red segment boundary.

   ☐ No - *Continue* ☐ Yes - *Enter Name* ______________
   (Assign tract on Part ID, go to Item 4.)

4. Do you operate land inside the red segment boundary under any other name or land arrangement than the one listed above?

   ☐ No - *Continue* ☐ Yes - *Assign another tract letter for other arrangement.*

SECTION A — ACREAGES OF FIELDS AND CROPS INSIDE BLUE TRACT BOUNDARY

How many acres are inside this blue tract boundary drawn on the photo . . . . . . **Acres** [.]

Now I would like to ask about each field inside this blue tract boundary and its use in 1981.

Field Number. . .	1	2	3	4
1. TOTAL ACRES in Field	.	.	.	.
2. Crop or Land Use *(Specify)*				
3. Woods, Waste, Roads, Ditches, etc.	.	.	.	.
4. Occupied Farmstead or Dwelling	843 .	.	.	.
5. Idle Cropland — Idle during 1981	857 .	857 .	857 .	857 .
6. Corn	530 .	530 .	530 .	530 .
7. Soybeans	600 .	600 .	600 .	600 .
8. Sugar Beets	691 .	691 .	691 .	691 .
9. Irish potatoes	552 .	552 .	552 .	552 .
Dry Edible Beans				
10. Navy (Pea)	010 .	010 .	010 .	010 .
11. Dark Red Kidney	020 .	020 .	020 .	020 .
12. Light Red Kidney	030 .	030 .	030 .	030 .
13. Cranberry	040 .	040 .	040 .	040 .
14. Yellow-eye	050 .	050 .	050 .	050 .
15. Pinto	060 .	060 .	060 .	060 .
16. Black Turtle	070 .	070 .	070 .	070 .
17. Other *(Specify __________)*	080 .	080 .	080 .	080 .
Small Grains				
18. Winter Wheat	540 .	540 .	540 .	540 .
19. Rye	547 .	547 .	547 .	547 .
20. Oats	533 .	533 .	533 .	533 .
21. Barley	535 .	535 .	535 .	535 .
22. Alfalfa and Alfalfa mixtures	653 .	653 .	653 .	653 .
23. HAY OTHER HAY Kind				
Acres	654 .	654 .	654 .	654 .
24. Other Crops	— — — .	— — — .	— — — .	— — — .

SECTION A — ACREAGES OF FIELDS AND CROPS INSIDE BLUE TRACT BOUNDARY (Cont'd)

Field Number. . .	5	6	7	8	9	OFFICE USE
						Total Acre
1. Total Acres in Field	.	.	.	.	.	840
2. Land Use Name						Sum of Woods, etc.
3. Woods, Waste, etc.	.	.	.	.	.	841
4. Occupied Farm-stead	.	.	.	.	.	
5. Idle Cropland	857 .	857 .	857 .	857 .	857 .	
6. Corn	530 .	530 .	530 .	530 .	530 .	
7. Soybeans	600 .	600 .	600 .	600 .	600 .	
8. Sugar Beets	691 .	691 .	691 .	691 .	691 .	
9. Irish potatoes	552 .	552 .	552 .	552 .	552 .	
Dry Edible Beans						
10. Navy (Pea)	010 .	010 .	010 .	010 .	010 .	
11. Dark Red Kidney	020 .	020 .	020 .	020 .	020 .	
12. Light Red Kidney	030 .	030 .	030 .	030 .	030 .	
13. Cranberry	040 .	040 .	040 .	040 .	040 .	
14. Yellow-eye	050 .	050 .	050 .	050 .	050 .	
15. Pinto	060 .	060 .	060 .	060 .	060 .	
16. Black Turtle	070 .	070 .	070 .	070 .	070 .	
17. Other *(Specify ________)*	080 .	080 .	080 .	080 .	080 .	
Small Grains						
18. Winter Wheat	540 .	540 .	540 .	540 .	540 .	
19. Rye	547 .	547 .	547 .	547 .	547 .	
20. Oats	533 .	533 .	533 .	533 .	533 .	
21. Barley	535 .	535 .	535 .	535 .	535 .	
22. Alfalfa Hay	653 .	653 .	653 .	653 .	653 .	
23. Other hay — Kind						
Other Hay—Acres	654 .	654 .	654 .	654 .	654 .	
24. Other Crops	— — — .	— — — .	— — — .	— — — .	— — — .	

	Acres
What are the total acres you operate under this land arrangement. Include all cropland, woodland, pastureland, wasteland, and rented land.	900 .

Considering all land you operate, what are the total acreages of Dry Edible Beans planted? Include land both inside and outside red segment boundary.

TOTAL DRY BEAN ACREAGE

Kind	Acres Planted *(Acres)*
Navy	110 .
Dark Red Kidney	120 .
Light Red Kidney	130 .
Cranberry	140 .
Yellow-eye	150 .
Pinto	160 .
Black Turtle	170 .
Other *(Specify _______)*	180 .
Total All Kinds	190 .

Refer to face page to check box. Is operation partnership or joint?

☐ **Yes** · *Continue* ☐ **No** · *Go to Response Code*

Now I would like to identify the other persons in this joint farming operation *(excluding landlords)* so that dry bean acreages you report are not duplicated.

Name ______________________________ **Telephone No.** ____________
(Last) *(First)* *(Middle)*

Address ____
(Rt. or Street) *(City)* *(State)* *(Zip)*

Is he a: Partner ______ **Corporate Member** ______ **Manager** ______ **Other** ______

How many acres of Dry Beans are in the Joint arrangement . . . . **Acres** ______

Partnership or Corporation Name ____________________________

Response Code *(Circle number)*

1 — **Completed by operator**
2 — **Completed by other** ____________ *specify*
3 — **Inaccessible** *(observed)*
4 — **Refusal** *(observed)*

ENDING TIME	R.C.
	10

☆ U.S. Government Printing Office: 1981—[illegible]

APPENDIX B

STRATIFICATION PROCEDURES FOR THE MICHIGAN DRY BEAN FRAME

I. GENERAL:

In the past our area frames have been generalized frames intended to do a fairly good job of estimating major crop and livestock items, but not highly efficient for any particular item.

The objective of a specialized area frame is to maximize the sampling efficiency for a specific item. For this project, the specific item is dry beans in a 16 county area in Michigan.

To accomplish this objective it will be necessary to achieve a finer stratification by classifying each frame unit according to its probability of containing dry beans and the amount of potential dry bean area. This is somewhat obvious from the amount of cultivated land present. However, the probability of the frame unit containing dry beans will have to be determined by an analytic approach using crop calendars, LANDSAT imagery, field travel, and any ancillary data (such as soil type, field pattern, presence or predominance of competing crops, etc.) that may be available.

II. PROJECT INFORMATION:

1. State: Michigan

2. Area(s): 16 Counties in Bay-Thumb region:

County	(000) Acres Dry Beans	County	(000) Acres Dry Beans
Huron	79	Isabella	15
Tuscola	66	Arenac	13
Gratiot	59	Midland	12
Saginaw	50	Lapeer	10
Bay	48	Ionia	10
Sanilax	46	Shiawassee	8
Eaton	20	Clinton	8
Montcalm	19	St. Clair	5

3. Primary Work Unit: County

4. Material:
    1. County Statistics
    2. Crop Calendars
    3. ASCS Photo Index Sheets
    4. 1:250,000 LANDSAT Imagery
    5. Maps and Map Transparencies

5. Requirements: Stratify area of land within a county into homogeneous frame units. The stratification variables (cultivated land, woods, pasture, urban areas, water, etc.) must be evenly distributed throughout the frame unit. Such a distribution of stratification variables will allow a breakdown of the frame unit such that all resulting segments will have approximately the same characteristics as the frame unit.

   Using multi-temporal LANDSAT coverage, along with crop calendars, assign the probability of the occurence of dry beans.

III. STRATIFICATION GUIDELINES:

1. Reliance on physical boundaries must be emphasized; especially important are section lines when bounded by roads or other observable boundaries.

2. Each county will have a frame worksheet on which stratification variables will be recorded by frame unit.

IV. STRATIFICATION WORK PROCEDURES:

A. Phase One

1. Obtain the photo index sheets for the county to be worked.

2. Prepare a frame worksheet for the county, filling in all essential information. (Stratifiers name, County Name, County Code, Ranges and Townships boundaries for the county).

3. Delineate homogeneous areas of land containing cultivation (field patterns visible), wood, water or houses on the clean overlay. Use an orange grease pencil. The frame units must have clearly visible boundaries. Where possible, section lines should form the frame unit boundaries. The minimum frame unit size is one section, and the maximum is ten sections. Generally, they should range from 4 - 8 sections.

4. After the stratification is completed on all PI sheets for the county, identify the frame units on the PI sheets. Start with PI sheet number one and begin numbering the frame units in the upper right hand corner using the standard serpentine procedure; then go to PI sheet 2, etc. The frame unit identification will be a four digit number incorporating the PI sheet number and a three digit frame unit number, thus:

PI Sheet Number	Frame Unit Identification
1	1001-1999
2	2001-2999
3	3001-3999
4	4001-4999

so that 2127 would be frame unit 127 on PI sheet number 2.

5. After the frame units have been identified, list them on the frame worksheet. Classify the frame units as to percentages of land cultivated, woods, urban, miscellaneous (factories, etc.), and water. The total of these items must equal 100 percent. These classification estimates must be accurate, so a grid should be used when possible.

6. If in classifying the frame units it becomes evident that a frame unit should be split, divide the unit and assign the new part the next unassigned frame unit identification number. Footnote this frame unit on the worksheet and note what frame unit it was split from. This should make it easier to locate the unit, if we need to look at it at some later date.

B. Phase Two

1. Obtain the appropriate Landsat scene(s), the county overlay for the Landsat, the overlay showing existing JES segments within the county and a clear overlay. Align the county overlay on the various Landsat scenes so that you feel comfortable in moving the overlay from scene to scene.

2. Place the clear overlay over the county overlay, tape them together and copy the stratification from the PI overlay to the Landsat overlay using a red fineline lumacolor.

3. The goal of this phase is to evaluate, using Landsat, the basic stratification done from the PI in phase one. Moving the county overlay and frame overlay from scene to scene as necessary, check the classification of the frame units as to percentages cultivated, woods, urban, miscellaneous, and water. Be certain that each frame unit looks as "unique" on the Landsat as it did on the PI. Keeping in mind the difference in dates flown between the PI and Landsat, look for cleared woods or grown over land. Any apparent differences should be checked and if necessary, split or combine frame units. If you can discern pasture or hay, enter a percentage for these on the worksheet.

4. If any frame units need to be split based on the Landsat analysis, follow the procedures outlined in step 6 of phase one. Before splitting any frame unit, be sure to look at that unit on the PI. Any changes made to a frame unit based on Landsat <u>must also be made on the PI overlay</u>.

5. Using soil data obtained from Michigan, assign a soil code to each frame unit and record it in the soil code box on the worksheet.

C. Phase Three

1. Assemble all the materials used in phases one and two.

2. Review the work for consistency and completeness. Discuss any problems with the appropriate person.

3. Review the percentages of cultivated, woods, urban, and water assigned. If you disagree with the classification breakdown, enter reviewer's percentage on the right side of the data boxes on the worksheet using a red pencil.

4. The last step of this phase is to enter a dry bean probability on the worksheet for each frame unit. In Feeso's research proposal, he states that " . . . dry beans are negatively correlated with hay . . ." Basically, the dry bean probability could be expressed as a percentage by taking 100 minus the percentages of woods, urban, miscellaneous, water, and hay/pasture within each frame unit.

   Before assuming that this basic formula is accurate, however, you should review the multitemporal Landsat coverage, crop calendar and any ancillary data available. Taking all this into consideration, assign a dry bean probability between 0-100.

D. Phase Four

Using the PI, draft the final frame onto the appropriate 1/2" = 1 mile county map, using an orange pencil. Be sure that each frame unit has an identification number entered.

E. Phase Five

Supervisor review final frame for completeness and accuracy.

APPENDIX C

Closed Estimate

The direct expansion estimate (closed segment approach) is computed as follows:

$$x' = \sum_{h=1}^{L} x'_{hp} = \sum_{h=1}^{L} \sum_{p=1}^{p} \frac{N_{hp}}{n_{hp}} \sum_{i=1}^{N_{hp}} x_{hpi} \tag{1}$$

where N_{hp} = total number of segments in stratum h, paper stratum p,

n_{hp} = sampled number of segments in stratum h, paper stratum p,

x_{hpi} = acres of dry beans in the i^{th} segment of paper stratum p, in land use stratum h

The variance of the closed estimate is computed as follows:

$$s^2_{x'} = \sum_{h=1}^{L} s^2_{x'_{hp}} = \sum_{h=1}^{L} \sum_{p=1}^{p} s^2_{x'_{hp}} \tag{2}$$

where

$$s^2_{x_{hp}} = \frac{N^2_{hp}}{n_{hp}} s^2_x = \frac{N^2_{hp}}{n_{hp}} \left(1 - \frac{n_{hp}}{N_{hp}}\right) \sum_{i=1}^{n_p} \left(\frac{x_{hpi} - \bar{x}_{hp}}{n_{hp} - 1}\right)^2 \tag{3}$$

and

$$\bar{x}_{hp} = \frac{1}{n_{hp}} \sum_{i=1}^{n_p} x_{hpi} \tag{4}$$

Weighted Estimate

In order to compute this estimate, entire farm data was collected for every tract operation. A weighted value of tract dry bean acres for the i^{th} segment was computed as follows:

$$x^*_i = \sum_{j=1}^{J_i} x_{ij} \left(\frac{t_{ij}}{f_{ij}}\right) \tag{5}$$

where

x_{ij} = reported total farm dry bean acres for tract operator j in segment i,

t_{ij} = acres inside the segment boundaries for tract operator j in segment i,

f_{ij} = total farm acres for tract operator j in segment i,

J_i = number of tracts in segment i.

The estimate, then, is computed using the same form as equation (1), i.e.

$$x^* = \sum_{h=1} x^*_{hp} = \sum_{h=1}^{L} \sum_{p=1}^{p} \frac{N_{hp}}{n_{hp}} \sum_{i=1}^{n_{hp}} x^*_{hpi}$$

The variance is computed in the same manner as in equation (3), substituting the weighted acreage (x^*_{hpi}) in place of the raw acreage (x_{hpi}).

Regression Estimate

In order to compute the regression estimate, auxiliary variables were assigned to every population sampling unit. The variable is defined as "predicted acres of dry beans" and is denoted by y_{hpi}. The sum of these in each paper stratum is denoted by:

$$Y_{hp} = \sum_{i=1}^{N_p} y_{hpi}. \tag{7}$$

The regression estimate is then:

$$x'_R = \sum_{h=1}^{L} \sum_{p=1}^{r} [x'_{hp} + b_{hp} (Y_{hp} - y'_{hp})], \tag{8}$$

where

x'_{hp} is defined by equation (1),
y'_{hp} is defined in the same way as x'_{hp} for the auxiliary variable,
Y_{hp} is defined in equation (7), and
b_{hp} is the regression coefficient.

The variance of the regression estimate is computed within paper strata as follows:

$$s^2_{x_{Rhp}} = (N^2_{hp}) \frac{1 - \frac{n_{hp}}{N_{hp}}}{n_{hp}} (s^2_{y'_{hp}} - 2b_{hp} s_{yx_{hp}} + b^2_{hp} s^2_{x'_{hp}}) \tag{9}$$

where

$s^2_{x'_{hp}}$ is defined in equation (3),

$s^2_{y'_{hp}}$ is defined in the same way for the auxiliary variable, and

$$s_{yx_{hp}} = \sum_{i=1}^{n_p} \frac{(x_{hpi}\ y_{hpi} - \bar{x}_{hp}\ \bar{y}_{hp})^2}{n_p - 1}$$

The variance of the regression estimate is then:

$$s^2_{x_r} = \sum_{h=1}^{L} \sum_{p=1}^{p} s^2_{x_{Rhp}} \qquad (10)$$

APPENDIX D

Regression Coefficients

The following table shows the regression coefficients for three cases: (1) pre-assigned by stratum, (2) computed from the sample by paper stratum, and (3) computed from the sample by stratum.

Stratum	Paper Stratum	Regression Coefficients: Pre-assigned	Computed from Sample: by paper stratum	Computed from Sample: by stratum
10	1	.70	.374	.550
	2		-.480	
	3		1.271	
12	1	.90	7.430	8.920
			18.026	
14	1	.75	.664	1.571
	2		1.567	
16	1	.70	.827	.685
	2		7.529	
29	1	.95	-3.835	.420
	2		.371	
30	1	.35	1.656	1.301
	2		1.601	
	3		.896	
32	1	.70	1.359	1.234
	2		.439	
36	1	.70	24.987	.635
	2		122.642	
	3		3.215	
	4		1.481	
34	1	1.20	.756	.756
40	1	.40	-.654	.119
	2		.033	
44	1	.40	-.695	-.572
	2		-.262	
48	1	.40	4.433	2.250
	2		-6.221	

Stratum	Paper Stratum	Regression Coefficients: Pre-assigned	Computed from Sample: by paper stratum	Computed from Sample: by stratum
50	1	.15	.296	.581
	2		1.694	
	3		.358	
	4		.788	
	5		.787	
52	1	.30	-.041	.313
	2		.000	
	3		-1.097	
55	1	.30	.047	-.141
	2		1.469	
56	1	.15	0.000	0.000
57	1	.15	0.000	.912
	2		.331	
58	1	.15	0.000	8.057
	2		0.896	
59	1	.35	0.000	0.000
	2		0.000	
60	1	.30	0.000	0.000