Adaptive Design Strategies:
Who do you target? And what do you do with them?

Jaki S. McCarthy, Kathleen Ott, and Melissa Mitchell
USDA/National Agricultural Statistics Service, Fairfax, VA

Proceedings of the 2013 Federal Committee on Statistical Methodology (FCSM) Research Conference

BACKGROUND

Different data collection strategies have always been used in survey data collection. However, these strategies are often unplanned or unknown to the survey designers and administrators. For example, interviewers given a caseload of sample units may prioritize them according to their own priorities, perhaps visiting those nearby first or last, or attempting contact with units perceived to be easier to enumerate before more difficult cases. Similarly, call centers conducting telephone interviews may schedule interviewers according to their own priorities, such as when interviewers are more likely to be available, or to provide steady work for interviewer staff.

Adaptive Design in survey data collection seeks to strategically allocate resources in data collection to improve data quality, reduce costs or reduce the Total Survey Error, not just serve goals important to the immediate survey production phase. The USDA's National Agricultural Statistics Service (NASS) has been trying to incorporate adaptive design strategies into several of its surveys. This paper will discuss tests of adaptive design in two NASS surveys, the Quarterly Crops/Stocks Survey (CS) and the third phase of the Agricultural Resource Management Survey (ARMS).

The CS collects crop acreage, inventory and productions of grain and oilseed crops. It is conducted in June, September, December and March. Estimates are produced at the national and state level (for major commodities) and additional specialty crop estimates differ by state. The samples sizes are typically large, in excess of 65,000, and the data collection period is short, less than 2 weeks. Due to the limited data collection period, most data collection is done by computer assisted telephone interviewing (CATI) in centralized phone centers. However, there is limited mail, online, and in person interviewing also conducted.

The ARMS collects production practices and cost of production data on selected commodities. The ARMS also collects detailed whole farm financial information from a representative sample of farms and ranches across the country. ARMS is collected in three data collection phases: The initial phase, ARMS Screening survey, collects general farm data such as crops grown, livestock inventory, and value of sales. Screening data are used to qualify (or screen) farms for the other phases. The second phase, (Phase II), collects data associated with agricultural production practices, resource use, and variable costs of production for specific commodities. The final phase, (Phase III) collects whole farm finance, operator characteristics, and farm household information and this final phase is the subject of this project.

In ARMS Phase III, farm operators provide data on farm operating expenditures, capital improvements, assets, and debt for agricultural production. In addition, operators provide data on farm-related income, government payments, the source and amount of off-farm income, and characteristics of themselves and their household. Estimates are produced at the National, regional and state level (for the 15 largest producing states). ARMS Phase III is a lengthy survey with average interview length just under 90 minutes and many interviews lasting much longer. Typically, an interviewer will call selected operators and set up an appointment for the in person interview. The data collection period is also much longer than for CS, lasting several months.

In addition to considering the data collection protocols for a survey it is also important to understand how non-response is handled in the survey. In the CS, the sample is selected using a multivariate probability proportionate to size design. However, nonresponse adjustments are made using traditional strata, based on the size of the farm operation in acres and the total grain storage capacity. In addition, in some states, additional strata are included for specialty crops such as potatoes, tobacco, etc.

In contrast, the ARMS nonresponse adjustments are handled by the use of calibration weighting. Calibration adjusts for nonresponse bias and measurement error by adjusting weights so that the reporting records' expanded data meet external calibration targets. A number of calibration targets are defined based on known external targets such as official estimates derived from other sources.

Another important consideration in establishment surveys is that certain sample units may have a large impact on population totals. Often, establishment survey populations may be highly skewed, with a small number of sample units dominating an estimate. In other cases, there may be a small number of operations that are critical to nonresponse weighting. For example, those in small nonresponse strata, producing specialty commodities or calibration targets for small subpopulations.

Given the characteristics of a particular survey, appropriate approaches to managing data collection may differ. For the CS, operations in the largest nonresponse strata, which is defined by size, and those in specialty commodity strata can be termed "impact operations". These are the ones likely to have the most influence on the survey's population estimates.

For the ARMS survey, which uses calibration weighting, the calibration targets are known in advance. These include items such as the total number of farms, corn, soybean, wheat, cotton, fruit and vegetable acres as well as cattle, milk production, hogs, broilers, eggs and turkeys. In order to use a particular variable in calibration you must have a minimum amount of the calibration target reported by respondents. That is, calibration weights are capped so that individual weights are not unrealistically large. For ARMS, we can then define impact operations as those that are large relative to the calibration targets. For example, large dairy operations who will contribute most to a milk production target can be defined as "impact operations." In addition, NASS has also developed non-response propensity models for the ARMS to identify those operations highly likely to be nonrespondents (see Earp, Mitchell, McCarthy, & Krueter (under review) and Mitchell & Ott (in preparation) for a more detailed discussion of those models). Those highly likely nonrespondents who are also impact operations can be targeted for additional attention in data collection.

NASS has also developed nonresponse propensity models for CS (McCarthy, Jacob, and McCracken, 2008; Mitchell and McCarthy, 2012). Thus for both the CS and ARMS, we can identify important operations and their a priori propensity to respond. But once operations have been targeted for additional attention, what strategies can be used in data collection to increase response?

ADAPTIVE DESIGN STRATEGIES IN DATA COLLECTION

For the CS, we used some of the nonresponse adjustment strata to define impact operations. Those in large strata (i.e. those likely to contribute more to population estimates) were defined as impact operations, for example, operations in strata with 5000 or more acres. In addition, we also included any specialty strata a state had (e.g. potatoes, tobacco, etc.) without regard to size, since these populations are typically small. We decided to restrict additional resource allocation to operations that were both "impact" operations and highly likely to be nonrespondents. In order to test the effectiveness of these procedures, selected NASS field offices were asked to participate in a test of systematic adaptive design procedures.

In the CS survey, the majority of data collection is CATI conducted by centralized call centers. Our adaptive design strategy for the identified operations was to assign responsibility for data collection to the local field office. It was felt that local interviewers would use local phone numbers to make the calls, would have more flexibility in timing calls, and might have more local knowledge about the sample units. Cases were assigned to local field interviewers, although in order to contain costs, we directed the field offices to have their field interviewers attempt contact by telephone before going in person to collect the data if necessary. Obviously, in person interviews will increase costs over telephone interviews. Operations that were likely to respond were assigned to the central call centers (even if field interviewers had previously collected their data), although at the end of the data collection period, they were to be reassigned to the field office for a last ditch effort to contact them with a local interviewer. In this test, several of our field offices volunteered to test these procedures. We compared the expected and actual response rates for these field offices to the expected and actual response rates for other field offices to evaluate the application of these adaptive design strategies.

In contrast, ARMS already uses primarily in person interviews, well known to produce higher response rates than mail or telephone interviews. So switching to a more effective mode is not an option in this case. However, the interviewers will often contact sample units by telephone to recruit them and set a time for an in person interview. For our ARMS adaptive design strategy we moved impact records who were also identified to be likely nonrespondents from regular field enumerators, and provided instructions for these to be contacted in person initially (instead of telephoned for appointments). We also directed the local field office director, or more senior or experienced staff to make the initial contact and recruit the respondent. The actual data collection interview was to be conducted only by experienced or supervisory interviewers, either in the initial contact or in a scheduled appointment. In addition, we also provided funds and instructed field offices to offer incentives to interviewers for completed responses from these targeted operations. The targeted records were marked and any interviewer who got a good completion would receive an additional $20 for each.

RESULTS OF INCORPORATING ADAPTIVE DESIGN STRATEGIES

For the CS we cannot directly compare response rates between field offices as we know that field offices do not historically have comparable response rates. Instead we used our nonresponse propensity models to compare each field office's expected response rate to its actual response rate. Our models for CS do reliably distinguish between respondents and nonrespondents (for both likely refusals and likely noncontacts) as shown by the response rates shown in table one.

The first column shows the response rate for the group of operations identified as being most likely to be either a refusal or noncontact. We can see that our models did identify the group that had a much lower response rate than other operations. The second column shows the response rates for the other nonresponse propensity groups (3 groups for refusals and 4 for noncontacts). The third and fourth columns show the difference between the actual and the predicted response rates, with higher numbers showing a more positive result. Overall, results were disappointing. In both September and December 2013, the states who were to participate in our experiment (treatment states) appeared to have slight gains over their expected response rates for the likely refusals, unlike the comparison states. But results were a bit more ambivalent for the expected noncontacts, with better response rates gains in the comparison states.

Table One: Expected and Actual Response Rates for CS

Crops/Stocks			**Difference (Actual – Predicted) %**	
Refusals	Likely REFUSAL Response Rate	Other Response Rates	Treatment States	Comparison States
Sept All	21%	52-78%		
Dec All	15%	46-73%		
Sept Impact Operations	18%	53-79%	2.8	-4.6
Dec Impact Operations	12%	47-74%	0.2	-12.0
Inaccessible	Likely NONCONTACT			
Sept	39%	34-75%		
Dec	30%	29-71%		
Sept Impact Operations	38%	32-74%	9.4	14.7
Dec Impact Operations	29%	26-70%	-5.1	5.5

We contacted our field staff again once the study was complete to gain more insight into how these procedures worked (or didn't) from their perspective. Overall, sending CS cases from the central data collection centers to field offices didn't appear to produce much improvement in response rates. Separately we have conducted research comparing response rates from cases called from local telephone numbers to those called from our call center in another state. This research has demonstrated that calls from a local telephone exchange do not increase telephone survey response rates (Ridolfo, Boone, Dickey, 2013) so this result is not surprising.

Cases identified as being likely nonrespondents persisted in being hard to collect data from. Follow up with field staff revealed problems in applying the adaptive design for CS. It was difficult to know whether our experimental procedures were at all effective, since our comparison group was not a true control group. In this case, we were comparing states that had agreed to follow procedures with a group of states that decided on their own individual data collection plans.

Results were also disappointing for the ARMS. Our models did identify operations more likely to be nonrespondents, but there was no difference in response rates in our experimental group with both groups having about a 55% response rate, as shown in table two.

Table Two: Expected and Actual Response Rates for ARMS

ARMS	Likely Nonrespondents	Others	Targeted Records	Control Records
Complete	55.6%	73.1%	55.3%	55.2%
Refusal	36.9%	21.8%	36.2%	37.5%
Noncontact	4.9%	4.2%	4.8%	5.2%
Office Hold	2.9%	0.9%	3.7%	2.1%

To our surprise, we discovered that the field staff had only minimally adhered to our instructions for handling ARMS. Several field offices commented that the field office or supervisory interviewer staff were simply not available to make the required in person contacts in the data collection time frame. This was understandable since our experiment unfortunately coincided with an internal agency reorganization which included many office staff moving between offices and many offices temporarily understaffed. In addition, many field staff did not utilize the ARMS interviewer incentives. Several commented that they did not want to offer additional pay to only a subset of interviewers if not all interviewers had targeted cases in their workloads. Some field offices reported that they were planning to pay the extra incentives but did not inform the interviewers ahead of time that these extra payments were available. Obviously, if interviewers do not know that they can receive extra money, this does not act as an incentive to motivate them to use more effort to complete these particular cases.

DISCUSSION

We attempted to embed adaptive design strategies in two of NASS's surveys in order to strategically manage data collection and allocate resources effectively to sample units. Our initial tests of these approaches were somewhat disappointing and taught us several important lessons. Overall, the most important lesson is that it is very difficult to conduct research and true experimental comparisons in an operational survey environment. Based on just our empirical results we would have concluded that our experimental procedures were not effective in increasing response from the targeted operations. However, in the case of the ARMS, the procedures were not uniformly applied, and no valid comparisons can be made. In the case of the CS, it was difficult to make a true evaluation of our proposed procedures. For the CS, the group of states that we compared to our experimental group was quite diverse. Each of these states designed its own data collection strategies. These strategies could have been quite different from each other, and importantly, these strategies could even have been exactly the same as the strategies we were testing, if that office decided to use those approaches.

NASS field offices have always decided on data collection strategies that they felt were optimal given the budget, time and resources available to them. Those field offices that felt that our strategies were equal to or worse than their current approaches, did not adopt any new adaptive design procedures. Similarly, field offices would not accept a set of true experimental control procedures. This is because the field offices felt that a control condition with assignment of generic data collection procedures without consideration of operations' importance or other information known about them would result in lower response rates and quality or they did not have the sufficient resources to fully implement our proposed data collection plans. For operational data collection, including procedures felt to potentially decrease response or exceed available cost or personnel time was not acceptable.

Going forward we plan to continue to test adaptive design strategies in our operational data collections. However, there are several things we plan to do differently. First, we must do more communication about the tests with all staff involved. In this first test, the research staff worked closely with the operational staff in NASS headquarters to develop the procedures and instructions. However, the research staff had limited communication with the field office staff carrying out the procedures. As a result, in some cases the field office staff did not understand what they were expected to do, why they were to do it or did not understand that adherence to the procedures was important. Second, we must make sure that any procedures we develop are acceptable to the field staff. Even with greater communication, field staffs do not want to test procedures that will negatively impact their performance, especially with respect to expectations from their direct line supervisors. It is important to have good communication with both HQ and field staff and work with everyone involved in carrying out the research.

In addition, it is critical to monitor what actually happens in data collection, NOT just what is proposed. If we had not followed up with our field staff to review how our experiments had worked, we would never have known that procedures were not actually followed and any conclusions we would have drawn would simply have been wrong. Finally, more consideration to how we will evaluate the results should be made. Reserving true control groups for comparison is ideal, but in operational settings, this is often unrealistic. If true control groups cannot be formed, we may have to be satisfied with more qualitative evaluations of the results. For example, with our CS experiment, we should likely only be looking at the actual and predicted response rates in our test states and not compare them to any of the other states where data collection procedures were not dictated.

One of the other lessons learned in these experiments is simply that applying adaptive design in survey data collection is quite difficult. Survey field managers are already using all the strategies they can to get the highest response rates. Groups that we predicted to be harder to obtain responses from, were indeed those with the lowest response rates. Operations that are hard to gain cooperation with remain nonrespondents. Going forward, it may be more effective to focus efforts on those cases that are not the most likely to be nonrespondents. Perhaps increases in response rate are more likely to come from groups with mid range likelihood of responding.

Also, we had great difficulty in creating new strategies to apply to our surveys. In the case of CS, much of what we proposed was already being done by our field staff. We merely formalized it and proposed that it be done systematically. For the ARMS survey, obvious alternative strategies such as switching mode are not options. Our suggestions to use senior staff and additional in person contacts are difficult to implement and are obviously still only strategies that can be used for small sets of operations (and at great cost).

Our initial attempts at applying adaptive design in NASS surveys illustrates just how difficult this is to accomplish and evaluate. As we move forward with efforts in this area, careful attention will be made to all aspects of planning, implementing and evaluating this research.

References

Earp, M., Mitchell, M, McCarthy, J.S., & Kreuter, F. (under review). Adaptive Data Collection in Establishment Surveys: Using Proxy Data and Tree Modeling to Identify Likely Nonrespondents and Reduce Bias.

McCarthy, J.S., Jacob, T, & McCracken, A. (2010). Modeling Non-response in National Agricultural Statistics Service Surveys Using Classification Trees, Research and Development Division Research Report Number RDD-10-05, US Department of Agriculture, National Agricultural Statistics Service.

Mitchell, M., &McCarthy, J.S. 2012. Voted "least likely to respond": Using Classification Trees to Identify Likely Non-respondents and Proactively Manage Data Collection in NASS's Quarterly Agricultural Survey. Presented at the Fourth Annual International Conference on Establishment Surveys, Montreal, Quebec, Canada.

Mitchell, M., & Ott, K. (In Preparation). Targeted Data Collection Efforts for 2011 and 2012 ARMS III.

Ridolfo, H., Boone, J. and Dickey, N. (2013). Will They Answer the Phone if They Know It's Us? Using Caller ID to Improve Response Rates. Research and Development Research Report Number RDD-13-01, US Department of Agriculture, National Agricultural Statistics Service.