

1 Supporting Statement – Part B

LOCAL FOOD MARKETING PRACTICES SURVEY

OMB No. 0535-NEW

B. COLLECTION OF INFORMATION EMPLOYING STATISTICAL METHODS

- 1. Describe (including a numerical estimate) the potential respondent universe and any sampling or other respondent selection method to be used. Data on the number of entities (e.g., establishments, State and local government units, households, or persons) in the universe covered by the collection and in the corresponding sample are to be provided in tabular form for the universe as a whole and for each of the strata in the proposed sample. Indicate expected response rates for the collection as a whole. If the collection has been conducted previously, include the actual response rate achieved during the last collection.**

The target population for the Local Food Marketing Practices Survey is any farm that sells its products directly to consumers or to retail outlets that in turn sell directly to consumers. In 1975 the USDA, the Office of Management and Budget, and the Census Bureau agreed on a definition of a farm that is still in use today: "A farm is currently defined, for statistical purposes, as any place from which \$1,000 or more of agricultural goods (crops or livestock) were sold or normally would have been sold during the year under consideration."

The sample for this survey will total approximately 56,000 units. Units will be drawn both from NASS's List Frame and a list developed independently from the NASS List Frame through a collaboration with the Multi-Agency Collaboration Environment (MACE). Response to this survey is voluntary.

- 2. Describe the procedures for the collection of information including:**
 - **statistical methodology for stratification and sample selection,**
 - **estimation procedure,**
 - **degree of accuracy needed for the purpose described in the justification,**
 - **unusual problems requiring specialized sampling procedures**

The target population for the Local Food Marketing Practices Survey is any farm that sells its products directly to consumers or to retail outlets that in turn sell directly to consumers.

The sample size determination is based on the assumption of simple random samples from the NASS list and the MACE list. This approach provides a conservative estimate of sample size because the stratified sampling design,

which will be described later, should provide more precise estimates than would be obtained from a simple random sample.

The sample size calculations are those derived for the two sample capture-recapture study. The original capture-recapture approach assumed a closed population (no farms entering or leaving the population during the course of the study) and that each farm had an equal probability of being caught. The assumption of a closed population is met here because the samples from the NASS list and the MACE list will be taken concurrently. The assumption of the probability of each farm being equally likely to be caught is not met. Large farms are more likely to be included on a list than small farms; new farms are less likely to be on a list than older farms; etc. The analysis will account for this “differential catchability” of farms, as discussed later.

The sample for the Local Food Marketing Practices Survey will be drawn from two sources: NASS’s List Frame (ELMO) and a list of potential local foods producers developed in collaboration with the Multi-Agency Collaboration Environment (MACE).

The ELMO sampling frame is comprised of:

1. All known farms (as defined in Part 1, above) on NASS’s List Frame.
2. Other entities on NASS’s List Frame identified as potentially in the target population (farms that sell their products directly to consumers or to retail outlets that in turn sell directly to consumers).

Within each state, records will be stratified into one of the following groups:

- a. Farms known to be in the target population.
- b. Farms that are *likely* in the target population.
- c. All other farms.
- d. Other entities that have been identified as potentially belonging in the target population.

Strata a, b, and c will be sub-stratified using ELMO control data, such as state and farm type. In contrast, stratum d, which does not have any control data, will not be sub-stratified. Within each stratum or sub-stratum, a simple random sample will be selected.

The MACE^a list of local food operations is derived from publically available web-based information. The MACE portion of the sampling frame consists of all records provided to NASS by MACE. The MACE records will be stratified by state and auxiliary data provided by MACE.

^aThe MACE website, David, Please put MACE’s internet site here, discusses performances metrics on past and ongoing projects.

A Mark-Recapture Sampling Design^b is used to derive sample sizes from the ELMO and MACE sampling frames.

NASS and MACE Sample Size Determination:

Let M be the NASS mark sample size.

Let C be the MACE recapture sample size.

Define R to be the relative sample size, i.e. $R = C/M$.

Using the Robson and Regeir (1964) method for determining sample size, the relative size of the capture and mark sample must be specified. It is anticipated that many of the operations on the MACE list will have relatively low sales values, and hence better estimates for total value of sales will be achieved by taking the value of the relative sample size to be less than one.

Define B to be an upper bound for the target population size, and let

$$D = \left(Z_{1-\frac{\alpha}{2}} \frac{(1+e)}{e} \right)^2$$

where e is the relative margin of error for a level of confidence $(1-\alpha)$.

Then the required sample size, M, is the positive root of:

$$R(B-D)M^2 + DB(R+1)M - DB^2 = 0$$

An R value of 0.80, an alpha (α) of 0.05, a relative margin of error (e) of 0.05, and an upper bound for the total population size (B) of 250,000^c were used to derive a target sample size. That target sample size (approximately 39,000) was then divided by a projected response rate of 0.70 to reach the effective sample size of 56,000.

- For a local foods farm to be captured, it must be (1) on the NASS list frame, (2) included in the sample, (3) respond to the questionnaire, and (4) correctly

^b Robson, D. S. and Regier, H.A. (1964), Sample Size in Petersen Mark-Recapture Experiments, Transactions of the American Fisheries Society, Vol. 93, No. 3, pp 215 – 226.

Evans, M.A., Bonett, D.G., McDonald, L. (1994), A General Theory for Analyzing Capture-recapture Data in Closed Populations, Biometrics, 50, pp 396 - 405.

^c USDA Economic Research Service (January 2015), Report to Congress on Trends in US Local and Regional Food Systems, Administrative Publication Number 068, p. 87.

classified as a local foods farm. The sampling weight accounts for sample inclusion. The sampling weight of each responding record on the NASS list frame will be adjusted to account for undercoverage, nonresponse, and misclassification. A logistic regression model of the probability of coverage will be developed using the sample records (from both the NASS and MACE lists) identified as being local food farms. A record on the NASS list will have a response of one, and a record not on the NASS list but on the MACE list will have a response of zero. The covariates in the model will be used to account for differential catchability, including the age of the farm, the land in the farm, and total value of production. The reciprocal of the estimated probability will be the coverage weight. A sample of operations responding as being in business and out-of-business will be drawn and will be re-contacted to verify their status. If misclassification is present, the probability of correct classification will be estimated for each record responding as a local foods farm. In addition to the coverage and misclassification adjustments, nonresponse weights will be calculated by a reweighting within each substratum or stratum based on the number of useable reports.

- The NASS list frame and the MACE list frame may have different sampled populations for the same target population. Consider two cases. First, a set of operations may be in the NASS-list-frame sampled population but not in the MACE-list-frame sampled population. The coverage probabilities for operations in this set will tend to be overestimated, resulting in population estimates that are biased downwards. In the second case, a set of operations may be in the MACE-list-frame sampled population but not the NASS-list-frame sampled population. The coverage probabilities for operations in this set will tend to be underestimated, causing population estimates to be biased upwards. At the conclusion of the study, all responding records will be reviewed to assess the extent to which a bias may be present.
 - All data will be analyzed for unusual values. Data from each operation will be compared to historical data (if available), as well as to trends from similar operations. Missing data for an operation will be estimated based on similar operations or historical data. Individual state and aggregated national estimates will be reviewed by NASS's Agricultural Statistics Board for reasonableness, then published (where sufficient data are available). If State-level data cannot be published due to confidentiality rules, the data will be published on either a regional or national level.
- 3. Describe methods to maximize response rates and to deal with issues of non-response. The accuracy and reliability of information collected must be shown to be adequate for intended uses. For collections based on sampling a special justification must be provided for any collection that will not yield "reliable" data that can be generalized to the universe studied.**

NASS will provide respondents with a variety of modes for completing the survey, including internet, mail, telephone, or personal interview. A customer service phone number is included at the top of the questionnaire in case respondents have any questions.

A cover letter will be mailed with questionnaires that will describe the importance of the data and how it will be used, as well as explain that individual data will be kept confidential. Instructions on how to access the internet questionnaire will also be provided.

Survey data are subject to non-sampling errors such as omissions and mistakes in reporting and in processing the data. Error is minimized by carefully reviewing all reported data for consistency and reasonableness.

4. Describe any tests of procedures or methods to be undertaken.

A maximum of nine cognitive interviews will be conducted with participants who are part of the target population in order to evaluate cognitive processes in understanding the survey questions and to assess ability to report the requested information accurately and reliably. Information from the cognitive interviews will be used to make non-substantive changes to the survey instruments, if necessary. In addition, the feedback will provide information that can be used in enumerator training. NASS will also do internal testing of the edit and summary programs to ensure accuracy before any publications are generated.

5. Provide the name and telephone number of individuals consulted on statistical aspects of the design and the name of the agency unit, contractor(s), or other person(s) who will actually collect and/or analyze the information for the agency.

Sample design was performed by NASS's Sampling, Editing, and Imputation Methodology Branch (Branch Chief is Mark Apodaca, 202-720-2857). Statistical summary programs are prepared by the Summary, Estimation, and Disclosure Methodology Branch (Branch Chief is Jeff Bailey, 202-720-4008).

Data collection will be carried out by NASS Regional Field Offices. The Director of Western Field Operations is Kevin Barnes (202-720-8220). The Director of Eastern Field Operations is Jay Johnson (202-720-3638).

The Survey Administrator responsible for coordination of questionnaire design, sampling, data collection, training, and Field Office support is Nate Vandermeer (202-720-0660).

January 2016

Revised March 2016