Using Nonresponse Propensity Scores to Improve Data Collection Methods and Reduce Nonresponse Bias

Morgan Earp[1]& Jaki McCarthy[2]
[1]National Agricultural Statistics Service (NASS), 3251 Old Lee Highway Room 305, Fairfax, VA 22030
[2]NASS, 3251 Old Lee Highway Room 305, Fairfax, VA 22030

Abstract
The USDA‘s National Agricultural Statistics Service (NASS) conducts the quarterly Crops/Stocks Survey. The Crops/Stocks Survey collects detailed data on crop acreage, yields, production, and quantities stored from selected agricultural operations. These data are used to set national and state level estimates of acres planted, harvested, production, and on-farm grain stocks. Nonresponse error is a huge concern, especially when setting estimates at the state level. Using administrative, previously reported data from other surveys and their history of response, we created nonresponse propensity scores. These propensity scores were provided to field offices, which are responsible for data collection, for each sampled operation, and field offices were asked to document any refusal/noncontact avoidance methods used. We analyzed six quarters of data to predict likely nonrespondents and establish baselines for assessing the effectiveness of state level treatments.

This paper discusses the operational use of nonresponse propensity scores (created from classification tree models) and assesses the utility of proactively employing them in data collection to reduce nonresponse rates and nonresponse bias. Overall, the use of the scores appeared to decrease the expected nonresponse. However, since field offices' use of the scores was not standardized, it was difficult to clearly evaluate their effectiveness.

This research will enable NASS to determine whether prior knowledge of nonresponse likelihood can be used to improve data collection methods and ultimately reduce nonresponse bias. However, the need for clear guidance in how to use them must be carefully considered in their operational use.

Key Words: Nonresponse; Bias; Characteristics; Classification Trees; Propensity Scores;

1. Introduction

As is the case for many surveys conducted by the Federal government and elsewhere, survey response rates have been declining and are requiring more resources to maintain. In order to reduce cost and better allocate resources, the National Agricultural Statistics Service (NASS) is interested in determining what characteristics are associated with survey nonresponse, and using that information to identify likely nonrespondents prior to data collection. Using a form of data mining known as classification trees, NASS began using auxiliary data collected from other surveys to identify likely refusals and noncontacts. Models were built to predict likely refusals and noncontacts in NASS's Quarterly Crops/Stocks September and December 2010 samples. In September, NASS field offices were provided with lists of likely nonrespondents, and were given the freedom to decide if and how they would use the scores. At the end of data collection, field offices were asked to report whether they used the scores, and if so, how they used them. After the initial field study in September, NASS compared the rate of refusals and noncontacts to previous quarters to determine which states experienced significant decreases in refusals and noncontacts. NASS also summarized and compared the most common uses of the refusal and noncontact scores. After the December 2010 survey, NASS provided field offices with a survey asking whether they used the scores and provided a list of the most common uses of the scores based on September 2010. An "other" response was made available for states to report on methods not listed on the survey. After data was collected for both September and December 2010, refusal rate and noncontact rate comparisons were made between states using the scores versus those not using the scores. Of the states using the scores, the researchers assessed and summarized data collection methods used. The methods were summarized and presented to field offices.

Background

NASS conducts hundreds of surveys each year on issues including agricultural production, economics, demographics, and environment. Every five years NASS also conducts the Census of Agriculture. One of the surveys conducted by NASS is the Crop/Stocks Survey also known as the Quarterly Agriculture Survey (QAS). The QAS provides detailed estimates of crop acreage, yields, production, and quantities of grain and oilseeds stored on farms. The QAS is conducted quarterly (March, June, September, and December), and targets producers of row crops and small grains and farms operations with grain storage capacity in all states.

NASS was interested in determining if it was possible to use auxiliary data to identify likely QAS refusals and noncontacts. NASS has an abundance of auxiliary data available for modeling nonresponse, including frame, administrative, proxy data, and their past reporting history on other NASS surveys. Frame data provides information about the operations' current operating status, farm type (grain, fruit, nursery, hog, cattle, etc.), location (state, district, county), year added to list frame, and operator's age. The Census of Agriculture includes all known agricultural operations, is mandatory, is conducted every five years, and includes collection of data on specific commodities raised, farm tenure (owner or tenant operated), government program participation, operator's race and gender, and size of the operation (total value and acres). Using zip code and county data, NASS can also determine the percent of county in farmland, county population, how rural or urban the county is, the percent of the county population that is foreign born, and education level within a county based on the Population Census.

Problem

By using a series of simple univariate comparisons, we can see that a number of variables are related to QAS refusal and noncontact (Table One). The problem with large datasets is many variables are significantly different between cooperators, refusals and noncontacts just due to the number of observations; however, these differences are small in practical terms. Furthermore, many of these variables are correlated, and will result in variance inflation and multicollinearity issues if they are all used in the same model. Using Table One alone is not useful for predicting individual establishment nonresponse or managing data collection. Therefore, classification tree models were developed to identify characteristics associated with as well as predicting nonresponse in future samples (McCarthy and Jacob, 2009, McCarthy, Jacob and McCracken, 2010). These classification tree models used the auxiliary data known for each operation to segment the dataset (in this case previous samples of the QAS) into mutually exclusive groups. A

CHAID type approach was used to build the models where the dataset was split using a series of simple rules. In this approach, a classification tree model is constructed by segmenting a dataset using a series of simple rules. Each rule assigns an observation to a segment based on the value of one input variable. One rule is applied after another, resulting in a hierarchy of segments within segments. The rules are chosen to maximally separate the sub-segments with respect to the target variable (in this case either refusal or noncontact). A node with all its successors is termed a branch of the node that created it. The final nodes are called leaves. In our analysis, we are interested in the leaves that contain a higher proportion of records with respect to the target variable. We created separate models to predict survey refusals and noncontacts as the target. Each leaf has a probability of nonresponse which can be ranked, ordering subgroups with respect to their likelihood of being nonrespondents. Details on how these models were constructed are discussed in McCarthy, Jacob and McCracken (2010).

Currently there are other studies underway using decision trees to identify nonrespondents (Phipps & Toth, 2011; Toth & Phipps, 2011); however, these studies focus on using propensity scores to calculate nonresponse adjustment weights as opposed to guiding data collection. While it is difficult to conduct a true scientific experiment in the field, we attempt in this report to compare the effectiveness of using propensity scores to target likely refusals and noncontacts in the hopes of improving data quality during the data collection phase.

Table One: Differences on Selected Variables between Survey Cooperators, Refusals and Noncontacts

Measure	Cooperator	Refusal	Noncontact
Farm Type: Grains, Oilseeds, Dry Beans and Dry Peas	37%	51%	39%
Farm type: Cattle and Calves	24%	18%	20%
USDA Conservation Reserve Program reported	18%	23%	18%
Soybeans reported	73%	85%	76%
Hours worked at an off farm job= 0	65%	72%	64%
Operator lives on operation	82%	84%	79%
Major occupation is farming	86%	91%	86%
Average % of the R’s county in farmland	63.00	73.73	64.97
Average % of R’s county population foreign born	3.5	3.07	3.86
Average number of people/sq. mile in the R’s county	95.09	68.12	94.08
2003 Rural Urban continuum Code = 8 or 9	19%	26%	18%

Purpose

Nonresponse scores for each sampled operation in the September and December 2010 QAS were provided for the NASS field offices (who are responsible for data collection.) Each field office decided whether to use these scores, and if they used them, how to incorporate this into their data collection. Field offices were asked to report whether they used the scores, and if so, how they used them. The purpose of this research is to determine the effectiveness of using the nonresponse indicators, as well as best practices for using them to reduce refusals and noncontacts.

2. Method

Research Design

A causal-comparative research design was used to compare field offices that reported using the nonresponse scores with field offices that did not report using the scores. Field offices using the scores were considered the treatment group, and those not using the scores were considered the comparison group. Each was assessed to determine if the amount of refusals or noncontacts significantly decreased in September and December compared to the model. After, comparisons were completed within field offices, the proportion experiencing significant decreases in refusals and noncontacts was compared across the treatment and control group to determine if offices that reported using the scores were more likely to see significant decreases in likeliest refusals and/or noncontacts.

Participants

Twenty-Nine field offices volunteered to use the scores (Table Two). Fifteen decided not to use the scores prior to data collection. New England decided not to use to the scores in September, but decided to use the scores in December.

Table Two: Comparison and Treatment States

Comparison States	Treatment States
Arkansas	Alabama
Colorado	Arizona
Florida	California
Georgia	Delaware
Indiana	Idaho
Kentucky	Illinois
Missouri	Iowa
Nebraska	Kansas
Nevada	Louisiana
New England[1]	Maryland
New Jersey	Michigan
New Mexico	Minnesota
North Carolina	Mississippi
Oregon	Montana
Tennessee	New England[1]
	New York
	North Dakota
	Ohio
	Oklahoma
	Pennsylvania
	South Carolina
	South Dakota
	Texas
	Utah
	Virginia
	Washington
	West Virginia
	Wisconsin
	Wyoming

[1] New England did not use the scores in September, but did in December.

Procedures

Likeliest refusals and noncontacts were identified using the decision trees built by McCarthy and Thomas (2009). The resulting decision tree used to identify likeliest refusals identified four subgroups; Group One being the most likely to refuse (likeliest refusals), and Group Four being the least likely to refuse (Figure One). The resulting decision tree used to identify likeliest noncontacts identified five subgroups; Group One being the most likely to be noncontacts (likeliest noncontacts), Group Five being the least likely to be a noncontact (Figure Two). Shown in Figure One is the model developed using operations in all states. This model was built using the 2006 and 2007 March, September, and December QAS data. The results of this model were applied to each field office separately to generate predicted probabilities of highest refusal and noncontact rates for each field office.

The criterion laid out in these trees was then used to assign the September 2010 and the December 2010 QAS samples to refusal and noncontact propensity groups. This research focuses on the likeliest refusals and likeliest noncontacts (group one in each tree). For refusals, as shown in the tree, this is defined as those sampled operations that had two or more survey refusals in the past three years, and within that group, also had fewer than two good survey completions in the last three years. As shown in Figure Two, the highest probability noncontacts were those that have 1 or more survey noncontacts in the past two years, an individual response rate of less than 25.5 percent on NASS surveys in the past two years and within that group, had 3 or more noncontacts on NASS surveys in the past three years.

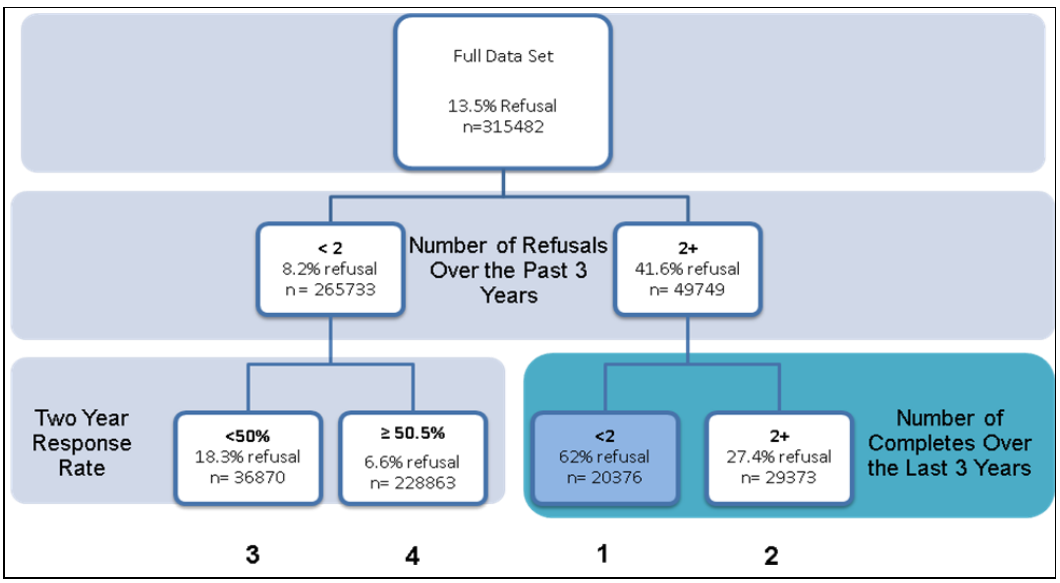

Figure One: Classification Tree Model for Refusals

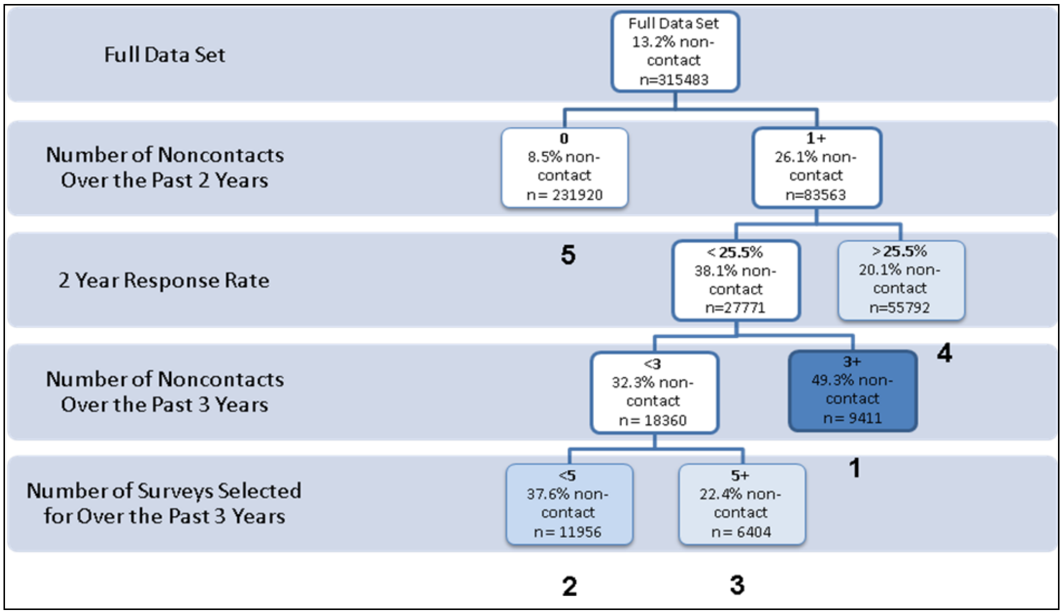

Figure Two: Classification Tree Model for Noncontacts

Data Analysis

Within states, we first compared the rate of likeliest group one refusals for September to the model and then December to the model using the Fisher Exact Test. We then compared the proportion of states exhibiting significant decreases in likeliest refusals or noncontacts both in September and December in the treatment group to the control group using a z test for proportion.

3. Results

Using a Fisher Exact Chi Square Test, we tested whether states experienced a significant decrease in the number of likeliest refusals and/or noncontacts. We were specifically interested in states that saw significant reductions in both September and December, since this would indicate the decrease was reliable over time; therefore, the results for New England are not reported here.

When we look at the percent of refusals and noncontacts overall, not accounting for state, we see that the treatment group had significantly less refusals in September (x =197.05 , df =1, p < .05) and December (x =52.91 , df =1, p < .05) and significantly less noncontacts in September (x = 221.55, df =1, p < .05) and December (x = 73.05, df =1, p < .05) according to the Fisher Exact Test; however, comparisons should be done at the state level, since that is the level at which treatments were administered.

When we assessed the results at the state level, we found that 21 percent (3/14) of the comparison group saw significant decreases in the number of likeliest refusals both in September and December, compared to 29 percent (8/28) of the treatment group (Table Three). While this difference was not significant according to the z test for a difference in two proportions (z =.50, p = .31), the rate of significant decreases in likeliest refusals was higher for the states using the refusal indicator (Table Three).

Fourteen percent (2/14) of the comparison group saw significant decreases in the number of likeliest noncontacts both in September and December, compared to 32 percent (9/28) of the treatment group (Table Three). Again, while this difference was not significant according to the z test for a difference in two proportions ($z = 1.24$, $p = .11$), the rate of significant decreases in likeliest noncontacts was higher for the states using the noncontact indicator (Table Three).

Because all of the states using the scores did not do the same thing, we wanted to look at what the states which did see significant reductions in their likeliest nonrespondents had done. The treatment states exhibiting significant decreases in likeliest refusals are shown in Figure Three, and Figure Four shows those with significant decreases in likeliest noncontacts. We assessed what methods they reported using to decrease nonresponse cases. In the case of likeliest refusals, field offices with significantly lower than predicted refusal rates reported assigning these operations to personal interview, assigning specific enumerators to collect the data, reviewing list frame comments[2], and one field office reported not attempting to collect data from these operations, instead holding them out of data collection. In the case of likeliest noncontact, field offices reported assigning these operations to personal interview, assigning specific enumerators to collect the data, reviewing list frame comments, and updating phone numbers.

Table Three: States Exhibiting Significant Reductions in Refusals and Noncontacts

Likeliest Refusals		Likeliest Noncontacts	
Comparison States	**Treatment States**	**Comparison States**	**Treatment States**
Arkansas*	Alabama	Arkansas	Alabama*
Colorado	Arizona	Colorado	Arizona
Florida*	California*	Florida	California
Georgia*	Delaware	Georgia	Delaware
Indiana	Idaho	Indiana	Idaho*
Kentucky	Illinois*	Kentucky	Illinois
Missouri	Iowa	Missouri	Iowa*
Nebraska	Kansas	Nebraska	Kansas*
Nevada	Louisiana*	Nevada	Louisiana
New Jersey	Maryland	New Jersey	Maryland
New Mexico	Michigan	New Mexico*	Michigan*
North Carolina	Minnesota	North Carolina	Minnesota
Oregon	Mississippi	Oregon*	Mississippi*
Tennessee	Montana	Tennessee	Montana
	New York		New York
	North Dakota*		North Dakota
	Ohio*		Ohio
	Oklahoma*		Oklahoma*
	Pennsylvania		Pennsylvania*
	South Carolina		South Carolina
	South Dakota*		South Dakota*
	Texas*		Texas
	Utah		Utah
	Virginia		Virginia
	Washington		Washington
	West Virginia		West Virginia
	Wisconsin		Wisconsin
	Wyoming		Wyoming

*Significant decrease in number of refusals/noncontacts compared to the model prediction for the state at the .05 level

[2] List frame comments are information kept on the list frame for that operation. Field office staff may enter anything into these comments, but will often include information on best methods to contact the operation, any special data collection handling, notes on previous contacts with the operation, etc.

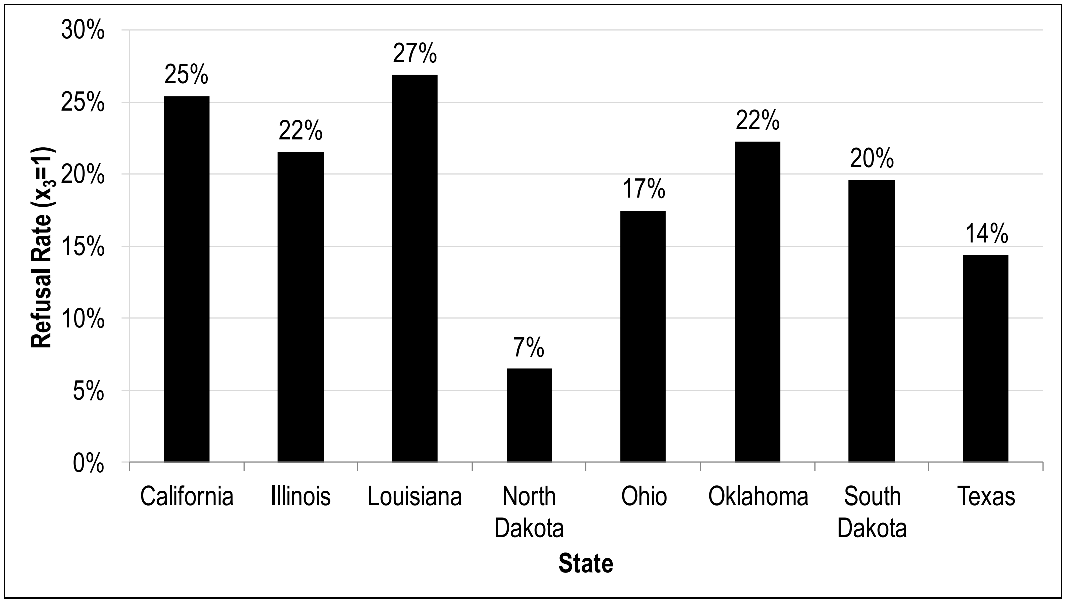

Figure Three: Mean Percent Decrease in September-December 2010 Likely Refusals Compared to the Model

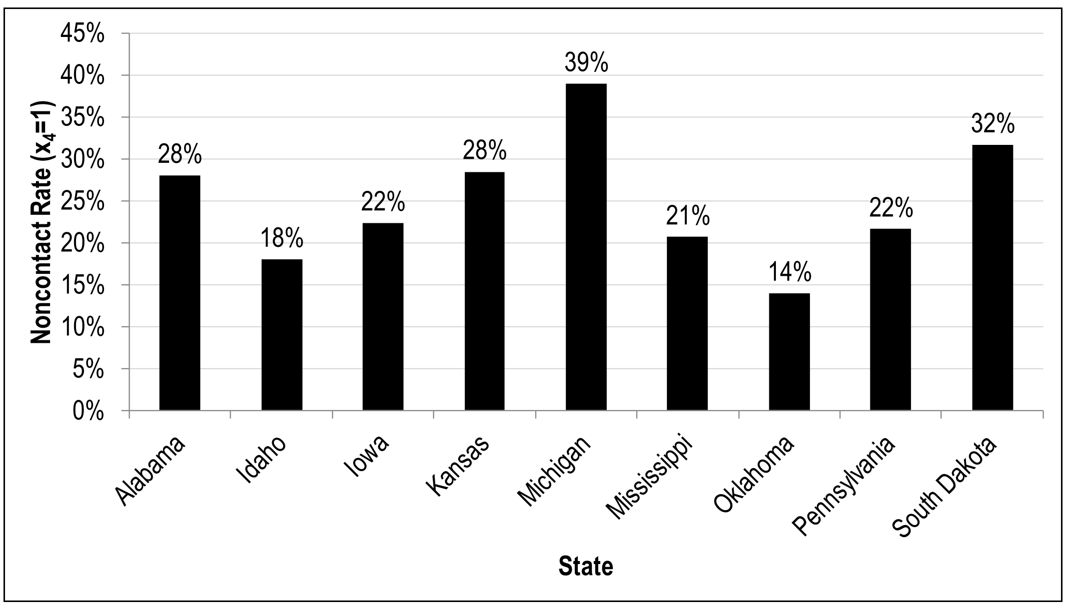

Figure Four: Mean Percent Decrease in September-December 2010 Likeliest Noncontacts Compared to the Model

4. Discussion

As is the case for many organizations that conduct periodic surveys, NASS has numerous sources of auxiliary data available for sample units. This information, in particular, the history of response in previous NASS surveys can be used to identify those agricultural operations most likely to refuse to participate or to never be contacted in the survey. Over the past several years, NASS has developed classification tree models to predict the likeliest nonrespondents. Once the most likely nonrespondents have been identified, the challenge is to decide how to use this information in field data collection. Our initial work with these nonresponse propensity scores relied on the field office staff to determine the best use of these scores. Our initial evaluation suggests that the scores can be helpful in reducing nonresponse. However, each field office chose a different approach to their use. Some offices assigned all of these cases to in person interviews. Other offices assigned these to specific interviewers. Still others assigned these cases to subject matter analysts for data collection. And in one case, the field office concluded that the effort required to gain cooperation from these cases was ineffective and did not attempt to collect data from them (clearly reducing their refusal rate in this group, but certainly not increasing their response rate.)

Although these offices all used the nonresponse models, it difficult to determine if "using" these scores was effective. We attempted to look at the methods used by the states whose nonresponse rates in the most likely nonrespondent groups were reduced compared to what the model predicted. These will form the basis for more specific instruction for how to use these scores in the future.

Going forward, there are a number of activities that will follow this evaluation. More specific guidance on how to use these scores should be developed. This will help to ensure that they are being used consistently across field offices. Consideration will be given to those strategies that appeared most effective, but in addition costs and the potential impact on the survey estimates should also be considered. While assigning all of the likeliest nonresponse cases to in person interviews may decrease nonresponse for this group, this will likely also increase data collection costs. In particular for NASS, long distances and additional time may be required to reach individual agricultural operations.

In addition, because this is a survey of establishments, some sample units may be contributing much more to the estimates of population totals. For example, it may be more important to obtain a response from someone farming 10,000 acres of corn than from someone growing only 10 acres. Additional work to see how these highly likely nonrespondents can impact the survey estimate and which are most likely to introduce nonresponse bias remains to be done. Work on another survey at NASS using a similar classification tree nonresponse model is also underway, including efforts to *both* identify likely nonrespondents and those with the biggest potential impact on nonresponse bias of key survey estimates (EARP REF here). Perhaps higher costs data collection methods should be directed only at a subsample of the groups we targeted.

Classification tree models to predict likely nonrespondents are an important, but only initial, step in the process of reducing nonresponse and ultimately improving the survey estimate. However, moving these models into field operations can be done in many ways, and knowing whether each particular use is effective is difficult to evaluate. True experimental comparisons of methods in the field are often difficult or impossible, so more ad hoc methods of evaluation may be useful. We will continue to use these models and hope that they can improve both the quality and efficiency of our surveys.

5. References

McCarthy, J.S. & Jacob, T. (2009). Who are You?: A Data Mining Approach to Predicting Survey Non-respondents. In *JSM Proceedings*, American Association for Public Opinion Research. Hollywood, Florida: American Statistical Association.

McCarthy, J.S., Jacob, T. and 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. (2011). Analyzing Non-Response and Non-Response Bias in the Occupational Employment Statistics Survey Using Regression Trees. In *JSM Proceedings*, American Association for Public Opinion Research. Phoenix, Arizona: American Statistical Association.

Toth, D. & Phipps, P. (2011). Exploration of Data Quality in Work with Administrative Records and Sample Surveys. In *JSM Proceedings*, American Statistical Association. Miami, Florida: American Statistical Association.