1Supporting Statement Part B

2022 CENSUS OF AGRICULTURE

OMB No. 0535-0226

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 for the universe as a whole and for each stratum in the proposed sample. Indicate expected response rates for the collection as a whole. If the collection, was conducted previously, include the actual response rate achieved during the last collection.

The target population for the Census of Agriculture is all farms. A farm is defined as any operation from which $1,000 or more of agricultural products were produced and sold or would normally be sold during the reference year (2022). This definition has been used since the 1974 Census of Agriculture (conducted by the Census Bureau). The National Agricultural Statistics Service (NASS) maintains a list sampling frame containing names and addresses of operations qualifying as farms under this definition. The list frame is continuously updated and supports the NASS agricultural estimates program as well as the Census of Agriculture Program. Processes are in place to identify and eliminate duplication and inactive operators (e.g., deceased, retired, and out-of-business farmers), and to evaluate outside list sources to find new and missing farm operators.

The list building effort leading up to the 2022 Census of Agriculture started the year following the previous Census. During the five-year time span between the two censuses, NASS performs intensive list building efforts to improve list coverage. List building is completed using two strategies: large general lists of potential farm operators available for all states and state-specific lists of targeted commodities.

The large general lists for all states that NASS processes each year include farm operators who utilize the Farm Service Agency, Risk Management Agency, and those that file federal tax forms related to agricultural production with the Internal Revenue Service. These lists are procured at the national level and produce a high success rate when

matched against the NASS list sampling frame.

The other list sources vary by state. Staff in field offices target list sources for specific crops or livestock commodities. The targeted commodities include major commodities as well as specialty commodities, particularly those which NASS considers to be underrepresented on the list frame. Examples of major commodity list sources include livestock and crop producer organizations along with lists from state and local government agencies. Smaller list sources procured in the past include lists of goat operators, beekeepers, and producers who may have maple syrup taps. These lists have varying degrees of success in identifying farm operators but are considered a necessity to building a representative list sampling frame.

The 2021 National Agricultural Classification Survey (NACS) will be used to screen approximately 1.35 million of these records to determine if they have any agricultural production and should be included on the census mail list. The list development effort is expected to produce a list of approximately 4.7 million names and addresses of farm operators and potential farm operators for the census of agriculture.

Response to the Census is required by law under the “Census of Agriculture Act of 1997,” Pub. L. No. 105-113 (7 U.S.C. 2204g). The overall response rate for the 2017 Census of Agriculture was 71.8 percent. Response rates will be monitored, and follow-up strategies will be used in order to achieve sufficient response rates at the state and county levels. Follow-up may use a 4-page questionnaire that collects the minimum amount of information needed in order to determine if the operation qualifies as a farm and to allow the data to be used collectively with the data collected on the full-length Census of Agriculture questionnaire.

The target population for the pilot data collection of agricultural decision makers’ sexual orientation and gender identification (SOGI), disability status, and expanded race data will be all confirmed farms on the NASS list frame. Because this list frame becomes the basis for the Census Mailing List, the target population should be similar to that for the Census of Agriculture. A statistical random sample will be drawn that will provide sufficient power to determine the efficacy of collecting these data from agricultural operations. The end product will be a report that will we used internally in the Department to inform program needs. The results may also be shared with external audiences with the aim of providing guidance on these topics.

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 goal of the estimation procedure for the census of agriculture is to produce agricultural totals for publication that are fully adjusted for list under-coverage, non-response, and misclassification at the county level. As was done in 2012 and 2017, a capture-recapture methodology will be used for the adjustments. The paper, The 2012 Census of Agriculture: A Capture-Recapture Analysis (*Journal of Agricultural, Biological, and Environmental Statistics* 22(4): 523-539), describes these procedures.

The primary assumptions in capture-recapture are (1) that the population is closed to the formation and deletion of farms between the two sampling periods, (2) that the probability of capture is the same for farms with a given set of characteristics, (3) that the probability of capture does not vary with sample, and (4) that being included in the June Area Survey (JAS) sample does not affect the probability of responding in the census relative to farms not included in the JAS survey, which is approved under OMB No. 0535-0213.

With the six-month time between data collection for the JAS and the census of agriculture, the population will have had some new farms formed and others that have gone out of business. Kendall (1999 *Ecology* 80: 2517-2525) showed that, if the farms enter and leave the population at random (violation of (1)), the estimates are still unbiased, but the precision is decreased. Further, if the capture probabilities vary with time (violation of assumption (3)), the estimates continue to be unbiased.

NASS assumes that being included in the JAS sample does not affect the probability of response on the census of agriculture. With only two samples, it is not possible to test this assumption. If responding to the JAS, increases the probability of response to the census of agriculture relative to farms that do not respond to the JAS, the estimate of the number of farms will be biased downwards; otherwise, if the probability of response to the census of agriculture decreases if a farm is in the JAS sample relative to other farms not in the sample, the estimate of the number of farms is biased upwards.

The capture probabilities do vary with type of farm (assumption (2)). To account for this variation in catchability in the models for coverage, response, and misclassification of Census Mailing List (CML) farms, stepwise weighted logistic regression with cross validation will be used for model development, with the selection made from the variables reported

on the census of agriculture. The key reporting variables (State, land in farm (in acres), operator's sex, operator's age, operator's race/ethnicity, type of farm, and total value of production) will be included in all models.

To implement this method these two surveys (JAS and Census) are used. It is assumed that the surveys are independent and that farms with a given set of characteristics are equally likely to be captured on both surveys. NASS uses the JAS to estimate the number and types of farms not on the CML. The tracts in the JAS that are not on the CML are said to be in the "Not on the Mail List" (NML) domain. If a tract in the NML domain is determined to be a farm during the census, it is an NML farm. The CML is used with the NML in the capture-recapture framework to represent all farming operations across all states in the JAS sample. Although much effort is expended to make the CML as complete as possible, inevitably the CML will not include all U.S. farms, resulting in list under-coverage. Some farm operators who are on the CML will not respond to the Census, despite numerous attempts to contact them. In addition, although each operation is classified as a farm or nonfarm based on their responses on the Census questionnaire, some will be misclassified (some non-farms will be classified as farms and some farms will be classified as non-farms).

Probability estimates for missing a farm due to nonresponse, misclassification, and under-coverage are obtained via logistic regression models using matched records from the CML and JAS. These probability estimates are used in the capture-recapture model to adjust weights. The weights for these estimates are general purpose in that they do not provide any control over expected levels of commodity production. To address this a calibration algorithm is used to create the final set of weights. The calibration algorithm adjusts the weights to align indications with previously published estimates and to reduce bias. (See the attached *Calibration for the Census of Agriculture*.) The calibration algorithm outputs integer weights to ensure that totals for all sub-domains and cross-tabulations will balance in the final data products.

The NASS area frame, which is used for the independent June Area Survey (JAS), covers all land in the U.S (excluding Alaska and Hawaii) and includes all farms. The land in the U.S. is stratified by characteristics of the land. A probability sample of segments is drawn within each stratum for the JAS. The JAS sample of segments is allocated to strata to provide accurate measures of acres planted to widely grown crops, farm numbers, and inventories of cattle. In June 2022, the operational sample will be increased to improve the farm counts for operations that produce specialty commodities or have socially disadvantaged or minority operators. The supplemental sample will also enhance the capture-recapture modeling.

Prior to the beginning of data collection, input will be solicited from field office staff with the specific objective to identify a small portion of records that need to be specially handled (tagged records). These records may be matches to other surveys, have multiple locations or unique operator structures, have existing data collection plans, or need to have data collected by statisticians or supervisory enumerators, many are typically large or complex operations which we have existing relationships and contact frequently.

In an effort to reduce response burden NASS coordinates data collection activities with concurrent surveys. A majority of these census of agriculture questionnaires are matched with records in the Agricultural Resource Management Survey Phase 3 (ARMS 3), (OMB approval number 0535-0218) sample. The ARMS 3 survey collects whole farm data and has many similarities to the census of agriculture questionnaire. NASS and the Economic Research Service (ERS) have collaborated on an 'integrated' questionnaire that incorporates all the necessary agriculture census data items into the ARMS 3 questionnaire. Effectively, the ARMS 3 data collection will run concurrent with most of the census of agriculture data collection period, with operations in the ARMS 3 sample targeted for face-to-face enumeration. Upon conclusion of ARMS 3 data collection, the necessary data captured in ARMS 3 will be used for the census of agriculture.

Some other tagged records will be of farms and ranches that have multiple locations and can be reported by one headquarters operation.

In the population census, which is conducted by the Census Bureau, households are treated equally, by this we mean that there are no adjustments for size for a larger home vs. a smaller home. However, with the census of agriculture we are dealing with farm or ranch operations. A small number of farmers could represent a very large portion of production of a certain commodity. These records may be tagged so that NASS staff can be sure that theses questionnaires are completed which may involve personal enumeration.

Finally, there are a very minimal number of records that we have a long history with and have an established enumerator contact. To honor an agreement, NASS staff tags these records for collection by that enumerator.

The consequences of normal handling of these records could manifest itself into increased response burden due to the potential need for follow-up contacts to verify data.

During 2019, 2020, and 2021 NASS staff worked with minority population groups to obtain lists of potential agriculture producers to receive a National Agricultural Classification Survey questionnaire (OMB No. 0535-0140). This effort will continue into 2022. During the collection period if an operation is discovered within one of the under-represented populations, NASS will include their report in the aggregate data.

NASS intends to continue efforts to obtain census of agriculture data from individual American Indian farmers and ranchers on reservations in the 2022 census. Historically, the Census Bureau and NASS treated most American Indian reservations in the U.S. as single farming operations for the census of agriculture. A single report was obtained for the entire reservation, including data for any tribally operated farm or ranch and all individual farms and ranches. In 2007, NASS expanded its efforts to reach individual American Indian farms and ranches, on and off reservations. For the majority, individual operators were represented in the agriculture census data in all states. In a few instances, operator counts were obtained from reservations which preferred to report aggregated reservation data. Similar procedures will be used in 2022. NASS will also use a customized questionnaire for American Indian operations in the southwestern United States.

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.

In an effort to reach typically under-covered populations, NASS has and will continue to work with community-based organizations (CBOs) to recruit and train members of the under-covered populations as enumerators. NASS also plans to send staff to attend events held by CBOs leading up to and throughout the 2022 Census of Agriculture data collection period. These staff will provide assistance in filling out census of agriculture questionnaires.

Farms that potentially represent a large proportion of the total for a given commodity will be identified prior to data collection and categorized as "must case" records. These will be the top 0.1% to 1.5% depending on the number of farms with the commodity. Non-response from these operations often cannot be accurately adjusted for, so special efforts will be made to encourage response. This may include sending a pre-survey letter to confirm the operation's address or having a NASS representative visit the operation prior to mailing out the 2022 Census of Agriculture

questionnaires. During the data collection phase, these “must cases” are tracked and will be among the first to receive phone follow-up or personal enumeration.

NASS places a high priority on obtaining comprehensive and uniform coverage of all farms. In order to ensure sufficient coverage in every county, NASS will utilize a Computer Assisted Telephone Interview (CATI) program specifically designed to target records in counties with lower coverage rates. Personal enumeration may also be used to target these counties.

Item non-response will be handled in one of four ways. First, deterministic imputation will be employed whenever the missing value can be derived from other cells on the questionnaire. Second, previously reported data from either a recent NASS survey or the previous census of agriculture, will be imputed, when appropriate. Third, a nearest neighbor donor (farm of similar type, size, and location) will be found and a value or relationship from the donor will be used for the recipient. When all of these automated options fail, the problem will be referred to a statistician for resolution. The statistician will utilize knowledge from training, existing agricultural knowledge, computer program analysis tools, and on some occasions will recontact a respondent. Unit non-response is handled through weighting, as described in item 2, above.

Relative standard errors for U.S. fully adjusted estimates of number of farms for major demographic items in 2017 ranged from 1.7 to 20.7 (see Table B in Appendix A of the 2017 Census of Agriculture United States Summary and State Data, Volume 1, Part 51). In 2022, NASS will survey a comparable number of JAS segments for census of agriculture adjustments. Thus, the relative standard errors are expected to be similar to 2017.

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

NASS used 2017 Census of Agriculture data to identify problem areas of the questionnaire by looking at edit and imputation rates by question and section. Comments and suggestions from field staff were also reviewed for potential problems with data collection and questionnaires.

NASS conducted a total of 50 cognitive interviews (OMB No. 0535-0248) for the census of agriculture questionnaires, focusing on new content as well as alternate ways to collect commodity data. In addition, NASS conducted a content test in 2020 (OMB No. 0535-0243) with a nationwide sample of approximately 30,000 operations. The content test incorporated new content (from the 2017 Census of Agriculture), some questionnaire

format changes, and several embedded experiments. These experiments focused on testing alternative data collection strategies and on providing one test group with some previously reported data. These experiments were intended to test ways of increasing response rates and lowering respondent burden.

A second, internet only census of agriculture test (OMB No. 0535-0243) will be conducted in 2022 for the 2021 production year. This test will further develop and review NASS's ability to provide respondents with their previously reported data.

Content test results were used to determine the final questionnaire content and design.

Processing systems have been updated based on experiences from the 2017 Census of Agriculture. All processing systems will be tested using previous agriculture census data, as well as data from the 2020 content test.

For the pilot study, of agricultural decision makers' sexual orientation and gender identification (SOGI), disability status, and expanded race data, NASS will divide the pilot study sample into two groups: a treatment group and a control group. For the treatment group, the questionnaire will include questions concerning agricultural decision makers' sexual orientation and gender identification (SOGI) and disability status, an expanded list of race options, and the questions on the traditional NACS questionnaire. The control group will receive the traditional NACS questionnaire, which does not include the SOGI, disability, and expanded race questions. Comparisons between the control group and the treatment groups will identify possible challenges with collecting SOGI, disability, and expanded race data. These challenges could include (but are not be limited to) differences in unit or item non-response and differences in the percent of recipients who contact the provided help telephone line.

As of the date of this submission, NASS does not have the questionnaire that will be used for the pilot data collection that will collect data on agricultural decision makers' sexual orientation and gender identification (SOGI), disability status, and expanded race data. To develop the questionnaire, NASS will consult with other Federal Statistical Agencies and other subject matter experts. Potential questions will be cognitively tested by NASS staff, with potential collaboration from other Federal Statistical Agencies. NASS will submit a change request to OMB that provides the questionnaire when it is complete.

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

Several NASS units contribute to developing census of agriculture methodology, each containing staff members with prior agriculture census experience. Contributing senior staff and unit leaders are:

- Mark Apodaca, Chief, Sampling, Editing, and Imputation Methodology Branch: (202) 720-2857),
- Jeff Bailey, Chief, Summary, Estimation, and Disclosure Methodology Branch: (202) 720-6468,
- Dan Beckler, Chief, Standards and Survey Methodology Branch: (202) 720-8858,
- Donald Buysse, Chief, Census Planning Branch: (202) 690-8747,
- Data collection is carried out by NASS Field Offices; Eastern Field Operation's Director is Jody McDaniel (202) 720-3638 and the Western Field Operation's Director is Troy Joshua, (202) 720-8220,
- Tony Dorn, Chief, Environmental, Economics, and Demographics Branch: (202)720-6146,
- Suzette Qualey, Deputy Director, National Operations Division: (314) 595-9502,
- Kathy Ott (202) 720-1114 and Heather Ridolfo (202) 692-0293 Survey Methodologists,
- Linda Young, Director, Research and Development Division: (202) 690-1401.

July 2021