Adaptive Design in an Establishment Survey: Targeting, Applying and Measuring 'Optimal' Data Collection Procedures in the Agricultural Resource Management Survey

Tyler Wilson[1], Jaki McCarthy[2], and Andrew Dau[3]

Abstract

For NASS's annual Agricultural Resource Management Survey (ARMS), targeted data collection procedures were developed based on input from field staff and tested in an adaptive design. For both 2014 and 2015, ARMS records with nonresponse propensities between 50 and 69% were identified. Half were randomly assigned alternative data collection strategies designed to increase response rates; the other half followed NASS's standard mail, nonresponse follow-up procedures. No significant increase in response rates was obtained in 2014. Neither were any differences found between edit and imputation rates between samples. The 2014 results and field office feedback on the experimental procedures led to slight changes to the procedures for 2015. These changes were not effective in increasing response rates either. Potential reasons as to why these alternative procedures on targeted hard-to-get respondents were ineffective at increasing response rates are provided.

Key Words: In-person Contact, Establishment Survey, Propensity Scores, Drop Off, Response Rates

1. INTRODUCTION

Many surveys are turning to adaptive or responsive survey designs in data collection to either minimize costs, reduce non-response bias or improve data quality. By identifying key design features affecting survey costs and errors, survey managers can either proactively adapt or dynamically respond by altering those features in the survey data collection process. Thus, multiple "phases" of a survey with differing conditions are combined in estimation and referred to as adaptive survey design (Groves and Heeringa, 2006). Targeting subsamples by propensity scores and the formulation of experimental data collection procedures for these groups were motivated by adaptive design strategies.

The Agricultural Resource Management Survey (ARMS) is administered annually by the National Agricultural Statistic Service (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 with commodities that have been targeted for that given year (targeted commodities change annually). The second phase (ARMS II) focuses on production expenses, chemical use,

[1] Tyler Wilson is a Survey Statistician & Methodologist with the National Agricultural Statistics Service, Research and Development Division, 1400 Independence Ave, Washington DC, 20520
[2] Jaki McCarthy is a Senior Cognitive Research Psychologist with the National Agricultural Statistics Service, Methodology Division, 1400 Independence Ave, Washington DC, 20520
[3] Andrew Dau is a Mathematical Statistician with the National Agricultural Statistics Service, Methodology Division, 1400 Independence Ave, Washington DC, 20520

and the targeted commodity. The third phase focuses on financial data, such as expenses and income as well as risk management practices. The focus of this research is on ARMS III, which will be referred to as ARMS. For 2014 and 2015 adaptive designs were applied to targeted subsamples during the data collection process.

Before 2012, the majority of ARMS data collection was by personal interview. However, as sample sizes continued to grow and resources were restricted, data collection methods evolved into a mail out with an in-person follow-up process. *Table 1* below provides a historical look at ARMS sample sizes and response rates.

Table 1: History of Response Rates

ARMS III Sample Year	Sample Size (n)	Respondents (n_r)	Nonrespondents (n_n)	Response Rates (%)
2000	17,903	11,295	6,608	63.09%
2001	13,313	8,500	4,813	63.85%
2002	18,219	13,484	4,735	74.01%
2003	33,861	21,282	12,579	62.85%
2004	33,908	22,966	10,942	67.73%
2005	34,937	24,684	10,253	70.65%
2006	34,203	23,227	10,965	67.91%
2007	31,924	22,251	9,673	69.70%
2008	36,388	24,052	12,336	66.10%
2009	33,328	22,753	10,595	68.27%
2010	35,421	23,285	12,146	65.73%
2011	34,070	22,130	11,636	64.95%
2012	32,096	21,638	10,458	67.40%
2013	33,315	17,428	15,887	52.30%
2014	43,770	29,733	14,037	67.90%
Total	466,656	308,708	157,663	66.16%

As prescribed by the Office of Management and Budget (OMB), response rates lower than 80% require U.S. Federal agencies such as NASS to conduct nonresponse bias analyses and make efforts to increase their response rates. The research described here is part of NASS's efforts to use adaptive design to increase response rates in ARMS and builds on previous work developing nonresponse propensity models for ARMS (Earp & McCarthy, 2011; Mitchell, Ott & McCarthy, 2015; Mitchell, Ott, Ridolfo & McCarthy, 2015).

1.1 Targeted Subsampling with Propensity Scores

Beginning in 2010, classification trees were developed to target specific samples of the ARMS based on their likelihood of responding. These classification trees segment a dataset by a set of simple rules. The rules assign an observation to a segment based on the input variable that maximizes the difference between two groups based on the target (in this case response and nonresponse). The dataset is sequentially subset by these rules until there are no more splits that can be created. No more splits can be produced when the sample size per segment is too small, no more significant splits can be generated, or the maximum depth of the tree is too large.

Classification trees create a hierarchy (tree) where the segments are called nodes. The first node, known as the root node, contains the entire dataset. From the root node, there

are branches or paths to and from nodes within a tree. Terminal nodes are nodes that have no branches coming from them and are known as leaves. Each record will appear in only one of the leaves, and the leaves will collectively contain all the records in the dataset. The leaves of interest are those that have the highest proportion of records with the target (in this case, refusal). An illustration using training data is shown below:

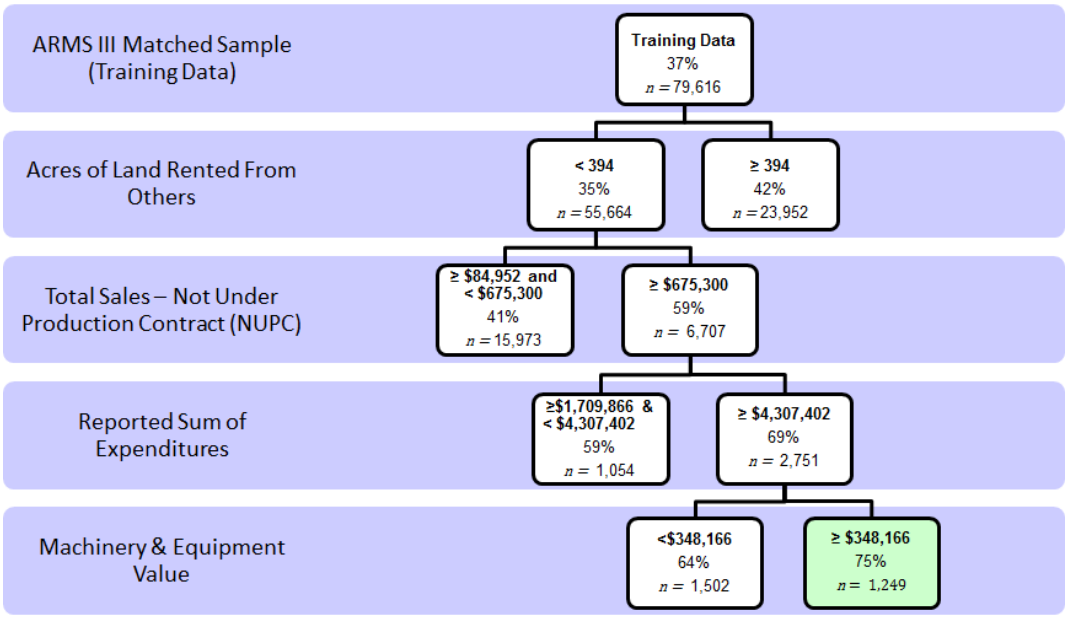

Figure 1: Example Classification Tree

In 2012, NASS began using these classification trees to identify ARMS respondents with nonresponse propensities above 70% (Mitchell, Ott & McCarthy 2015; Mitchell, Ott, Ridolfo & McCarthy 2015). In addition, respondents were assigned to impact groups based on their importance in nonresponse weighting and the final, overall estimate (Earp, McCarthy, Porter, & Kott 2010). Split subsamples of records with high nonresponse scores and high impact scores were identified. Approximately half were given to high level NASS staff for contact (i.e. Directors, Deputy Directors and State Statisticians) and half went through standard mail and field office follow-up.

After several years of split sample testing, results of these data collection methods on response rates for highly unlikely survey respondents with high impact scores were inconclusive. Although some increases in response rates were found, the samples were too small to say anything conclusive, and the overall impact of these changes were minimal on survey estimates. Thus, in 2014, classification trees identified a larger subsample of operations with a midrange 50-69% nonresponse propensity. These records were *not* stratified into impact groups. This larger subsample, while still predicted to be nonrespondents, were thought to be more likely to cooperate if given more 'persuasive' data collection methods than those predicted highly unlikely to respond.

1.2 Data Collection Strategies

The more 'persuasive' data collection methods tested in 2014 were based on a drop off, pick up (DOPU) interviewer, on-the-ground approach to data collections. As shown in the research by Melevin et al. 1999 and Steele et al. 2001, this approach has resulted in

some increases in response rates when compared to more traditional mail-out and follow-up processes. The approach relies solely on personal contact to administer and/or deliver the survey. Even when administration by the interviewer is not immediately available, it is possible to leave the survey with the respondent for self-administration.

The process involves the use of token items and trusted sponsorship (branding) within the immediate context of the survey itself. This is often accomplished with a labeled bag where the survey and items of interest are placed, resulting in an all-in-one data collection package. This package is used during in-person contact to solicit cooperation. Such a package facilitates communication and personal interaction during the initial contact period. After establishing contact, the survey can be immediately administered, rescheduled, or left with the respondent to complete on their own time. These increases in personal contact, token items, sponsorship, and flexibility in the DOPU process may be responsible for increased cooperation and response rates.

Cooperation rates and response rates are often used interchangeably; however, these rates are not the same. Response rate is the number of eligible respondents who complete the survey. Cooperation rate is the number of eligible respondents *personally contacted* who complete the survey. Measuring and comparing both rates in this type of data collection process is critical because oftentimes response rates between samples do not reflect the effect of in-person interaction. For example, Melevin et al.'s (1999) found that when respondents were personally handed the data collection packet 65.6% completed the survey; whereas, only 54% responded when the packet was given to someone else in the household and 17.4% responded when no contact was established. Any type of survey research that relies solely on a boots-on-the-ground approach to data collection benefits from reporting both response and cooperation rates.

2. METHODS

In 2014 and 2015, all records with nonresponse propensity scores between 50% and 69% were identified in the ARMS sample. 1558 records were identified in 2014 while 1170 records were identified in 2015. At the start of each ARMS, these groups were randomly split into two subsamples – alternative (treatment) and standard (control).

For both years, the control subsamples received standard ARMS data collection procedures involving an initial mail out. All mail nonresponse records were then sent to field offices to be called and possibly visited by an interviewer. For the treatment subsamples, initial mail out was held and NASS interviewers were required to visit each possible respondent in-person (more details in following section).

Response and cooperation rates between samples were the measures of interest in this study. However, edit and imputation rates between samples were also examined the first year to evaluate whether changes in data collection methods affected the accuracy of reported data. In addition, qualitative field office feedback was collected between years to understand how the alternative data collection methods could be better implemented the following year. The following section outlines the motivations used to generate the alternative data collection methods.

2.1 Alternative Data Collection Procedures for the Treatment Samples

Motivated by adaptive design and the DOPU framework of data collections, four basic guidelines were used when forming 2014 and 2015 alternative data collection procedures:

1. Increase face-to-face interaction and interviewer support during survey distribution and administration
2. Increase the use of informational materials and token items of interest around the survey
3. Limit the number of opportunities of easy refusals
4. Increase the flexibility of response by allowing immediate administration or self-administration by respondent

The first element required interviewers to visit the operation in-person and locate the operator. This visit would be the first time the respondent sees the ARMS survey. If the interviewer failed to establish contact on the first visit, they were required to leave a NASS-sponsored door hanger with contact information and a date indicating when they would be returning to try again. It is possible the respondent could call and refuse; however, interviewers were instructed to revisit the operation if at all possible.

The second element to the experimental procedures increased the use of informational and token items surrounding the survey, providing specific talking points for the interaction to help survey cooperation. The personal visit and use of token items were also used to convey that cooperation was expected. The token and informational items are listed and illustrated below:

1. *Statistical fact sheet from previous studies*
2. *A cover letter signed by the Administrator of NASS*
3. *Incentive item (pens, notebooks, gloves, rain gauges, and snow scrapers depending on region)*
4. *Privacy and mail-back envelope*
5. *Postcard with contact information*
6. *Information regarding how to take the survey online*

These items were prepackaged by field offices in a plastic USDA/NASS-sponsored bag and given to interviewers prior to delivery.

The third element required only in-person contacts (i.e. no initial contact by mail or telephone), which was intended to reduce the number of easy refusals. Standard ARMS data collection starts with a mail out where a respondent can choose to complete, refuse, or ignore. If ignored, oftentimes phone calls are used to establish contact. During the planning stages of this research many interviewers and field office staff voiced that respondents decided to refuse before being contacted by interviewers and were sometimes covertly refusing the survey by saying they 'plan' on filling out the survey via mail-back or web form. The alternative data collection methods eliminate the overt

refusal via mail and the covert refusal via telephone, starting the process with face-to-face interaction.

The final element sought to increase flexibility for the respondent by allowing immediate administration or self-administration. Once face-to-face interaction was achieved, all modes of survey administration were immediately available. For example, if a farmer is baling hay when the interviewer arrives, the interviewer can say, *"If you would like to complete it now, I am available. If not, my contact information is in there if you need help filling it out or to schedule an appointment. There is also information on how to fill it out online or mail it back. If I don't hear back from you, I'll be back to pick up the survey in two weeks. Enjoy the gloves, they are in the bag."* Most field offices also included a mail-back envelope as well as information regarding where and how to take the survey online.

2.2 Tracking and Analysis

As was done in years past, a supplement form was used to verify if contact was being made within the treatment sample and the overall feasibility of the new procedures. These forms allow for quick adaptive changes to be verified before next year's data collection begins. As is common in establishment surveys, some sample units may have special data collection handling procedures in place prior to the current survey. It was assumed that some records in the subsamples had these special arrangements or were coordinated and done jointly with other surveys. Supplement sheets allowed us to track and filter out these records when doing the final analysis.

Other questions within the two supplement sheets tracked whether interviewers were able to establish contact with the respondent. In addition, questions tracked how well interviewers were able to execute the treatment procedures, because, unlike other methods, this process required personal contact.

2.3 Field Office Feedback

In addition to the supplement sheets, between years 2014 and 2015 a qualitative survey feedback was requested from survey administrators and interviewers. Eight of NASS's 12 Regional Field Offices (RFOs) provided feedback from those who directly handled the research records (a total of 22 people).

The first question asked how many in-person attempts they felt were needed to establish face-to-face contact with the interviewee. The majority (13) answered that 3-4 attempts were needed. Five answered that only 2 attempts were needed, while three others said 5 or more attempts were needed.

The second question asked if all the bags, surveys and auxiliary items were received within a reasonable amount of time – 'reasonable' meaning that RFOs and interviewers were able to sort and put together items for alternative treatment records *before* data collection began. Twenty-one out of 22 answered yes. These answers reflect a generally high level of feasibility for the alternative data collection procedures.

Two of the feedback questions were open-ended and allowed personnel to offer their comments, concerns, and suggestions. The most common theme within these answers pertained to valuing the time of the respondent. These types of comments related to how busy a farmer is – conveying the value of the operator's time, and that when an

opportunity arises to administer the survey, just do it. “The farmer needs to feel we know how valuable his time is.” “I don’t think dropping it off and coming back is such a good idea…doing it right away if they can…is still the best.” “Most of the better responses they got were because the operator said let’s finish this now or never.” These types of comments raise some concern over the understanding and communication of the new alternative procedures. It appears that some interviewers thought the alternative methods limited them to only dropping the survey off and not immediately administering the survey if possible. One supervisor commented, “More than one person was very upset that the enumerator would not do the interview at the first contact.” During 2015 ARMS, interviewers were retrained to administer the survey immediately upon contact whenever possible.

Other comments suggested pre contacting respondents via the telephone and using response history instead of propensity scores to subsample records. “I think we should have our enumerators set up an appointment to complete the interview rather than just a drop off.” “Operators should be pre-selected as previous refusals or inaccessibles. Several cases were operators who normally cooperate with us and complete the surveys.” Telephone calls were not allowed in 2014 so that respondents could not refuse before actually seeing the survey alongside other token items of interest. However, in 2015 ARMS, we allowed phone calls prior to interviewer visits to help increase the in-person contact rate. Comments about subsampling by response history instead of propensity scores were of little concern because response history was one of the many variables from which the propensity scores were derived.

In the last question, personnel were asked if they thought delivering or administering ARMS in-person without mailing it first made a difference. Most answered ‘yes’ (n=16), some answered ‘not sure’ (n=5), and only 1 answered ‘no.’ Those who did answer ‘not sure’ only handled 1 or 2 cases, and felt they lacked the evidence to answer the question. “I didn’t have enough of these to make a real decision (only 1).” Finding that a majority agreed that no pre-mail out was good for survey cooperation is also not a surprise. This strategy was taken directly from suggestions from past research as well as the Supervisory Enumerator Advisory Council (SEAC) who often acts as the interviewers’ voice at NASS.

3. RESULTS

Before ARMS data collection began, some records within the subsamples were deemed ineligible for alternative treatment due to establishment restrictions. These were identified and removed prior to testing. In 2014, 659 out of 774 records remained eligible to receive alternate data collection methods. In 2015, 574 out of 597 records remained eligible to receive alternate data collection methods.

3.1 Response and Cooperation Rates

The following table shows the response disposition of the treatment and control cases:

Table 2: Response Rates

Type of Response	2014 Eligible Treatment (n=659)	2014 Control (n=778)	2015 Eligible Treatment (n=574)	2015 Control (n=573)
Completion	59%	58%	38%	37%
Refusal	34%	33%	46%	46%
Inaccessible	6%	6%	12%	12%
Out-of-Scope	1%	2%	5%	5%

The response rates for both 2014 and 2015 differed by only 1 percentage point between samples with no significant difference in response rates for the alternative data collection methods.

Cooperation rates were the secondary measure of interest between samples. In 2014, using a maximum of 2 contact attempts, NASS interviewers were able to make in-person contact with 57% of the respondents in the treatment sample. The following year, when contact attempts (4) were doubled and pre-visit phone calls were made, in-person contact remained at 57% in the treatment sample.

Overall response rates in the treatment group were higher in 2014[4]. However, cooperation rates within the treatment group were still higher for both 2014 and 2015. The similarity of the escalation in cooperation rates by year is represented in the graph below.

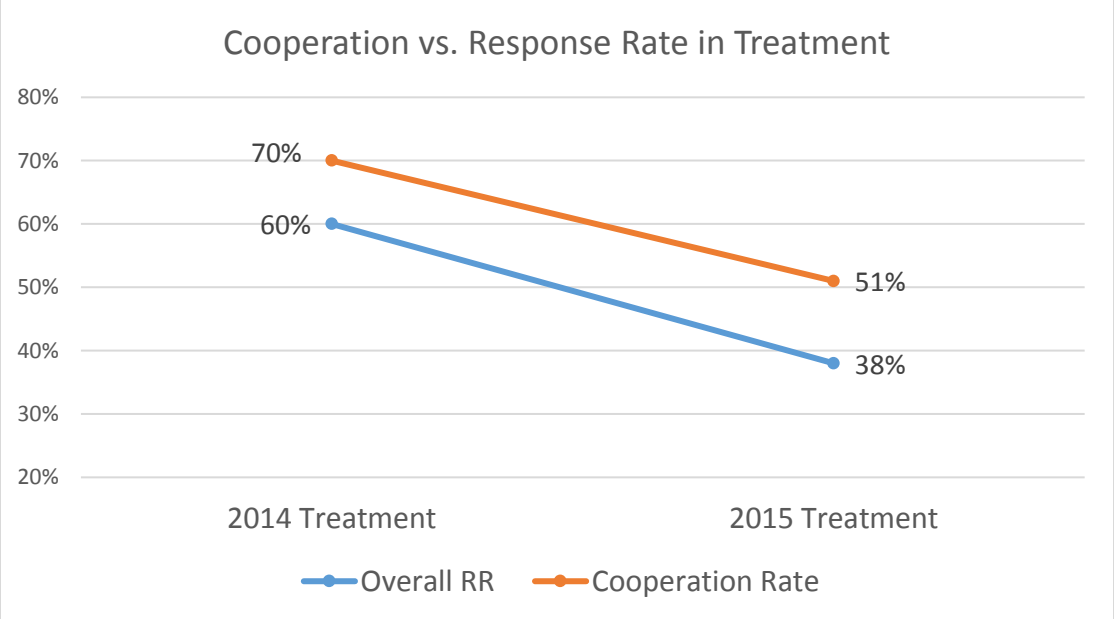

Figure 2: Cooperation Rates

[4] In 2014, ARMS was combined with another NASS survey that carried mandatory reporting authority. Only during the Census of Agriculture (every 5 years) is ARMS combined with another survey.

When in-person contact was established, completion rates were ten percentage points higher in 2014 than for the treatment sample overall; similarly, completion rates were thirteen percentage points higher for records that were contacted in 2015.

3.2 Edit and Imputation Rates for 2014

It was thought that alternative data collection procedures might also impact reported data quality. For example, if respondents have the assistance of in-person interviewers, they may be more motivated to provide quality data or have assistance for difficult items. However, because standard data collections use a mixed-method approach and some respondents do receive assistance from interviewers, we questioned whether any differences between samples would be evident. In 2014, edit and imputation rates within a few critical sections of ARMS were examined.

Three variables (questions) were selected: *Operator Seed Expense* (IC600), *Value of Owned Land* (IC854) and *Operator Depreciation* (IC756) for an initial evaluation of data quality. The answers to these questions are highly important to data users and are derived through different edit and imputation methods. *Seed Expense* allows no statistical imputation (only manual by analyst), while *Value of Owned Land* has moderate statistical imputation rates. *Operator Depreciation* has a very high rate of statistical imputation. For each of these three items, the number of times the respondents' reported data were either edited or imputed was calculated.

Table 3 below shows that 2.63% of the usable records[5] in the treatment sample had adjustments made to the reported data on *Seed Expense*, while 5.03% of the records in the control sample had adjustments made.

Table 3: Seed Expense

Mode	Treatment			Control		
	N	*Edits/Imputations*	*Percent*	*N*	*Edits/Imputations*	*Percent*
Mail	64	2	3.13%	135	10	7.41%
Phone	7	0	0.00%	9	0	0.00%
Face	344	8	2.33%	256	10	3.91%
Web	4	1	25.00%	37	2	5.41%
Total	**419**	**11**	**2.63%**	**437**	**22**	**5.03%**

Table 4 below shows that 21.96% of usable records in the treatment sample had adjustments to the reported data from the *Value of Owned Land*. Once again, a higher proportion (24.49%) of records in the control group had adjustments to the reported data.

[5] Number of usable records are completions in each sample that were 'in-business' and had over $1000 of sales or the potential for over a $1000 of sales.

Table 4: Value of Owned Land

Mode	Treatment			Control		
	N	*Edits/Imputations*	*Percent*	*N*	*Edits/Imputations*	*Percent*
Mail	64	18	28.13%	135	45	33.33%
Phone	7	3	42.86%	9	4	44.44%
Face	344	70	20.35%	256	50	19.53%
Web	4	1	25.00%	37	8	21.62%
Total	**419**	**92**	**21.96%**	**437**	**107**	**24.49%**

Lastly, the third variable examined was *Operator Depreciation*. *Table 5* below shows that 32.22% of records in the treatment sample had adjustments made from the reported data. A similar proportion (32.27%) of reported responses in the control group were altered. *Operator Depreciation* is generally a hard figure for respondent to accurately report since it is only calculated once a year for tax filing purposes. This type of question is a good reminder that the accuracy of reported data is often a function of the type of question asked and not how the answer was collected.

Table 5: Operator Depreciation

Mode	Treatment			Control		
	N	*Edits/Imputations*	*Percent*	*N*	*Edits/Imputations*	*Percent*
Mail	64	18	28.13%	135	39	28.89%
Phone	7	1	14.29%	9	6	66.67%
Face	344	114	33.14%	256	92	35.94%
Web	4	2	50.00%	37	4	10.81%
Total	**419**	**135**	**32.22%**	**437**	**141**	**32.27%**

A Wilcoxon test of significance was conducted on the totals of each question revealing no statistical difference between samples at the 5% level.

Table 6: Wilcoxon Two-Sample Test Two Sided P-Values

Variable	Normal Approximate	T Approximate
Seed Expense	.1752	.1847
Value of Land	.5785	.5785
Depreciation	.6503	.6506

In summary, alternative data collection methods do not appear to lead to differing amounts of edit and imputation rates but other reported data quality indicators are being evaluated.

4. CONCLUSION AND RECOMMENDATIONS

In 2014, this research was considered a pilot study involving: (1) a new alternative approach to data collection for a targeted subsample, (2) adaptive data collection methods emphasizing in-person contacts, and (3) an increased effort of communication and feedback from field personnel. Records with nonresponse propensity scores between 50 and 69% were targeted because they provided larger sample sizes than records with higher nonresponse scores and were a group that might be more likely to respond with new data collection procedures than the highest nonresponse propensity group. It was also thought that this group would be less likely to respond by mail, and thus benefit more from an in-person data collection strategy. Response rates were not significantly increased in 2014 and were followed by gathering feedback from field personnel. The alternative data collection methods were revised and retested in 2015. The methods for subsampling records by their propensity score (50-69% nonresponse) remained the same.

Overall response rates did not significantly increase in 2015. In 2014, the in-person contact rates (57%) suggested that interviewers had trouble locating and making face-to-face contact with the respondent. This remained the case in 2015, even when the amount of contact attempts were doubled and pre-visit telephone calls were allowed. Thus many of the records never received some of the adaptive design and DOPU components critical to survey cooperation. This becomes more evident when examining the difference between overall response rates and cooperation rates in both the 2014 and 2015 treatment samples.

The 2014 feedback received from field personnel provided anecdotal support for the adaptive changes to the 2015 alternative data collection methods. However, these changes had little effect on empirical results. For example, a majority (73%) of field staff stated that no pre mail-out helped avoid survey refusal and gain cooperation; however, no increase in overall response rates was found when no pre-mail out was sent.

Adapting alternative data collection procedures for unlikely ARMS respondents performed no better than NASS's standard, mixed-method approach to data collection. It is recommended that additional evaluation be given to the targeting procedures involving nonresponse propensity scores to determine whether other strategies can be used to target specific hard-to-get records. Also, additional research will need to examine why many interviewers were not able to locate and/or interact with respondents within the treatment subsamples. Proactive targeted enumeration based on coverage importance is also being investigated for 2016 ARMS. In addition, survey organizations may want to use adaptive design strategies to target resources to those cases where alternative strategies may have the potential to improve the quality of the data collected. In the case of increasing response rates using an adaptive design in an establishment survey, finding more effective data collection methods than what are already in place remains a difficult, yet worthwhile, task.

5. REFERENCES

Earp, M.S., McCarthy, J.S., Porter, E., & Kott, P.S. (2010). Assessing the Effect of Calibration on Nonresponse Bias in the 2008 ARMS Phase III Sample Using Census 2007 Data. In *JSM Preceedings*, American Statistical Association. Vancouver, British Columbia: American Statistical Association.

Earp, M. and McCarthy, J. (2011). Using Nonresponse Propensity Scores to Improve Data Collection Methods and Reduce Nonresponse Bias. In *JSM Proceedings*, American Association of Public Opinion Research. Phoenix, AZ: American Statistical Association.

Groves, R. M. and Heeringa, S. G. (2006), Responsive design for household surveys: tools for actively controlling survey errors and costs. Journal of the Royal Statistical Society: Series A (Statistics in Society), 169: 439–457.

Melevin, P. T., Dillman, D. A., Baxter, R., & Lamiman, C. E. (1999). Personal delivery of mail questionnaires for household surveys: A test of four retrieval methods. *Journal of Applied Sociology*, *16*(1), 69-88.

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., 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.

Steele, J., Bourke, L., Luloff, A. E., Liao, P. S., Theodori, G. L., & Krannich, R. S. (2001). The drop-off/pick-up method for household survey research. *Community Development*, *32*(2), 238-250.