

# **Obstacles in Planning Establishment Survey Experiments - Census of Agriculture Content Test and Agricultural Resource Management Survey**

**Tyler Wilson and Kathy Ott**

U.S. Department of Agriculture, National Agricultural Statistics Service 1400 Independence Ave. S.W.

Proceedings of the 2018 Federal Committee on Statistical Methodology (FCSM) Research and Policy Conference

## **Abstract**

Experiments in establishment surveys require consideration of factors unseen in household surveys. For example, consideration of the burden on establishments, tailored data collection, wide variety in size of sampled establishments, availability of unit level control data, and coordination across survey programs. In agricultural surveys, sample units can range from small farms to multimillion dollar operations. The largest operations may be sampled repeatedly over time and across survey programs.

This paper will discuss how these issues impacted the planning, operations, and results for two experiments, one stand-alone experiment for the 2017 Census of Agriculture (COA) Content Test and another embedded in a large-scale production survey referred to as the Agricultural Resource Management Survey (ARMS).

Alternatives to data collection procedures for ARMS intended to increase response rates were tested in several split sample experiments. Because ARMS collects complex detailed information, the National Agricultural Statistics Service (NASS) did not want to impose this burden in a stand-alone experiment. However, planning involved concessions such as removing large sample units or those receiving special handling in data collection and using the large amount of auxiliary data to selectively apply procedures.

Because the COA is conducted only every five years, experiments designed to improve response rates or data quality were stand-alone experiments. Prior to the 2017 COA, several split sample experiments were conducted to evaluate alternative questionnaires and data collection procedures. Considerations unique to establishments were also made in planning these experiments such as removing large, complex or unique units or those sampled in other NASS surveys in the months surrounding the experiment due to concerns about their level of burden. In addition, NASS was also able to use auxiliary data to tailor the sample.

The results of the experiments, the related tradeoffs made in planning experiments, and the implications will be discussed.

## **Introduction**

Experiments in establishment surveys require consideration of factors unseen in household surveys. For example, consideration of the burden on establishments, tailored data collection, wide variety in size of sampled establishments, availability of unit level control data, and coordination across survey programs. In agricultural surveys, sample units can range from small farms to multimillion dollar operations. The largest operations may be sampled repeatedly over time and across survey programs.

Two experiments conducted by NASS highlight several of the obstacles and considerations unique to establishment surveys. One of these experiments is a stand-alone experiment for the 2017 Census of Agriculture Content Test and one is embedded in a production survey (the Agricultural Resource Management Survey).

Agencies that conduct household surveys in the United States have information about geography and characteristics of communities that help to identify households to select for surveys. Households are rarely in more than one survey over the course of several years because of the large universe of households available for sampling.

Agencies, like NASS, that conduct establishment surveys in the U.S. have to maintain more than just geographic data in order to create the universe and sampling for surveys in order to target the correct types and sizes of establishments. NASS maintains an area frame as well as a detailed list frame of all known and potential farming and ranching operations in the United States. The list frame contains hundreds of geographic, demographic, and control data items for operations and is used to create the universes and samples for different agricultural surveys conducted by NASS. This list frame is crucial to the universe creation and sampling used for NASS surveys. The experiments in this paper only include NASS list frame records.

## **Census of Agriculture (COA) Content Test**

### *Background*

The COA is conducted every five years by NASS and collected production, practices, demographics, and economic data for all farm and ranch operations in the United States. Because the COA is conducted only every five years, experiments designed to improve response rates and data quality for 2017 were conducted as stand-alone experiments. Prior to the 2017 COA, the COA Content Test was conducted to evaluate alternative questionnaires and data collection procedures.

### *Methods*

The main purpose of the COA Content Test was to test the question content and format of several versions of the COA paper questionnaire. No weighting or imputation was done to account for missing records or data, and no estimates were produced for any commodities or other variables on the questionnaire.

In addition to testing all questions and sections on the questionnaire, there were also questionnaire issues that were tested during the COA Content Test, resulting in six questionnaire versions. The issues were the placement of demographic information in the form, modified land and acreage sections, the use of a code listing in the questionnaire vs. an instruction booklet, and the use of preprinted items in data tables. In addition, we wanted to test the use of a presurvey contact card. Specific results from these different versions are discussed in McCarthy et al (2017), but are not discussed in detail in this paper.

There were two universes created for the COA Content Test to mimic the two universes planned for the 2017 COA, one for the long form and one for the shorter form. The long form asked about all commodities and was 24 pages long. The short form was designed to address a common complaint from operators that questionnaires contain information that does not pertain to them. The short form was 20 pages and targeted operations that only have certain commodities (cattle, horses, hay, and a limited number of field crops) and did not ask about other commodities. The operations in each universe were targeted by using the list frame control data for each operation.

A sample of 30,000 records was used for the COA Content Test, given our budget. The six forms, two universes, and the contact card data collection strategy were combined to create ten treatment groups. Detailed information about the questionnaire versions and treatment groups can be found in Ott et al (2017), but is not discussed in detail in this paper.

### *Obstacles/Considerations Related to Establishment Surveys*

The methodological components of the COA Content Test with the biggest consideration unique to establishments were in the universe composition and sampling strategies. As stated earlier, NASS maintains a list frame that contains control data for farm and ranch operations in the United States. This allows NASS to target certain types and sizes of farms and ranches for surveys and experiments.

For the COA Content Test, the major universe composition issue was to divide the entire universe of farm operations into a long form universe and a short form universe. Before this division, several operations were removed from consideration. The first group of operations removed was large, complex or unique farms and ranches. These operations are often critical to one or more agricultural production estimates that NASS releases each year and therefore, are often selected with certainty for commodity specific surveys, or are in several surveys

throughout the year. It is not unusual for NASS to use some sort of customized data collection strategy to collect data for large or complex operations. These operations have a large impact on the commodity specific surveys, so NASS decided that in order to reduce respondent burden, they would be removed from the universe of eligible operations for the COA Content Test. While potentially advantageous from the operation's perspective, this is not ideal for the conclusions that can be drawn from the Content Test results. NASS decided however, that participation in the commodity survey programs throughout the year and across years was more important than participation in the COA Content Test which might jeopardize a respondents' future participation. No survey estimates were generated from the COA Content Test, but any conclusions drawn from the results about the questionnaire have the caveat that none of the largest operations were included.

The second group of operations removed from the universes were those operations that were sampled many times throughout the year, or that were in other NASS surveys in the months surrounding the experiment, namely December 2015 through March 2016. The reason for this decision was concern about these operations' level of burden during the year and during the data collection period. Similar to removing the large operations, while this decision is advantageous from the operation's perspective, it does not allow us to draw conclusions about them when analyzing results. In addition, this decision caused certain types of operations to be almost completely removed from the sample. For example, any operation that was selected for a monthly survey (such as *Cattle on Feed*) would have met the number of surveys threshold to be removed from our universe. Therefore, cattle on feed operations were likely to be removed.

The COA Content Test tested two different forms, a long form and a short form. The populations for those two forms was different. The short form targeted operations that had cattle, horses, hay, and a limited number of field crops, but no other commodities, according to the list frame. In order to divide the overall universe, control data on the types of commodities reported in the past for each operation were used. NASS's list frame contains numerous control data items about the types and amount of crop and livestock commodities an operation has, as well as some production practices used by the operation. Operations that only had control data for cattle, hay, horses, and a small number of field crops were eligible for the short form. Dividing the sample allowed NASS to test both versions of the questionnaire with the population of interest for that questionnaire. Without the use of historical data from the list frame, the option of creating a targeted form would not have been a consideration.

NASS was also able to use auxiliary data to tailor the sample for the needs of the COA Content Test, particularly for the types of operations selected. Because the COA includes all types of operations, we wanted to ensure that the COA Content Test sample met a minimum threshold for demographic items, specialty commodities such as bees, cotton, and aquaculture, and practices that are somewhat rare, such as production contracts. This allowed us to have potential operations in sample that could fill out every section of the questionnaire. This is a huge benefit of being able to use list frame data for this survey. If we don't see enough potential respondents in the sample for a particular item, we could modify the sample.

### *Results*

The numbers on NASS's list frame are considered administratively confidential, however, rounded percentages are provided to show how using the list frame potentially impacted the COA Content Test findings and results.

As described above, NASS removed large operations, operations that had been contacted many times throughout the year, and those with a recent survey contact. This removed approximately 20% of the population from the overall farm operation frame.

The remaining records were then divided by whether they were eligible for the shorter form or not, based on the types of commodities that the operation grew or raised in the past. Operations that only had control data for cattle, hay, horses, and a small number of field crops were eligible for the short form. After applying these criteria, approximately two-thirds of the remaining records were eligible for the short form, and one-third had to receive the long form in order to collect data about all of their agricultural production.

In order to test all sections of the form, we set a minimum threshold for many demographic items and specialty commodities. After the sample was drawn for each form, NASS looked at the frame data for these variables to see if we hit our thresholds or not, to determine if the sample should be modified. We didn't quite hit the targets for

aquaculture, cotton, or operators of Hispanic origin, but all of these were within 10% of the threshold. However, we did not come close to our target for some types of production contracts, such as cattle on feed and “other crops.” As discussed above, the low number of cattle on feed contract operations was likely because of the elimination of operations that were in sample for many surveys during the year. One of these surveys is cattle on feed. Given this, we did not modify the sample to put more cattle on feed operations in sample. We did modify the sample for “other crops” production contracts. Upon looking at the sample selected for operations with these contracts, we realized that we had very few in sample. Before fielding the survey, we included more of these operations in sample to ensure that the production contracts section of the questionnaire could be adequately tested.

### *Implications*

Modifying the universe and sample selection for this experiment had both positive and negative implications. While reducing the burden placed on operations by eliminating them from sample is potentially positive for the operation, it does not allow us to provide analysis, findings, and results from some types or sizes of operation in our experiment. Any conclusions drawn from the results, or decisions made using the results, must be made cautiously and consider what they represent and do not represent.

Using control data to ensure representation of specialty commodities and relatively rare types of operations can be beneficial for experiments such as this to ensure testing of all parts of the questionnaire and to include all types of crops and livestock of interest. Knowing this information ahead of time allowed us to modify the sample to include operations with “other crops” production contracts, which would otherwise have not been included.

## **Agricultural Resource Management Survey (ARMS)**

### *Background*

The Agricultural Resource Management Survey (ARMS) is administered annually by NASS in partnership with the Economic Research Service (ERS). It is conducted in three phases over the course of the year. The first phase (ARMS I) is a screening phase that provides a sample of in-business operations that have commodities that have been targeted for that given year (targeted commodities change annually). The second phase (ARMS II) focuses on production expenses, chemical use, and the targeted commodity. The third phase (ARMS III) focuses on financial data such as expenses and income as well as risk management practices. In recent years, response rates for ARMS III have dropped below 80% requiring a coordinated effort to target and test strategies that can increase response rates. However, due to the three phases, the ARMS survey is essentially in motion all year. This adds to the unique set of considerations and obstacles when performing experimental design tests within an operational environment for an establishment survey.

### *Methods*

During the initial years of research (2011-2013), nonresponse propensity scores were developed for ARMS and used to target farm operations predicted as highly unlikely to respond (Mitchell, M., Ott, K., & McCarthy, J. 2015; Ridolfo, H., & McCarthy, J. 2015). Experimental designs were used to test the effect of using high-status personnel, such as Deputy Directors or State Statisticians, to solicit survey cooperation. Some increases in response rates were found using these methods but statistical significance was difficult to obtain due to small sample sizes.

In 2014, research methods evolved to increase sample sizes and target operations identified as *unlikely* to respond, as opposed to *highly unlikely* to respond. Half of these sample units received an increased amount of personal interaction and incentives via the “drop and collect” method of data collection (Ibeh, Brock, and Zhou 2004; Allred and Ross-Davis 2011). Targeting this midrange group of unlikely responders not only provided a larger sample of operations to test but were, in theory, more likely to cooperate when given alternative data collection procedures.

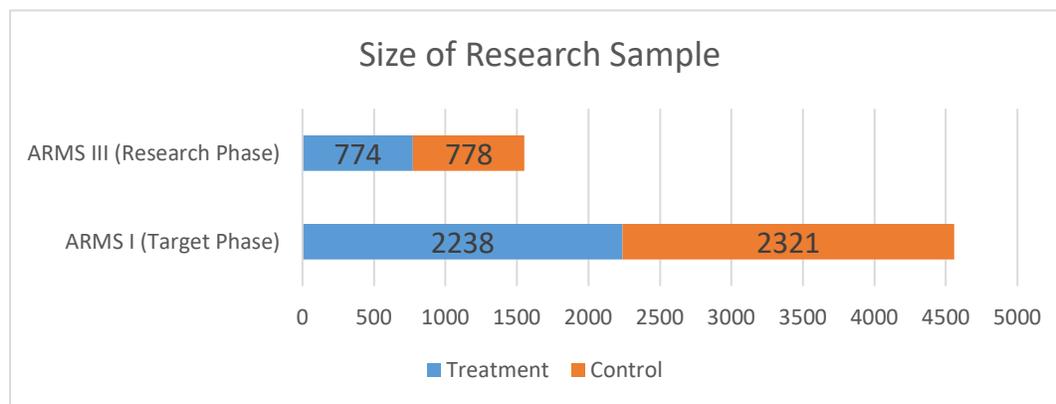
The next section briefly outlines a few of the considerations and obstacles that were encountered during these tests in ARMS.

### Obstacles and Considerations for Establishment Surveys

For obvious reasons, modeling nonresponse and assigning propensity scores in both household and establishment surveys is more determinate on what information can be obtained for nonrespondents than those who respond. From our experience in establishment surveys, the characteristics of the operation, and not the respondent or their location, are what makes effective propensity scores. As was done at NASS, this information is often found within control data or other establishment surveys such as the COA. Unfortunately finding this information can be the easy part; whereas linking it all together can be a test of patience.

Due to the longitudinal nature of ARMS I, II, and III, assigning propensity scores to records and selecting records to test must be done almost a year in advance. Because of this, the exact size of the treatment and control group cannot be determined ahead of time. The graph below illustrates this.

Figure 1: Size of Samples



Without this knowledge, checklist items such as budgeting extra resources for treatment sample, acquiring OMB approval, and training field enumerators, became an arduous task of estimation and best guesses.

Farms and ranches are asked by NASS for similar information, multiple times a year. At NASS, this often leads to coordination of data collection efforts. It is not uncommon for surveyors to carry several different surveys when traveling to rural areas to collect survey responses. This operationally efficient practice is difficult to predict and even more difficult to assign alternative data collection to specific records that may or may not be coordinated with other surveys. In addition, this difficulty brings up questions of overburdening an establishment.

In 2014, several field offices had to hold a number of test cases out of the survey (and thus, the experiments) because operations were receiving other surveys running at the time and didn't want to risk overburdening the respondent. One field office stated, *"Our cases were held because they were all marked as extreme operators (EOs) for the March Agriculture Survey. We didn't want to mess up the research sample so we held them out and marked alternative methods."* These cases, as well as others, contributed to 15% of our initial test sample not receiving the alternative data collection methods being tested. Because of the large sample size (700+) in recent years, this was a surmountable obstacle, however, if this had happened in prior years where sample sizes were 30-40 respondents, it would have been much harder.

Other than the Centennial Census, household surveys are generally voluntary. Whereas discerning whether an establishment survey is voluntary or mandatory can be challenging for a respondent. Case in point, every year after the COA, some respondents received an ARMS that was considered mandatory because it was a follow-on to the COA, whereas others received the voluntary version of ARMS. This occurred during one of the alternative data collection tests in 2013 and response rates increased significantly across the board. These increases made it difficult

to assess the effect of alternative data collection procedures because the baseline (control) group's response rate was so high.

#### *Results:*

Although there were many obstacles, the results of these studies made it clear that unlikely responders to ARMS are not easily influenced by alternative data collection procedures. However, without multiple years of testing, especially within an operational environment, this result could not be fully realized. These studies, conducted with an establishment survey within an operational environment, would have been much less impactful if only conducted one time over the course of one year. Because multiple tests were done over the course of several years, the research adapted to the limitations found in the past and the circumstances of the present.

#### **Summary**

There are several obstacles and implications in planning for experiments in establishment surveys, in both stand-alone experiments and those that are embedded in a production survey. In the two studies discussed here, the implications and obstacles were related to universe creation, sample selection, finding available establishment data on nonrespondents, and data collection coordination and considerations. Although household survey organizations consider burden within a survey, they do not generally consider burden across surveys, something establishment surveys often do. Along with this is the issue of individualizing tailored data collection or coordination of survey programs, particularly for large or complex establishments.

#### **References**

- Allred, S. B., & Ross-Davis, A. (2011). The drop-off and pick-up method: An approach to reduce nonresponse bias in natural resource surveys. *Small-Scale Forestry*, 10(3), 305-318.
- Ibeh, K., Brock, J. K. U., & Zhou, Y. J. (2004). The drop and collect survey among industrial populations: theory and empirical evidence. *Industrial Marketing Management*, 33(2), 155-165.
- Mitchell, M., Ott, K., & McCarthy, J. (2015). Targeted Data Collection Efforts for the 2012 ARMS III. Research and Development Division Research Report Number RDD-15-01, US Department of Agriculture, National Agricultural Statistics Service. Mitchell, M., Ott, K.,
- McCarthy, J., Ott, K., Ridolfo, H., McGovern, P., and Moore, D. (2017). "Multi-method questionnaire testing in establishment surveys." Proceedings of the Fifth International Conference on Establishment Surveys, Geneva Switzerland, June 20-23, 2017.
- Ott, K., McGovern, P., and Sirkis, R. (2017). "Using Analysis of Field Test Results to Evaluate Questionnaire Performance." Proceedings of the Fifth International Conference on Establishment Surveys, Geneva Switzerland, June 20-23, 2017.
- Ridolfo, H., & McCarthy, J. (2015). Targeted Data Collection Efforts for the 2013 ARMS III. Research and Development Division Research Report Number RDD-15-01, US Department of Agriculture, National Agricultural Statistics Service.