



United States
Department of
Agriculture



National
Agricultural
Statistics
Service

Research and
Development Division
Washington DC 20250

RDD Research Report
Number RDD-14-02

November 2014

Using Nonresponse Propensity Scores to Set Data Collection Procedures for the Quarterly Agricultural Survey

Melissa Mitchell
Kathy Ott
Jaki McCarthy

This report has been prepared for limited distribution to the research community outside the U.S. Department of Agriculture (USDA). The views expressed herein are not necessarily those of National Agricultural Statistics Service or USDA.

EXECUTIVE SUMMARY

Starting in 2009, national nonresponse propensity score models were used for the Quarterly Agricultural Survey (QAS) to identify likely nonrespondents. Initially, these scores were given to Field Offices (FO) for use at their own discretion. Over time, FOs were given guidelines on how to use the scores in their data collection efforts. In 2012, the national-level model was replaced with state-specific models to give more accurate nonresponse propensity scores.

Two questions drove this research:

- Can the models identify potential nonrespondents in the QAS?
- Can standardized data collection procedures be developed that use the nonresponse scores to increase response rates for the QAS?

Structured data collection protocols were developed and tested in September and December 2012 for the QAS. The data collection protocols used the nonresponse propensity scores in conjunction with strata to assign data collection method codes. Nine treatment states used the new data collection protocols in September and six states in December to test whether the protocols increased response rates. Non-participating states served as the comparison states.

Using the overall response rates from the survey, we determined that the models do identify potential nonrespondents for the QAS, identifying both refusals and inaccessibles well.

Results for increasing response rates were mixed. Predicted response rates were compared to actual response rates for the treatment and comparison states. The results showed the treatment states were more effective at targeting refusal operations than the comparison states, but the converse was true for inaccessible operations. Interpreting the differences in response rates can be misleading, however, since non-treatment states could also have used the same data collection methods as prescribed by the treatment, and treatment states did not always use the prescribed methods, making it difficult to evaluate specific treatments.

Note: The name of the Quarterly Agricultural Survey changed to the Crops Acreage and Production Survey (APS) after this research was completed, but for the purposes of this report, the original name will be used.

RECOMMENDATIONS

1. Continue to use nonresponse propensity scores for the Quarterly Agricultural Survey (QAS).
2. Continue to develop and evaluate data collection method procedures to increase response rates for both refusals and inaccessibles/noncontacts. The procedures tested in this study were inconclusive, so more work on effective ways to increase response rates, while containing costs, is needed.
3. Expand the use of nonresponse propensity scores to other surveys. Other surveys may find they are good predictors of nonresponse, and be able to develop tailored data collection strategies to specific subgroups.
4. Investigate a more formal way to identify impact operations to further target operations.
5. Standardize data collection procedures agency-wide to facilitate rigorous and clean evaluations of specific data collection protocols. Unless this is done, tests and experiments will not provide conclusive results.

TABLE OF CONTENTS

1. INTRODUCTION	1
2. METHODS	3
2.1 Targeted Data Collection	3
3. RESULTS	7
3.1 Can we use the models to identify potential nonrespondents in the Quarterly Agricultural Survey (QAS)?	7
3.1.1 September Results.....	8
3.1.2 December Results	10
3.2 Can we develop standardized data collection procedures that use the nonresponse scores to increase response rates for our surveys by targeting potential nonrespondents? ..	14
3.2.1 Following procedures.....	14
3.2.2 Predicted vs Actual Response Rates	15
4. CONCLUSIONS AND RECOMMENDATIONS	17
4.1 Recommendations.....	18
5. REFERENCES	20
APPENDIX A: Memos regarding targeted data collection.....	A-1
APPENDIX B: Response rates for treatment and comparison states.....	B-1
APPENDIX C: State level predicted vs actual response	C-1

Using Nonresponse Propensity Scores to Set Data Collection Procedures for the Quarterly Agricultural Survey

Melissa Mitchell, Kathy Ott, and Jaki McCarthy¹

Abstract

In order to target nonrespondents proactively, the National Agricultural Statistics Service (NASS) began using propensity nonresponse models for the Quarterly Agricultural Survey (QAS). Since 2009, NASS has used national models to identify likely nonrespondents for the QAS. For the September and December 2012 surveys, targeted data collection procedures were created, partly based on the nonresponse propensity scores, and evaluated. We compared response rates between treatment and comparison states to determine the scores' efficacy. In addition, we examined the difference in actual and predicted response rates for the treatment and comparison states to determine whether the targeted data collection procedures increased response rates.

Key Words: nonresponse, classification trees, targeted data collection

1. INTRODUCTION

Survey nonresponse has been a growing concern for many years. Survey response rates are declining in both public and private sectors. Approaches to reduce survey nonresponse include using face-to-face interviews, sending a personalized letter, or using incentives (Dillman, 1978; Groves and Couper, 1998). Calibration weighting is one of the ways to adjust for survey nonresponse (Earp et al, 2008; Kott, 2005). In 2009, NASS started developing a technique to proactively identify likely nonrespondents by using classification trees. Previous research by McCarthy and Jacob (2009) and McCarthy, Jacob, and McCracken (2010) modeled survey nonresponse for the Quarterly Agricultural Survey (QAS) with classification trees. Earp and McCarthy (2011) used classification trees to identify likely nonrespondents for the Agricultural Resource Management Survey Phase III (ARMS III). Identifying likely nonrespondents before data collection allows the use of more effective data collection strategies to maximize response, reduce nonresponse bias, and contain costs.

This paper describes an evaluation of a set of field offices using standardized procedures compared to a set of field offices using a variety of data collection protocols. This project involved the development and field testing of a set of standard procedures used by selected states, but because this was done as was part of the operational data collection, the other states continued using the procedures they had in place. This evaluation was not set up as a controlled experiment, but we can compare the actual response rates to the predicted response rates for our

¹ Melissa Mitchell and Kathy Ott are Mathematical Statisticians, and Jaki McCarthy is the senior cognitive research methodologist with the National Agricultural Statistics Service, Research and Development Division, 3251 Old Lee Highway, Room 305, Fairfax, VA 22030.

selected states and for all the other states. We did not examine or evaluate how targeting these operations could impact or bias estimates from the QAS.

Our research questions were:

- Can models identify potential nonrespondents in the QAS?
- Can standardized data collection procedures be developed that use the nonresponse scores to increase response rates for the QAS?

Data mining is a machine-learning technique that is able to handle large datasets like the one used for this study. Within data mining, classification (or decision) trees are used to predict the outcome of a binary variable, such as survey response/nonresponse, from auxiliary data. The primary objective of classification trees is classification of groups (in our case respondent/nonrespondent operations). Classification trees are non-theoretical and completely data driven; no hypothesis is proven or disproven, the only concern is how well they can classify into groups (survey response/nonresponse in this case).

Logistic regression is another common technique used to predict the outcome of a binary variable. However, classification trees are preferred to logistic regression in this situation for multiple reasons. First, in logistic regression, operations with missing data are usually removed from the analysis, whereas classification trees retain such operations. Especially in cases where we are interested in survey response/nonresponse, missing values can be indicative of what we are predicting. Second, a limited number of predictors can be used in a regression analysis. Our data have over 60 predictor variables. It is not ideal to use all of these variables (and possible interactions) in a regression model but all of these variables can be used in a classification tree analysis. By using classification trees, we do not have to specify the variables in the model beforehand; it automatically detects significant relationships and interaction effects without pre-specification which reduces the risk of selecting the wrong variables or other model specification errors (Schouten, 2007). Logistic regression requires pre-specification of the variables to include in the model which leaves the user at risk for selection bias, inclusion of the wrong variables, and other issues. Also, the more variables we include in our model, the harder it is to avoid issues of multicollinearity in logistic regression. This is less of a problem for classification trees (Phipps & Toth, 2012). Variables that do not predict survey response/nonresponse will simply be dropped from the classification tree model.

Once developed, the models identified likely respondents and nonrespondents. This allowed the possibility for targeted data collection methodology to use limited funds effectively. For example, for likely respondents, cheaper data collection methods, such as mail or telephone, may be used. This allows saving more costly personal interviews for likely refusals and inaccessible (i.e., noncontacts). In addition, knowing the likelihood of specific operations to respond could help during data collection by allowing changes to data collection strategies part way through the data collection period. For example, near the end of data collection, changing the mode or interviewer for the remaining records that are more likely to respond may positively impact response rates and data collection costs.

Widely used data collection methods that effectively solicit response include the use of incentives, personal enumeration, and multiple calling attempts on different days and times. These methods can be tailored to specific operations based on their relative likelihood of responding. Managing these data collection methods can help contain monetary costs while attempting to increase (or maintain) response rates.

2. METHODS

McCarthy and Jacob (2009) and McCarthy, Jacob, and McCracken (2010) developed national nonresponse models for NASS surveys using classification trees. These trees have been used since their creation to identify and flag likely nonrespondents for the QAS. Records are assigned a rank order “refusal” score ranging from one to four and an “inaccessible” score ranging from one to five. Records with refusal/inaccessible scores of one are the most likely to be a refusal/inaccessible, and those with a score of four or five are the least likely to be a refusal/inaccessible.

2.1 Targeted Data Collection

As mentioned earlier, targeting data collection for records (i.e., operations) based on their propensity to respond can possibly help increase response rates and/or contain data collection costs. These gains can be maximized if the operations’ impact on the survey results are also taken into account.

In addition to possible gains in response rates and decreased costs, managing data collection based on an operation’s impact on the survey results could help reduce nonresponse bias. The distribution of agricultural production across operations is often highly skewed, with some operations contributing a large percentage of a particular commodity. With this in mind, we wanted to identify the impact an operation has on survey estimates so we could use our resources to maximize response rates for those operations that cover the most production.

The QAS uses a multivariate probability proportional to size (MPPS) sample design. However, sampled operations are also assigned to strata, which are used in post data collection for item-level imputation and nonresponse weighting adjustments. Strata definitions vary by state, but generally indicate either gross measures of the operation’s size or presence of significant (or rare) commodities within the state.

The strata value was used to determine the impact of an operation on the QAS estimates, which has advantages and disadvantages. Strata are readily available on the NASS sampling frame for all operations. Because strata are clearly defined within each state, no development time was needed to use them. However, the strata differ by state, making cross-state comparisons difficult. Also, specialty strata that target operations with control data for commodities important to state level estimates such as potatoes, cherries, or pineapple differ by state, so they cannot uniformly help identify operations that have a high impact on a particular commodity

across states. In addition, the QAS strata only take into account the value of sales or the acreage for a commodity, but no other variables. Given the pros and cons, as well as time and resource constraints, strata were chosen as the best indicator of impact for this study. The largest operations are assigned strata values of 90 and above, while specialty operations are assigned values of 70-79 in the states that use them.

NASS assigns each operation a data collection method code (DCM) to control data collection. The DCM codes used for the September and December 2012 QASs were:

- (1) Mail only,
- (2) Mail with data collection center (DCC) or National Operations Center (NOC) phone follow-up,
- (3) Field Office (FO) handling, with cases sent to field enumerators to call first and then attempt in-person,
- (4) Mail by print mail center (PMC) with FO follow-up,
- (5) Office hold, and
- (6) Coordinated surveys.

For the September 2012 QAS, nine treatment states (KS, NE, ND, MD, NH, FL, IA, MI, SC) followed our instructions using the strata (our proxy for the impact of the operation), the propensity scores, and other variables to assign DCM codes. For the December 2012 QAS, six of the nine original states (KS, NE, ND, MD, NH, FL) continued to use the specific instructions. The remaining states were the comparison states.

For both the September and December 2012 QASs, high impact operations were defined as operations in strata 90 or above (typically very large producers in terms of value of sales) or in specialty commodity strata. Although only the treatment states were required to use our method code assignment criteria, nothing prohibited the comparison states from following a similar protocol.

The DCM assignment criteria are shown in Tables 1 and 2 for September and December.

Table 1: September Data Collection Method Code Criteria

Data Collection Method Code	Records assigned to that code
01 - Mail Only	<ul style="list-style-type: none"> • No specific records assigned.
02 – Mail with DCC Follow-up	<ul style="list-style-type: none"> • Low or mid impact operations with a comparatively low chance of being a nonrespondent (i.e., all records in strata less than 90 with a refusal nonresponse propensity score = 2- 4 or an inaccessible nonresponse propensity score =2-5)
03 - FO Handling	<ul style="list-style-type: none"> • Operations highly likely to be nonrespondents (records with a refusal nonresponse propensity score =1 or an inaccessible nonresponse propensity score= 1)

Data Collection Method Code	Records assigned to that code
	<ul style="list-style-type: none"> • Operations with the same target/primary operator (Opdom 85/45) • All high impact records (Records in strata 90 or above) • Records where a partner switch was done sometime throughout the year. • Special contact arrangements made with the operator.
04 - Mail by PMC with FO Follow up	<ul style="list-style-type: none"> • No specific records assigned
05 - Office Hold	<ul style="list-style-type: none"> • Known zeros • Previous contact agreements • Special situations (dangerous or possible violent situation, etc.)
06 - Coordinated with surveys	<ul style="list-style-type: none"> • Matches with another NASS survey
<p>Call-out</p> <p>(Call-out is the end of the survey cycle when the sample is looked at to see if changes in data collection mode could help increase response for specific subgroups in the population if needed.)</p>	<p>Use the Nonresponse Propensity Scores and Strata to determine which records will have the highest impact on estimates and are most likely to respond. Pull those records back from the DCCs and send them to the field for calling. The DCC/NOC will keep the rest of the records for calling. Records that should be pulled back include:</p> <ul style="list-style-type: none"> • Records in strata 90 and 70-79 with a refusal nonresponse propensity score =4 or an inaccessible nonresponse propensity score =5 meet this criteria • See footnote*

*Unrelated to this project, county estimate coverage needs were considered (see Appendix A)

Based on state office feedback about their experience with the procedures in September, changes were made to the instructions for December, as reflected in Table 2. During the September study, treatment states gave feedback on the data collection protocol resulting in slight changes to the instructions. The modifications for December included: specialty commodities were assigned to personal visit (PV); EO strata and highly likely to refuse/be inaccessible can be mailed first; disconnected phone numbers were considered; and records with an existing appointment were held in the DCC during call-out. These changes are shown in Table 2 in italics.

Table 2: December Data Collection Method Code Criteria
(changes from September are shown in italics)

Data Collection Method Code	Records assigned to that code
01 - Mail Only	<ul style="list-style-type: none"> No specific records assigned.
02 – Mail with DCC Follow-up	<ul style="list-style-type: none"> All low or mid impact farms with a comparatively low chance of being a nonrespondent (i.e., All records in strata less than 90 with a refusal nonresponse propensity score = 2- 4 or an inaccessible nonresponse propensity score =2-5), <i>except specialty commodities.</i>
03 - FO Handling	<ul style="list-style-type: none"> Related records (85/45 records) Records where a partner switch was done sometime throughout the year. Special contact arrangements made with the operator. <i>Disconnected phone number</i>
<i>03 - FO Handling or 04 - Mail by PMC with FO Follow up</i>	<ul style="list-style-type: none"> <i>All records that are highly likely to be nonrespondents (i.e., All records with a refusal nonresponse propensity score =1 or an inaccessible nonresponse propensity score =1)</i> <i>All high impact records (i.e., Records in the EO Strata (90 or above))</i> <i>All records in specialty commodity strata</i>
05 - Office Hold	<ul style="list-style-type: none"> Known zeros (records that we know don't have the commodity of interest) Previous contact agreements (records that we have data collection agreements with) Special situations (dangerous or possible violent situation, etc.)
06 - Coordinated with surveys	<ul style="list-style-type: none"> Matches with another NASS survey
Call-out	<p>Use the Propensity Scores and Strata to determine which records will have the highest impact on estimates and are most likely to respond. Pull those records back from the DCCs and send them to the field for calling. The DCC/NOC will keep the rest of the records for calling. Records that should be pulled back include:</p> <ul style="list-style-type: none"> Records in strata 90 and 70-79 with a refusal nonresponse propensity score =4 or an inaccessible nonresponse propensity score =5 meet this criteria See footnote* <p><i>Keep records with existing appointments in the DCC</i></p>

*Unrelated to this project, county estimate coverage needs were considered (see Appendix A)

See Appendix A for the memoranda and specifications sent to the field offices each quarter.

Measures of impact and the likelihood of responding were used to guide data collection methods in an effort to maximize response and coverage while minimizing costs. Hence the farms most likely to be nonrespondents and the high impact farms were targeted for field enumeration by the field offices (DCM codes = 3 and 4). This mode of data collection has the highest response rates, but also the highest cost. Cases that were sent for field enumeration were completed by field enumerators, either on the phone or in person. These cases were not sent to the DCC or the NOC, but were called or visited in person by field enumerators. Often, enumerators are familiar with the local farms, particularly large ones, because those are interviewed for multiple surveys. Therefore, field enumerators may be better equipped to perform the data collection.

In general, operations that were not as important to the estimates and operations that were highly likely to respond were mailed questionnaires and had phone follow-up. This method of data collection is much cheaper than field enumeration, so it was a good choice for ‘easier’ (likely to respond) operations and those that had less potential impact on the estimates.

Towards the end of data collection, state offices asked the NOC or DCC to send some of the cases that were not yet complete back to the state office for one last attempt at getting responses. NASS refers to this process as “callout.” During callout for the September and December 2012 QASs, treatment states took back operations that were highly likely to respond and had the most anticipated impact on the estimates. These were operations in strata 90 or above and specialty strata, if applicable (70-79), with a nonresponse propensity score of 4 for refusals or 5 for inaccessibles. These cases were returned to the state offices for two reasons. First, they were fairly likely to respond. The notion was that a state office may have information that could help obtain a response from these impact operations. Second, these cases were in the defined impact group, so they were most likely to impact at least one estimate for the state. The offices may also have field enumerators who were particularly good at obtaining response in-person or who knew more information about the operation.

3. Results

3.1 Can models identify potential nonrespondents in the Quarterly Agricultural Survey (QAS)?

To address this question, response rates were compared across the nonresponse propensity groups for the September and December 2012 QASs. The rates for the treatment and comparison states were similar, so only the overall response rates are shown in this section. The separate response rates for the treatment states and the comparison states are shown in Appendix B for informational purposes.

3.1.1 September Results

Tables 3 and 4 and Figure 1 show the overall response rates for records flagged as highly likely nonrespondents, both refusals and inaccessibles, within the nonresponse propensity groups for the September 2012 survey. Again, separate response rates for the treatment states and the comparison states are shown in Appendix B.

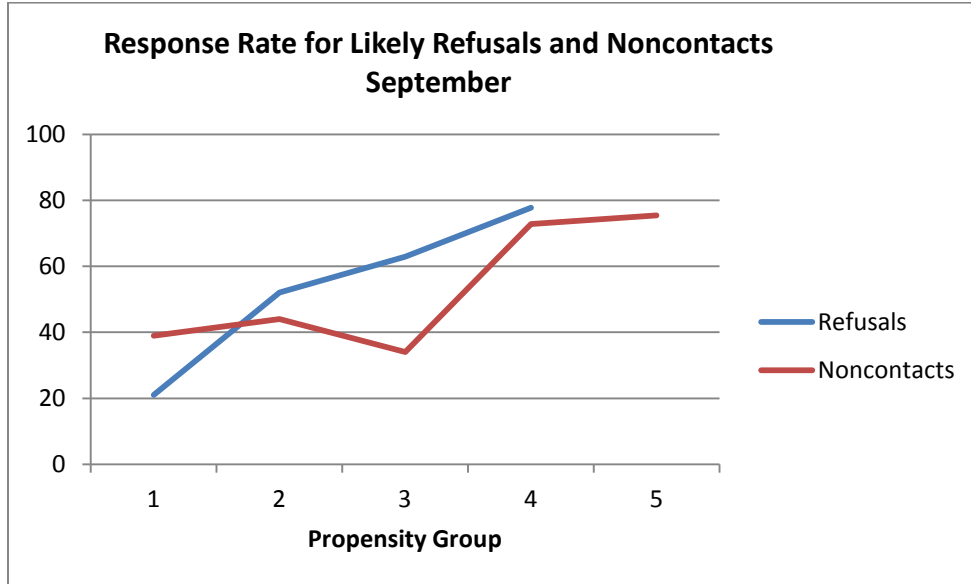
Table 3: September 2012 response rates for flagged refusals within propensity groups

Refusal	Overall response rate (%)
Score =1 (most likely to refuse)	21.06
Score=2	52.06
Score=3	62.92
Score=4 (least likely to refuse)	77.78
Total N	60,320

Table 4: September 2012 response rates for flagged inaccessibles within propensity groups

Inaccessible	Overall response rate (%)
Score=1 (most likely to be inaccessible)	38.94
Score=2	44.04
Score=3	34.02
Score=4	72.80
Score=5 (least likely to be inaccessible)	75.42
Total N	60,320

Figure 1 September 2012 Overall Response Rate by predicted outcome



Tables 3 and 4 and Figure 1 show that the models worked fairly well at predicting nonresponse for the 2012 September Crop/Stocks survey because, in general, response rates for the records that were most likely to be refusals or inaccessible were indeed much lower than for the records that were predicted less likely to be nonrespondents.

Overall response rates for operations that were predicted most likely to be nonrespondents are very low – 21.06 percent for refusals and 38.94 percent for inaccessibles. The response rate increases as predicted for refusals for all propensity groups. The response rate for inaccessibles generally increases with propensity group, but decreases between operations in propensity groups 2 and 3. It is unclear why this dip occurs in response rates for these two groups. The decrease could be somewhat artificial, as the operations with a noncontact propensity score equaling 1 were the operations states focused on in the study. It is possible that this additional attention to the operations most likely to be inaccessible increased the response rate for that group. This could explain why operations in that group have a similar and higher response rate than those with scores equaling 2 and 3.

High impact operations for this study were those in strata greater than or equal to 90 (the largest operations) and those in the specialty strata, i.e., strata 70-79, if applicable. Response rates were calculated for the high impact operations to see if the scores accurately identified potential nonrespondents in that group. Again, response rates for the treatment states and comparison states were done separately, but showed no difference in the response trend, so those response rates are broken out in tables in Appendix B. Tables 5 and 6 below show the response rates for the high impact operations across all states for September 2012.

Table 5: September 2012 response rates for all high impact operations within propensity score groups for flagged refusals

Refusal	Overall response rate (%)
Score=1 (highly likely)	18.46
Score=2	52.53
Score=3	65.08
Score=4 (least likely)	79.42
Total N	30,581

Table 6: September 2012 response rates for all high impact operations within propensity score groups for flagged inaccessibles

Inaccessible	Overall response rate (%)
Score=1 (highly likely)	38.09
Score=2	44.72
Score=3	31.63
Score=4	73.39
Score=5 (least likely)	74.32
Total N	30,581

The same pattern shown in Tables 3 and 4 emerge in Tables 5 and 6. In general, response rates for the high impact records that are mostly likely to be refusals and inaccessibles are indeed lower than for the high impact records that were predicted to be less likely to be nonrespondents.

3.1.2 December results

Tables 7 and 8 show the overall response rates for the 2012 December Crop/Stocks survey by propensity group. Appendix B shows the breakdown by the treatment and comparison states.

Table 7: December 2012 response rates for flagged refusals within propensity groups

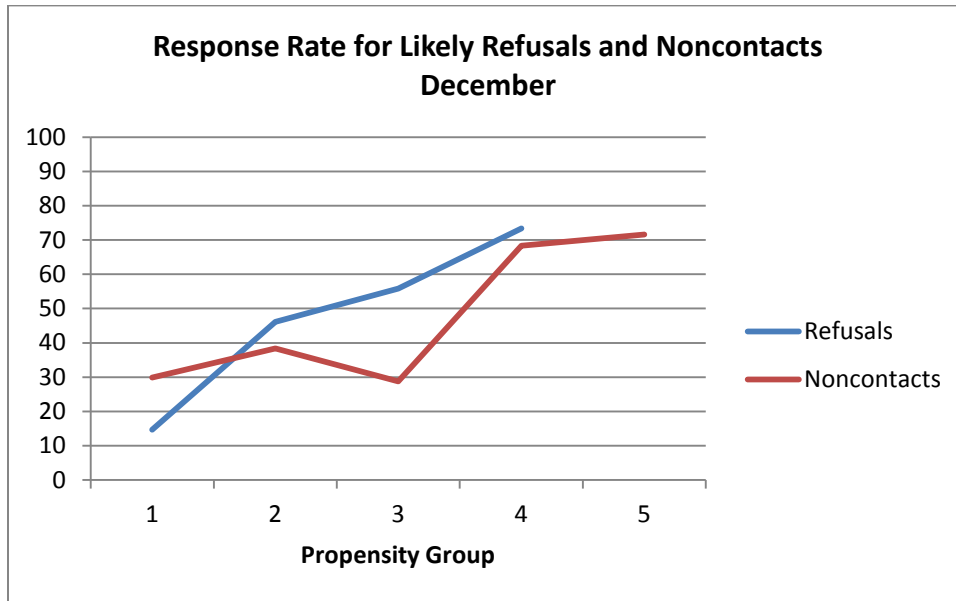
Refusal	Overall response rate (%)
Score =1 (most likely to refuse)	14.55
Score=2	45.91
Score=3	55.58
Score=4 (least likely to refuse)	73.35
Total N	73,027

Table 8: December 2012 response rates for flagged inaccessible within propensity groups

Inaccessible	Overall response rate (%)
Score=1 (most likely to be inaccessible)	29.80
Score=2	37.98
Score=3	28.66
Score=4	68.18
Score=5 (least likely to be inaccessible)	71.25
Total N	73,027

As in September, Tables 7 and 8 for December show the models are fairly good at predicting nonresponse. In general, response rates for the records that are mostly likely to be refusals and inaccessible are indeed quite a bit lower than for the records that were predicted to be less likely to be nonrespondents. This is shown graphically for the December survey in Figure 2.

Figure 2 December 2012 Response Rate by predicted outcome



As in September, high impact operations in December were those with strata values greater than or equal to 90 and the specialty strata (70-79). Overall, 12.28 percent of the high impact operations flagged as highly likely refusals responded to the survey and 29.26 percent of the high impact operations flagged as highly likely inaccessible responded to the survey. As seen in September, in December, the response rate for inaccessibles generally increases with propensity group, but decreases between operations in propensity groups 2 and 3. It is unclear why this dip occurs in response rates for these two groups.

Tables 9 and 10 show the response rates for the high impact operations by propensity score. Appendix B shows the breakdown for the treatment and comparison states.

Table 9: December 2012 response rate for all high impact operations within propensity groups for flagged refusals

Refusal	Overall rate
Score=1 (highly likely)	12.28
Score=2	46.68
Score=3	56.38
Score=4 (least likely)	74.43
Total N	36,607

Table 10: December 2012 response rates for all high impact operations within response propensity groups for flagged inaccessibles

Inaccessible	Overall rate
Score=1 (highly likely)	29.26
Score=2	34.28
Score=3	25.70
Score=4	68.77
Score=5 (least likely)	70.16
Total N	36,607

The same patterns shown in Tables 7 and 8 emerge in Tables 9 and 10. In general, response rates for the high impact records that were most likely to be refusals and inaccessibles are indeed lower than for the records predicted to be less likely to be nonrespondents.

Tables 3-10 show that the nonresponse propensity scores predict nonresponse fairly accurately for the Quarterly Agricultural Surveys for both the entire sample and the high impact operations.

3.2 Can standardized data collection procedures that use nonresponse scores increase response rates for the Quarterly Agricultural Surveys?

As stated earlier, although the treatment states agreed to follow the data collection procedures provided, other states may also have used the same procedures (or used the same procedures for some subset of their records), but there was no systematic way to determine how similar the procedures used by other states were to the treatment procedures. Also, based on the debriefing questions with the treatment states after September 2012, the treatment procedures were very similar to those normally used during data collection to set data collection method codes. These two issues could cause effects in the treatment group to be small, and make it difficult to interpret the results comparing the treatment and comparison states.

3.2.1 Following procedures

Method Codes used for the September QAS were summarized for four treatment states (IA, MD, MI, and NE) for which we had Method Code data. Treatment assignments were given for Method 2 (mail with phone follow up) and Method 3 (FO handling, to be sent to field enumerator to call first, then do personal visit). Table 11 shows that the treatment protocol was often not followed.

Table 11: September Records targeted for Method Codes 2 and 3

	IA	MD	MI	NE
Should use Method Code 2	2,246	880	1,163	2,123
Used Method Code 2	1,722 (76.7%)	519 (59.0%)	307 (26.4%)	1,489 (70.1%)
Should use Method Code 3*	655	293	314	672
Used Method Code 3	122 (18.6%)	174 (59.4%)	306 (97.5%)	538 (80.0%)

* records with MPROFREF=1 (most likely to be a refusal) or MPROFINN=1 (most likely to be an inaccessible) or strata>=90 (typically large operations; large capacity/sales/land etc), or odstat=85 or odstat=45 (related record; operator status is either 85 or 45)

Table 11 does not take into account operations that should not have been assigned Method Code 2 or 3 and were assigned them anyway. From Table 11, we can see that IA, MD, and NE assigned at least half of the operations that the procedures specified to be assigned Method Code 2 to the treatment method. All of these states used Method 3 (field enumerators) to some extent when they should have used Method 2 (mail with phone follow up). There were much fewer operations that should have been assigned to Method Code 3 compared to Method Code 2. With the exception of IA, at least 50 percent of the operations that were supposed to be assigned Method Code 3 were assigned according to the treatment procedures.

Generally, data collection by mail with phone follow up does not get as high of a response as personal visit, but given the relatively high propensity to respond and relatively low impact of the identified operations, the tradeoff of response rates for costs was deemed appropriate for the treatment protocol. However, apparently field offices often disagreed and were apprehensive about using mail with phone follow up for several operations. Iowa, Maryland, and Nebraska also used data collection efforts defined by other Method Codes (not 2 or 3) to operations that should have used Method Codes 2 and 3.

Operations were also identified that should have been targeted for the callout (if necessary), but since there was no Method Code for follow up procedures, we could not determine which received follow-up.

3.2.2 Predicted vs. Actual Response Rates

Because states have different baseline response rates, it may be misleading to directly compare any given state's response rate to another. Therefore, each state's actual response rate was compared with that state's predicted response rate (from the state level models) to determine if that state's response rate was higher or lower than predicted by the model.

We used the national model to flag these records. However, in order to compare state-level response rates, we applied the model rules to each state's samples to obtain a state specific response rate predicted by the model. State level models varied in the number of levels they have (i.e., more terminal nodes). Therefore, all states have a group of highest likelihood nonrespondents (NRP=1), but the number of rank order groups after that varied by state models. Therefore, we focused this comparison on the highest likelihood group.

For the operations that were the most highly likely to be refusals and the operations that were most highly likely to be inaccessible, we calculated the difference between the actual response rates and the predicted response rates for each state and averaged them together to get one number for the treatment states and one number for the comparison states. Using this calculation, the higher the positive value, the better the increase in response rate. Negative values indicate decreases in response, with larger negative numbers showing larger decreases. Table 12 shows the average difference for the treatment and comparison states. Appendix Tables C1 and C2 provide the actual and predicted response rates for each of the treatment and comparison states, respectively, along with some discussion. As shown in those appendix tables, several states in both the treatment and comparison states had better than predicted response rates, but several had lower than predicted response rates.

Table 12. September 2012 summary of response rates for treatment and comparison states (for highly likely nonrespondent operations)

	Difference (Actual %-Predicted %) (%)	
	Treatment	Comparison
Highly Likely Refusal	2.78	-4.60
Highly Likely Inaccessible	9.36	14.67

Table 12 shows that during the September 2012 QAS, the treatment states had better response for predicted refusals, but worse response for likely inaccessibles when compared to the comparison states. As mentioned earlier, however, the treatment effect should be small because all states could use the treatment procedures, and those treatment procedures closely matched what states already do.

As in September, we also compared the actual response rates and predicted response rates for the most highly likely nonrespondents for December because of possible different baseline response rates among states. Again, although we used the national model to flag these records, we applied the model rules to each state's samples to obtain a state specific response rate predicted by the model. Also similar to September, we only compared the records in the NRP=1 (most highly likely nonrespondents) group because it was the only group that was the same across all state models. Using this calculation, the higher the positive value, the better the increase in response rate. Negative values indicate decreases in response, with larger negative numbers showing larger decreases. Appendix Tables C3 and C4 provide the actual and predicted response rates for each of the treatment and comparison states, respectively. As shown in those appendix tables, and consistent with September results, several states in both the treatment and comparison states had better than predicted response rates, but several had lower than predicted response rates.

Table 13. December 2012 summary of response rates for treatment and comparison states (for highly likely nonrespondent operations)

	Difference (Actual %-Predicted %) (%)	
	Treatment	Comparison
Highly Likely Refusal	0.20	-11.95
Highly Likely Inaccessible	-5.09	5.51

Table 13 shows that the difference between the actual and predicted response rates were better for the treatment state refusals, but worse for inaccessibles. This is similar to September results.

4. CONCLUSION AND RECOMMENDATIONS

In general, the prediction models accurately identified refusals and inaccessibles for the Quarterly Agricultural Survey. Operations in categories with higher predictions of nonresponse had consistently lower response rates. Inaccessibles were a bit mixed, but generally, the same trend held. The research question regarding whether targeting the nonrespondents increases response rates is more complicated.

As stated earlier, this was an evaluation of a set of field offices using standardized procedures compared to a set of field offices using a variety of data collection protocols, but was not set up as a controlled experiment. Data collection methods were developed with input from the participating states, so the treatment protocol was similar to what many states already do. The treatment data collection methods may have also been used by comparison states. Finally, the treatment protocol was not always followed. For these reasons, it is difficult to compare the results between the treatment and comparison states and to generalize the findings.

Results from the September and December experiments were mixed when comparing the differences between predicted and actual response rates. For both the September and December surveys, when compared to the comparison states, the difference between the actual and predicted response rates was higher for the treatment states' predicted refusals, but lower for the predicted inaccessibles. However, improvements for refusals were small and, as discussed, the treatment protocol was not always followed, so we do not know if the treatment states were more effective at targeting the refusal operations than the comparison states.

Due to the caveats already mentioned, we cannot determine whether substantial gains in response rates were attained using the standardized data collection methods developed for this research. However, the nonresponse models were shown to accurately predict nonrespondents. It is possible that different procedures for setting data collection method codes might be effective at increasing response rates. Also, it may be advantageous to create different procedures for operations that are highly likely to refuse versus those that are highly likely to be inaccessible. The procedures and methods used for these groups to increase response are likely different.

Although almost all records were sent to the field, the response rate for the high impact records (those in the largest and specialty commodity strata) that were flagged as the most likely to be nonrespondents was only 10-17 percent. NASS should consider spending this money on other groups that might show better increases such as the highly likely to be inaccessible group or the groups that have refusal or inaccessible propensity scores of 2 or 3. For example, it may be better to target operations with propensity to refuse scores of 2 or 3 (i.e., medium likely to refuse) with Field Office special handling since it may be easier to gain their cooperation.

One area that still needs additional work is defining high impact operations. For this study, high impact only involved operations in strata 90 or above and specialty commodities, if applicable. However, other criteria may better identify high impact operations. For example, certain control data may more accurately indicate an operation's impact on a particular estimate. This could be

taken into account up front to classify operations into impact groups. NASS should look into what variables would be the best impact indicators based on the variables of interest.

NASS should continue to develop standardized data collection procedures that target nonrespondents by exploring alternatives to the procedures used here. Ideas include analyzing the relationship between response rates, mode, and nonresponse score; targeting the operations that are somewhat likely to be nonrespondents; and considering special data collection procedures for the largest farms.

We could not make clean comparisons of the treatment states and comparison states for several reasons. The main reason is that the data collection procedures across state offices or regional offices are not standardized. In most cases, we did not know what procedures state offices used, but only what we asked them to do. It is difficult to evaluate new or different procedures if it is unclear exactly what is being compared. Also, as is often the case during research projects that use production samples and data, we could not completely separate the treatment and comparison state's procedures or enforce the procedures in the treatment states and regions. If we want to evaluate changes to procedures, we need to move as an agency to standardized procedures that can be modified for specific states or regions for testing. NASS is currently undergoing that process.

While this study was ongoing, Research and Development Division developed state-level models and implemented them starting in the June 2013 QAS data collection period. The state models were built using the same variables as the national models, which include demographic information, list frame variables, response history indicators, joint burden indicators, as well as QAS and Census of Agriculture variables. These models vary in size and complexity; some states have small, condensed trees while other states have much larger, expansive trees. This means that some state trees have only 1 level (only one important splitting variable) while other state trees have 5 or 6 levels. Although the models vary in size and complexity across states, the only types of variables in all the models are response history variables and joint burden indicators. Although the trees vary in size and complexity, a rank score of 1 still means that an operation is a highly likely refusal or inaccessible. The national model works well broadly, but it is possible that specific models for particular states will perform better than the national model at predicting likely refusals and inaccessibles. No evaluation has been done on the state-level models at this time. In addition, no evaluation has been done to examine the impact on estimates or nonresponse bias of using these targeted data collection techniques.

4.1 Recommendations

Based on the research done, the following are recommended:

1. Continue to use nonresponse propensity scores for the Quarterly Agricultural Survey (QAS).

2. Continue to develop and evaluate data collection method procedures to increase response rates for both refusals and inaccessible/noncontacts. The procedures tested in this study were inconclusive, so more work on effective ways to increase response rates, while containing costs, is needed.
3. Expand the use of nonresponse propensity scores to other surveys. Other surveys may find they are good predictors of nonresponse, and be able to develop tailored data collection strategies to specific subgroups.
4. Investigate a more formal way to identify impact operations to further target operations.
5. Standardize data collection procedures agency-wide to facilitate rigorous and clean evaluations of specific data collection protocols. Unless this is done, tests and experiments will not provide conclusive results.

REFERENCES

- Dillman, D. (1978). *Mail and Telephone Surveys: The Total Design Method*. New York: Wiley & Sons.
- Earp, M. and McCarthy, J. (2011). Using Nonresponse Propensity Scores to Improve Data Collection Methods and Reduce Nonresponse Bias. JSM Proceedings, American Association of Public Opinion Research. Phoenix, AZ.
- Earp, M.S., McCarthy, J.S., Schauer, N.D., & Kott, P.S. (2008). Assessing the Effect of Calibration on Nonresponse Bias in the 2006 ARMS Phase III Sample Using Census 2002 Data. Research and Development Division Staff Report RDD-08-06, United States Department of Agriculture, National Agricultural Statistics Service.
- Groves, Robert M., and Mick P. Couper. (1998). *Nonresponse in Household Interview Surveys*. New York: Wiley.
- Kott, P.S. (2005). Using Calibration Weighting to Adjust for Nonresponse and Coverage Errors. *Survey Methodology* 32:133-142.
- Maxwell, Scott E., and Delaney, Harold D. (2004). *Designing Experiments and Analyzing Data: A Model Comparison Perspective, 2nd edition*. New Jersey: Lawrence Erlbaum Associates, Inc.
- McCarthy, J.S., & Jacob, T. (2009). Who Are You? A Data Mining Approach to Predicting Survey Non-respondents. JSM Proceedings, American Association for Public Opinion Research. Hollywood, Florida: American Statistical Association.
- 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.
- Phipps, P. & Toth, D. (2012). Analyzing Establishment Nonresponse Using an Interpretable Regression Tree Model with Linked Administrative Data. *Annals of Applied Statistics*, 6 (2), 772-794.
- Schouten, B. (2007). A Selection Strategy for Weighting Variables Under a Not-Missing-at-Random Assumption. *Journal of Official Statistics* 23(1): 51-68.

Appendix A
Memos regarding targeted data collection

September 2012 Memo (Although South Dakota is included on this memorandum, they opted out of the study before data collection started).

***** OFFICIAL NOTICE *****

DATE: August 2, 2012

TO: Acting Regional Directors
State Directors
Deputy Directors
Survey Statisticians
Field Offices: KS, NE, ND, SD, MI, MD, NH, FL, SC, IA

THROUGH: Chris Messer
Chief
Program Administration Branch

FROM: Everett Olbert
Commodity Surveys Section
Program Administration Branch

Subject: Non-Response Propensity Score Pilot Study for September 2012

ACTION: Utilize the Outlined procedures to set data collection methods for the September Crops/Stocks

DUE DATE: A Final Decision on participation in the pilot is needed by the COB on Monday, August 6, 2012

Pilot Study Evaluation Questions are due by the COB on Friday, October 5, 2012

The Program Administration Branch needs your assistance in evaluating the use of Non-response propensity scores. Some of you assisted HQ with this last quarter and it was very much appreciated. The survey team has been asked to conduct this test again this quarter. HQ has been asked to work with states that had lower response rates and those who volunteered. Ultimately, senior management would like for states to follow the same procedures for the survey process right down to how initial method codes are set. This pilot is intended to help set the rules for those procedures.

Appendix A
Memos regarding targeted data collection

The goal of this pilot study is to see if the instructions provided are helpful in the establishment of data collection plans for September. We also want to determine if the procedures are any different from what you are doing now. Try to follow the instructions provided as closely as possible when setting method codes and completing the CALLOUT process for September Crops/Stocks.

Note, the survey instructions were written from the perspective of a survey statistician. Field Office management is asked to discuss the outlined procedures with your survey statisticians prior to making a final decision on whether to participate in the pilot.

RDD has provided evaluation questions for your completion after method codes have been assigned for each record (see page 3 of the survey instructions). Please send an email containing your response to these questions to Leslee Lohrenz (with a cc to Scott Cox) by the COB on October 5, 2012.

Please send an email to Everett Olbert (with a cc to Scott Cox) indicating whether your state is willing to participate in the pilot by the COB on Monday, August 6, 2012. I apologize for the last minute notice.

Please call Everett Olbert at 202-720-4332 or Scott Cox at 202-720-4028, if you have any questions.

Non-Response Propensity (NRP) Scores Instructions
September 2012 Aq Surveys

The non-response propensity scores will be on the Sample Master. Field Offices will no longer need to download the indicators in their extract. However, FOs will need to make sure these varnames are pulled from the Sample Master into SMS.

The Varnames are as follows:

X3 will become MPROPREF

Propensity Score – Refusal
(1 - most likely to refuse; 4 - least likely to refuse)

X4 will become MPROPINN

Propensity Score – Inaccessible
(1 - most likely to be inaccessible; 5 - least likely to be inaccessible)

Appendix A Memos regarding targeted data collection

Records that are high impact on estimates and highly likely to be inaccessible or refusal should be sent to the field. (Unless they have an arrangement with the Operator or they are dangerous/violent/threatening.) The rest of the records will be sent to the DCC/NOC for calling.

Suggestions for coding your Method Codes:
Method 01 - Mail Only

Method 02 – Mail with Phone Follow-up by DCC:

- All records in strata less than 90 with MPROPREF = 2- 4 or MPROPINN=2-5

Method 03 - FO Handling Only; send to field enumerators to call first:

- All records with MPROPREF=1 or MPROPINN=1
- Records in the EO Strata (90 or above)
- 85/45 records
- Records where a partner switch was done sometime throughout the year.
- Special contact arrangements made with the operator.
- These records should be sent to field enumerators and called at least 3 times before attempting a personal interview.

Method 04 - Mail by PMC – FO Follow up:

- PMC will label 2 questionnaires: 1st set mailed from PMC, 2nd set sent back to the FO and data collected by Field Enumerators.

Method 05 - Office Hold:

- Known zeros
- Previous contact agreements
- Special situations (dangerous or possible violent situation, etc.)

Method 06 - Coordinated with surveys:

- Matches with surveys

During the September Crops/Stocks CALL-OUT:

The Report generated in Blaise and sent to you by your DCCs will have the Propensity Score for each record. This was put on the Sample Master with the Varname shown in parenthesis. It is shown on the report at REF and INACC scores.

REF Score and INACC Score

REF Score (MPROPREF)

Propensity Score – Refusal

Appendix A Memos regarding targeted data collection

(1 - most likely to refuse; 4 - least likely to refuse)

INACC Score (MPROPINN)

Propensity Score – Inaccessible

(1 - most likely to be inaccessible; 5 - least likely to be inaccessible)

CALLOUT SUMMARY FOR STATE = 21										PAGE: 1				
DATE: 5/23/2012										TIME: 10:08 AM				
										TIME OF ATTEMPTS				
ST	STR	ID	T	S	DISCTY	STATUS	NUM OF ATTEMPTS	6am - 10am	10am - 2pm	2pm - 6pm	6pm - 10pm	REF SCORE	INACC SCORE	Enum Rmks
21	2004	100093550	1	1	50 137	NO ANSWER	1	0	1	0	0	0	0	

Use the Propensity Scores and Strata to determine which records will have the highest impact on estimates and are most likely to respond.

- Records in strata 90 and 70-79 with MPROPREF=4 or MPROPINN=5 meet this criteria.
- County estimates coverage: you may also want to handle records in other strata based on consideration of how many records have been completed by county for the county estimates program.

Pull those records back from the DCCs and send them to the field for calling.

The DCC/NOC will keep the rest of the records for calling.

Appendix A
Memos regarding targeted data collection

December 2012 Memo (Although Iowa is included on this memorandum, they opted out of the study before data collection started).

***** OFFICIAL NOTICE *****

DATE: November 7, 2012

TO: State Directors
Deputy Directors
Survey Statisticians
All NASS Field Offices except AK

THROUGH: Barbara Rater
Chief
Survey Administration Branch

FROM: Everett Olbert
Commodity Surveys Section
Survey Administration Branch

Subject: Continuation of Non-Response Propensity Score Pilot Study for December 2012

ACTION: Utilize the Outlined procedures to set data collection methods for the December Crops/Stocks

DUE DATE: Pilot Study Evaluation Questions are due by the COB on Friday, January 04, 2013

Pilot States (KS, NE, ND, MD, NH, FL, IA):

The Survey Administration Branch needs your assistance in evaluating the use of Non-response propensity scores. Some of you assisted HQ in June and September and it was very much appreciated. The survey team will conduct this test again this quarter, working with the same states that participated in September. Ultimately, senior management would like for states to follow the same procedures for the survey process right down to how initial method codes are set. The continuation of this pilot study is intended to help set the rules for those procedures.

The goal of this pilot study is to see if the instructions provided are helpful in the establishment of data collection plans for December and whether the procedures, as outlined, help reduce non-response. We also want to continue looking at how the

Appendix A
Memos regarding targeted data collection

procedures are different from what you normally do. Follow the instructions provided as closely as possible when setting method codes and completing the CALLOUT process for December Crops/Stocks, noting any deviations in the feedback provided in January.

RDD has provided evaluation questions for your completion after method codes have been assigned for each record. These questions are located in the EDC Web Form located at the following link: <http://edc/webforms/form.asp?formid=387> and should be completed by January 4, 2013.

Non-Pilot States:

This documentation is also being provided to non-pilot States for use in developing your data collection strategy for December Crop/Stocks. These scores may help increase your response rates by assisting in assigning data collection method codes to records with high probability of being non-respondents.

Field Office management is asked to discuss the outlined procedures with your survey statisticians prior to making a decision on whether or not to utilize these instructions. Non-response propensity scores should still be evaluated regardless of whether or not the documentation is followed this quarter.

Please call Everett Olbert at 202-720-4332 or Suzanne Avilla at 202-720-5389, if you have any questions.

Non-Response Propensity (NRP) Scores Instructions
December 2012 Ag Surveys

In September, as part of a pilot research study, we asked some states to use the propensity scores to determine the data collection method used for the September Ag Survey sample. We are continuing that research study in December.

The non-response propensity scores will be on the sample master. Field Offices will no longer need to download the indicators in their extract. However, FOs will need to make sure these varnames are pulled from the sample master into SMS.

The varnames are as follows:

X3 will become MPROPREF	propensity score – Refusal (1 - most likely to refuse; 4 - least likely to refuse)
X4 will become MPROPINN	propensity score – Inaccessible

Appendix A

Memos regarding targeted data collection

(1 - most likely to be inaccessible; 5 - least likely to be inaccessible)

Refer to comments to determine operation-specific instructions for data collection (such as arrangements with the operator or a dangerous/violent/threatening situation. If there are no comments regarding special data collection, follow the coding recommendations below. Basically, records that are high impact on estimates and highly likely to be inaccessible or refusal should be sent to the field. The rest of the records will be sent to the DCC/NOC for calling.

Suggestions for coding your Method Codes:

Method 01 - Mail Only

Method 02 – Mail with Phone Follow-up by DCC:

- All records in strata less than 90 with MPROPREF = 2- 4 or MPROPINN=2-5 (except operations with specialty commodities in your state that should be assigned Method 03 or 04).

Method 03 –OR- Method 04 (“FO Handling Only; send to field enumerators to call first” OR “Mail by PMC – FO Follow up”)

- Records with MPROPREF=1 or MPROPINN=1
- Records in the EO strata (90 or above)
- Operations with specialty commodities that need ‘special attention’

Method 03 - FO Handling Only; send to field enumerators to call first:

- All records with MPROPREF=1 or MPROPINN=1 (or Method 04)
- Records in the EO strata (90 or above) (or Method 04)
- 85/45 records
- Records where a partner switch was done sometime throughout the year.
- Special contact arrangements made with the operator.
- Unless otherwise noted in the comments, these records should be sent to field enumerators and called at least 3 times before attempting a personal interview.

Method 04 - Mail by PMC – FO Follow up:

- All records with MPROPREF=1 or MPROPINN=1 (or Method 03)
- Records in the EO strata (90 or above) (or Method 03)
- PMC will label 2 questionnaires: 1st set mailed from PMC, 2nd set sent back to the FO and data collected by Field Enumerators.

Method 05 - Office Hold:

Appendix A Memos regarding targeted data collection

- Known zeros
- Previous contact agreements
- Special situations (dangerous or possible violent situation, etc.)

Method 06 - Coordinated with surveys:

- Matches with surveys

During the December Crops/Stocks CALL-OUT:

The report generated in Blaise and sent to you by your DCCs will have the propensity score for each record. This was put on the sample master with the varname shown in parenthesis. It is shown on the report at REF and INACC scores.

REF Score and INACC Score

REF Score (MPROPREF) propensity score – Refusal
(1 - most likely to refuse; 4 - least likely to refuse)

INACC Score (MPROPINN) propensity score – Inaccessible
(1 - most likely to be inaccessible; 5 - least likely to be inaccessible)

DATE: 5/23/2012										CALLOUT SUMMARY FOR STATE = 21					PAGE: 1 TIME: 10:08 AM		
										TIME OF ATTEMPTS							
ST	STR	ID	T	S	DISCTY	STATUS	NUM OF ATTEMPTS	6am - 10am	10am - 2pm	2pm - 6pm	6pm - 10pm	REF SCORE	INACC SCORE	Enum Rmks			
21	2004	100093550	1	1	50	137	NO ANSWER	1	0	1	0	0	0	0			

Use the propensity scores and strata to determine which records will have the highest impact on estimates and are most likely to respond.

- Records in strata 90 and 70-79 with MPROPREF=4 or MPROPINN=5 meet this criteria.
- County estimates coverage: you may also want to handle records in other strata based on consideration of how many records have been completed by county for the county estimates program.

Unless there is an existing appointment, pull those records back from the DCC and send them to the field for calling.

Appendix A
Memos regarding targeted data collection

The DCC/NOC will keep the rest of the records for calling, including all records that have an existing appointment.

Field Office Feedback (Due Date – 10/05/2012):

FOs will need to provide feedback to Research and complete the following questions.

Did you understand these instructions? If not, what was unclear?

Were you able to use the refusal and inaccessible propensity scores to determine a data collection plan for each group of records as described? If not, what factors impacted your ability to implement the plan?

Using the instructions, was the workload reasonable for your office?

If you did not (or could not) use these instructions as written, why not?

Are the instructions similar to what you would have done anyway?

What other information would have been useful on the instructions?

Appendix B

Response Rates for treatment and comparison states

The tables in this appendix expand on Tables 3-10 in the body of this report by breaking out response rates for the treatment states and the comparison states. The table numbers correspond to the table numbers in the body.

September Results

There were 13,056 total records for the treatment states. 1,507 of the operations in these states were flagged as highly likely to be a refusal (a refusal nonresponse propensity score of 1). 1,056 of the operations in these states were flagged as highly likely to be an inaccessible (an inaccessible nonresponse propensity score of 1). For the comparison states, there were 47,264 total records. 3,336 of the operations in these states were flagged as highly likely to be a refusal. 4,291 of the operations in these states were flagged as highly likely to be inaccessible.

Tables B3 and B4 show the response rates for refusals and inaccessibles within the nonresponse propensity groups for the September 2012 survey. In the treatment states, 22.96 percent of the 1,507 flagged highly likely refusal operations responded to the survey, and 44.32 percent of the 1,056 flagged highly likely inaccessible operations respond to the survey. In the comparison states, 20.20 percent of the 3,336 flagged likely refusal operations responded to the survey and 37.61 percent of the 4,291 of the flagged likely inaccessible operations responded to the survey.

Table B3: September response rates for refusals within propensity groups

Refusal	N	Response Rate – Treatment	N	Response Rate – Comparison	N	Overall response rate
Score =1 (most likely to refuse)	1,507	22.96% (n=346)	3,336	20.20% (n=674)	4,843	21.06% (n=1,020)
Score=2	1,925	46.65% (n=898)	5,382	53.99% (n=2,906)	7,307	52.06% (n=3,804)
Score=3	1,468	59.95% (n=880)	4,460	63.90% (n=2,850)	5,928	62.92% (n=3,730)
Score=4 (least likely to refuse)	8,156	73.12% (n=5,964)	34,086	78.89% (n=26,890)	42,242	77.78% (n=32,854)
Total	13,056		47,264		60,320	

Appendix B
Response Rates for treatment and comparison states

Table B4: September Response Rates for inaccessibles within propensity groups

Inaccessible	N	Response Rate – Treatment	N	Response Rate – Comparison	N	Overall response rate
Score=1 (most likely to be inaccessible)	1,056	44.32% (n=468)	4,291	37.61% (n=1,614)	5,347	38.94% (n=2,082)
Score=2	505	34.85% (n=176)	1,575	46.98% (n=740)	2,080	44.04% (n=916)
Score=3	769	31.86% (n=245)	2,015	34.84% (n=702)	2,784	34.02% (n=947)
Score=4	2,376	65.24% (n=1,550)	10,147	74.57% (n=7,567)	12,523	72.80% (n=9,117)
Score=5 (least likely to be inaccessible)	8,350	67.65% (n=5,649)	29,236	77.63% (n=22,696)	37,586	75.42% (n=28,347)
Total	13,056		47,264		60,320	

High impact operations for the purpose of this study were those in strata greater than or equal to 90 (the largest operations) and those in the specialty strata, i.e. strata 70-79, if applicable. There were 1,108 high impact operations that were flagged as highly likely refusals in the treatment states, with 19.49 percent responding to the survey. There were 688 high impact operations that were flagged as highly likely inaccessible in the treatment states. Of the 688 highly likely inaccessible operations, 41.57 percent responded to the survey. For the comparison states, 1,953 high impact operations were flagged as highly likely refusals, with 17.87 percent completing the survey. 2,418 high impact operations were flagged as highly likely inaccessibles in the comparison states, with 37.10 percent completing the survey. See Tables B5 and B6 for direct comparisons.

Appendix B
Response Rates for treatment and comparison states

Table B5: September response rate for all high impact operations within propensity score groups for refusals

Refusal	N	Response Rate – Treatment	N	Response Rate – Comparison	N	Overall response rate
Score=1 (highly likely)	1,108	19.49% (n=216)	1,953	17.87% (n=349)	3,061	18.46% (n=565)
Score=2	1,484	47.24% (n=701)	3,629	54.70% (n=1,985)	5,113	52.53% (n=2,686)
Score=3	848	63.80% (n=5,41)	2,331	65.55% (n=1,528)	3,179	65.08% (n=2,069)
Score=4 (least likely)	3,885	75.32% (n=2,926)	15,343	80.46% (n=12,345)	19,228	79.42% (n=15,271)
Total	7,325		23,256		30,581	

Table B6: September response rates for all high impact operations within propensity score groups for inaccessibles

Inaccessible	N	Response Rate – Treatment	N	Response Rate – Comparison	N	Overall response rate
Score=1 (highly likely)	688	41.57% (n=286)	2,418	37.10% (n=897)	3,106	38.09% (n=1,183)
Score=2	190	40.53% (n=77)	416	46.63% (n=194)	606	44.72% (n=271)
Score=3	587	29.30% (n=172)	1,196	32.78% (n=392)	1,783	31.63% (n=564)
Score=4	1,525	66.10% (n=1,008)	6,206	75.19% (n=4,666)	7,731	73.39% (n=5,674)
Score=5 (least likely)	4,335	65.54% (n=2,841)	13,020	77.25% (n=10,058)	17,355	74.32% (n=12,898)
Total	7,325		23,256		30,581	

Appendix B

Response Rates for treatment and comparison states

3.1.2 December results

In December, there were 12,265 records for the treatment states. 1,674 of the operations in these states were flagged as highly likely to be a refusal (a refusal nonresponse propensity score of 1). 988 of the operations in these states were flagged as highly likely to be an inaccessible (an inaccessible nonresponse propensity score of 1). For the comparison states, there were 60,762 records. 4,311 of the operations in these states were flagged as highly likely to be a refusal. 5,267 of the operations in these states were flagged as highly likely to be inaccessible.

Overall, 14.55 percent of the highly likely to refuse operations responded to the survey. In the treatment states, 13.26 percent of the 1,674 flagged highly likely refusal operations responded to the survey and 23.99 percent of the 988 flagged highly likely inaccessible operations responded to the survey. In the comparison states, 15.05 percent of the 4,311 flagged likely refusal operations responded to the survey and 30.89 percent of the 5,267 of the flagged likely inaccessible operations responded to the survey. See Table B7 and B8 for a direct comparison.

Table B7: December 2012 response rates for refusals within propensity groups

Refusal	N	Response Rate – Treatment	N	Response Rate – Comparison	N	Overall response rate
Score =1 (most likely to refuse)	1,674	13.26% (n=222)	4,311	15.05% (n=649)	5,985	14.55% (n=871)
Score=2	1,706	41.15% (n=702)	6,121	47.23% (n=2,891)	7,827	45.91% (n=3,593)
Score=3	1,258	50.40% (n=634)	4,874	56.91% (n=2,774)	6,132	55.58% (n=3,408)
Score=4 (least likely to refuse)	7,627	67.13% (n=5,120)	45,456	74.40% (n=33,818)	53,083	73.35% (n=38,938)
Total	12,265		60,762		73,027	

Appendix B
Response Rates for treatment and comparison states

Table B8: December 2012 response rates for inaccessible within propensity groups

Inaccessible	N	Response Rate – Treatment	N	Response Rate – Comparison	N	Overall response rate
Score=1 (most likely to be inaccessible)	988	23.99% (n=237)	5,267	30.89% (n=1,627)	6,255	29.80% (n=1,864)
Score=2	401	25.49% (n=104)	1,932	40.48% (n=782)	2,333	37.98% (n=886)
Score=3	780	24.74% (n=193)	2,626	29.82% (n=783)	3,406	28.66% (n=976)
Score=4	2,061	59.58% (n=1,228)	10,954	69.79% (n=7,645)	13,015	68.18% (n=8,873)
Score=5 (least likely to be inaccessible)	8,035	61.18% (n=4,916)	39,983	73.27% (n=29,295)	48,018	71.25% (n=34,211)
Total	12,265		60,762		73,027	

As in September, high impact operations for the purpose of the December study were those operations with strata values greater than or equal to 90 and the specialty strata (70-79). Overall, 12.28 percent of the high impact operations that were flagged as highly likely refusals responded to the survey. There were 1,093 high impact operations that were flagged as highly likely refusals in the treatment states. Of the 1,093 highly likely refusal operations, 11.99 percent responded to the survey. There were 628 high impact operations that were flagged as highly likely inaccessible in the treatment states. Of the 628 highly likely inaccessible operations, 21.66 percent responded to the survey. For the comparison states, 2,612 high impact operations were flagged as highly likely refusals. Of those 2,612 operations, 12.40 percent completed the survey. 3,104 high impact operations were flagged as highly likely inaccessible in the comparison states. Of those 3,104 operations, 30.80 percent completed the survey. See Tables B9 and B10 for direct comparisons.

Appendix B
Response Rates for treatment and comparison states

Table B9: December 2012 Response Rate for all high impact operations within propensity groups for refusals

Refusal	N	Response Rate – Treatment	N	Response Rate – Comparison	N	Overall rate
Score=1 (highly likely)	1,093	11.99% (n=131)	2,612	12.40% (n=324)	3,705	12.28% (n=455)
Score=2	1,230	41.06% (n=505)	4,204	48.33% (n=2,032)	5,434	46.68% (n=2,537)
Score=3	739	50.88% (n=376)	2,642	57.91% (n=1,530)	3,381	56.38% (n=1,906)
Score=4 (least likely)	3,418	66.68% (n=2,279)	20,669	75.72% (n=15,650)	24,087	74.43% (n=17,929)
Total	6,480		30,127		36,607	

Table B10: December 2012 response rates for all high impact operations within response propensity groups for inaccessibles

Inaccessible	N	Response Rate – Treatment	N	Response Rate – Comparison	N	Overall rate
Score=1 (highly likely)	628	21.66% (n=136)	3,104	30.80% (n=956)	3,732	29.26% (n=1,092)
Score=2	164	18.29% (n=30)	609	38.59% (n=235)	773	34.28% (n=265)
Score=3	521	22.84% (n=119)	1,592	26.63% (n=424)	2,113	25.70% (n=543)
Score=4	1,247	58.46% (n=729)	6,868	70.65% (n=4,852)	8,115	68.77% (n=5,581)
Score=5 (least likely)	3,920	58.09% (n=2,277)	17,954	72.79% (n=13,069)	21,874	70.16% (n=15,346)
Total	6,480		30,127		36,607	

Appendix C
State level predicted vs. actual response rates

Table C1: September Actual and Predicted Response Rates for the Treatment States (for highly likely to nonrespond operations)

State	Refusals			Inaccessibles		
	Predicted RR for NRP=1 (%)	Actual RR for NRP=1 (%)	Difference (Actual-Predicted) (%)	Predicted RR for NRP=1 (%)	Actual RR for NRP=1 (%)	Difference (Actual-Predicted) (%)
KS	28.55	18.12	-10.43	32.9	40.93	8.03
NE	21.9	32.15	10.25	48.55	45.56	-2.99
ND	17.5	11.61	-5.89	21.55	35.54	13.99
MI	31.3	53.01	21.71	40.05	60.87	20.82
MD	22.05	30.16	8.11	35.35	43.72	8.37
NH (NewEng)	14.3	16.67	2.37	36.9	48.65	11.75
FL	20.15	29.41	9.26	39.5	60.71	21.21
SC	33.4	25.93	-7.47	39.5	35.19	-4.31
IA	28.25	25.40	-2.85	35.25	42.59	7.34
Average	-	-	2.78	-	-	9.36

- The predicted response rate was calculated by averaging training and validation for the state models

In five out of the nine treatment states (56 percent), actual response rates were higher than the predicted nonresponse rates for refusals, suggesting that these states were able to increase response for their likeliest refusals. Typically, the difference between the predicted and actual response rates for these states was at or below ten percent. The exception is Michigan with a difference of about +22. However, four of the states (44 percent) had a decrease in response rates, all about 10 percent or less, with the average difference for refusals of 2.78.

In seven of the nine treatment states (78 percent), the actual response rates were higher than the predicted nonresponse rates for inaccessible cases, while two states (22 percent) had a decrease in predicted response rates for the inaccessibles. The average magnitude of the difference for inaccessible cases is 9.36.

Appendix C
State level predicted vs. actual response rates

Table C2: September 2012 Actual and Predicted Response Rates for Comparison States (for highly likely to nonrespond operations)

State	Refusals			Inaccessibles		
	Predicted RR for NRP=1 (%)	Actual RR for NRP=1 (%)	Difference (Actual-Predicted) (%)	Predicted RR for NRP=1 (%)	Actual RR for NRP=1 (%)	Difference (Actual-Predicted) (%)
AL	26.5	22.22	-4.28	22.22	62.07	39.85
AZ	28.1	23.08	-5.02	23.08	33.33	10.25
AR	19.2	19.54	0.34	19.54	34.50	14.96
CA	15.85	54.69	38.84	54.69	53.76	-0.93
CO	19.05	19.01	-0.04	19.01	29.29	10.28
GA	30.1	41.86	11.76	41.86	57.45	15.59
ID	41.45	8.74	-32.71	8.74	23.96	15.22
IL	36.65	23.16	-13.49	23.16	35.0	11.84
IN	19.25	25.37	6.12	25.37	51.47	26.1
KY	25.15	33.33	8.18	33.33	39.47	6.14
LA	34.65	10.64	-24.01	10.64	43.08	32.44
MN	25.7	11.11	-14.59	11.11	32.61	21.5
MS	37.9	28.13	-9.77	28.13	58.49	30.36
MO	17.85	17.18	-0.67	17.18	25.38	8.2
MT	26.9	14.29	-12.61	14.29	26.76	12.47
NV	5.75	28.57	22.82	28.57	18.18	-10.39
NJ	22.05	22.22	0.17	22.22	34.62	12.4
NM	47.45	86.36	38.91	86.36	75.93	-10.43
NY	28.2	29.17	0.97	29.17	49.41	20.24

Appendix C
State level predicted vs. actual response rates

State	Refusals			Inaccessibles		
	Predicted RR for NRP=1 (%)	Actual RR for NRP=1 (%)	Difference (Actual-Predicted) (%)	Predicted RR for NRP=1 (%)	Actual RR for NRP=1 (%)	Difference (Actual-Predicted) (%)
NC	28.5	25.81	-2.69	25.81	43.53	17.72
OH	33.35	25.0	-8.35	25.0	34.86	9.86
OK	49.3	28.68	-20.62	28.68	39.16	10.48
OR	38.8	7.69	-31.11	7.69	23.16	15.47
PA	30.2	41.03	10.83	41.03	59.04	18.01
SD	25.45	21.34	-4.11	21.34	37.97	16.63
TN	36.05	22.58	-13.47	22.58	49.23	26.65
TX	23.0	18.66	-4.34	18.66	33.56	14.9
UT	21.45	26.92	5.47	26.92	49.21	22.29
VA	17.85	18.42	0.57	18.42	33.61	15.19
WA	15.7	9.21	-6.49	9.21	18.46	9.25
WVA	55.35	33.33	-22.02	33.33	40.0	6.67
WI	28.1	6.96	-21.14	6.96	20.89	13.93
WY	52.6	8.62	-43.98	8.62	22.22	13.6
Average			-4.60			14.67

Predicted response rates were calculated by averaging training and validation for the state models

For highly likely to nonrespond operations, the model predicted a national response rate of 37.98 for refusals and a national response rate of 52.37 for inaccessibles.

Appendix C
State level predicted vs. actual response rates

Table C3: December 2012 Actual and Predicted Response Rates for the Treatment States (for highly likely to nonrespond operations)

State	Refusals			Inaccessibles		
	Predicted RR for NRP=1 (%)	Actual RR for NRP=1 (%)	Difference (Actual-Predicted) (%)	Predicted RR for NRP=1 (%)	Actual RR for NRP=1 (%)	Difference (Actual-Predicted) (%)
KS	28.55	11.14	-17.41	32.9	20.06	-12.84
NE	21.90	16.71	-5.19	48.55	23.71	-24.84
ND	17.5	12.38	-5.12	21.55	25.41	3.86
MD	22.05	8.33	-13.72	35.35	23.98	-11.37
NH (NewEng)	14.3	20.83	6.53	36.9	23.08	-13.82
FL	21.05	57.14	36.09	39.5	68.00	28.5
Average	-	-	0.20	-		-5.09

From Table C3 above, we can see that typically the actual response rate for refusals and inaccessibles are less than the predicted response rate with the exception of the New England states and Florida for refusals and North Dakota and Florida for the inaccessibles.

Table C4: December 2012 Actual and Predicted Response Rates for Comparison States (for highly likely to nonrespond operations)

State	Refusals			Inaccessibles		
	Predicted RR for NRP=1 (%)	Actual RR for NRP=1 (%)	Difference (Actual-Predicted) (%)	Predicted RR for NRP=1 (%)	Actual RR for NRP=1 (%)	Difference (Actual-Predicted) (%)
AL	26.5	20.97	-5.53	22.22	43.06	20.84
AZ	28.1	22.22	-5.88	23.08	33.33	10.25
AR	19.2	29.17	9.97	19.54	51.41	31.87
CA	15.85	16.94	1.09	54.69	33.16	-21.53
CO	19.05	8.61	-10.44	19.01	24.09	5.08

Appendix C
State level predicted vs. actual response rates

State	Refusals			Inaccessibles		
	Predicted RR for NRP=1 (%)	Actual RR for NRP=1 (%)	Difference (Actual-Predicted) (%)	Predicted RR for NRP=1 (%)	Actual RR for NRP=1 (%)	Difference (Actual-Predicted) (%)
GA	30.1	18.97	-11.13	41.86	25.61	-16.25
IA	28.25	16.57	-11.68	35.25	21.15	-14.10
ID	41.45	6.67	-34.78	8.74	20.46	11.72
IL	36.65	16.85	-19.80	23.16	31.54	8.38
IN	19.25	21.69	2.44	25.37	36.97	11.6
KY	25.15	23.18	-1.97	33.33	28.85	-4.48
LA	34.65	21.05	-13.6	10.64	53.76	43.12
MI	31.3	20.00	-11.3	40.05	6.12	-33.93
MN	25.7	9.45	-16.25	11.11	25.68	11.47
MS	37.9	8.10	-29.8	28.13	49.13	21.0
MO	17.85	17.30	-0.55	17.18	26.16	8.98
MT	26.9	15.86	-11.04	14.29	30.56	16.27
NV	5.75	11.11	5.36	28.57	0.00	-28.57
NJ	22.05	28.57	6.52	22.22	47.82	25.6
NM	47.45	36.00	-11.45	86.36	57.58	-28.78
NY	28.2	16.36	-11.84	29.17	16.35	-12.82
NC	28.5	11.36	-17.14	25.81	36.92	11.11
ND	17.5	12.38	-5.12	11.61	25.41	13.80
OH	33.35	22.92	-10.43	25.0	47.87	22.87
OK	49.3	17.55	-31.75	28.68	28.98	0.30

Appendix C
State level predicted vs. actual response rates

State	Refusals			Inaccessibles		
	Predicted RR for NRP=1 (%)	Actual RR for NRP=1 (%)	Difference (Actual-Predicted) (%)	Predicted RR for NRP=1 (%)	Actual RR for NRP=1 (%)	Difference (Actual-Predicted) (%)
OR	38.8	10.14	-28.66	7.69	19.31	11.62
PA	30.2	24.24	-5.96	41.03	47.66	6.63
SC	33.4	34.38	1.02	39.50	57.44	17.44
SD	25.45	6.79	-18.66	21.34	8.82	-12.52
TN	36.05	7.84	-28.21	22.58	51.52	28.94
TX	23.0	17.34	-5.66	18.66	32.28	13.62
UT	21.45	10.71	-10.74	26.92	15.38	-11.54
VA	17.85	11.54	-6.31	18.42	28.93	10.51
WA	15.7	12.35	-3.35	9.21	18.03	8.82
WVA	55.35	14.29	-41.06	33.33	13.64	-19.69
WI	28.1	14.29	-13.81	6.96	19.58	12.62
WY	52.6	17.91	-34.69	8.62	32.08	23.46
Average	-	-	-11.95	-	-	5.51

For highly likely to nonrespond operations, the model predicted a national response rate of 37.98 for refusals and a national response rate of 52.37 for inaccessibles.

