Using Propensity Scores to Inform Respondent Incentive Escalation

Kate Bachtell[1], Catherine Haggerty[1], Becki Curtis[1], Shannon Nelson[1], Joanne Hsu[2], Jesse Bricker[2], Kevin Moore[2]
[1]NORC at the University of Chicago, 55 E. Monroe St., Chicago, IL 60603
[2]Board of Governors of the Federal Reserve System, 20th and C Streets NW, Washington, DC 20551

Abstract

The triennial Survey of Consumer Finances (SCF) has maintained response rates over time while collecting highly sensitive financial information. However, gaining cooperation from SCF sample members has presented challenges that often necessitate extending the data collection period. A new strategy for the 2016 SCF involves developing an "escalation need" score using aggregated, real-time case management data – including indicators of gated communities, locked buildings, and the number of past contact attempts – along with a measure of low response propensity from the Census Planning Database. In this paper we first share our methodology for developing the escalation need score – that is, the degree to which the household should be considered for incentive escalation based on factors historically associated with tough-to-reach households. Next, we describe how we are using this score to inform strategies for interviewing and offering respondent incentives in the 2016 SCF. Finally, we present early results from the first eleven weeks of data collection. We close with a brief discussion for fellow survey researchers, practitioners, and other stakeholders challenged with the careful orchestration of field data collection in an era of declining response rates for U.S. household surveys.

Key Words: Incentives, survey response propensity

1. Background

To meet high data quality standards amidst a national trend of declining participation in U.S. household surveys, many research protocols must rely on costly data collection methodologies. Funding agencies and research organizations employing in-person, interviewer-assisted data collection are particularly motivated to find efficiencies in their research. One method of improving efficiency is to apply response propensity modeling to inform differential treatment to potential survey respondents in the form of non-contingent monetary incentives. Propensity modeling is an empirical process by which a multivariate statistical model is developed to predict the likelihood that an entity in a sample will participate in a survey. The statistical model may then be used to allocate differential incentives (Lavrakas et al 2016).

Using response propensity modeling to inform the use of incentives is a relatively new avenue of pursuit in survey research. The standard approach to the use of monetary incentives has been a one-size fits all approach, even though respondents vary in their degree of interest in and resistance to survey participation. This approach may exacerbate differential response and nonresponse bias if those inclined to participate without any incentive participate at a greater rate with an incentive, and if those participants that respond in even greater numbers are different from others (Lavrakas et al 2016).

Using data from the 2014 National Household Education Survey (NHES), Jackson et al (2016) conducted research to determine whether data available *prior to data collection*—from the sampling frame and/or from publicly available sources—can be used to identify cohorts of households expected to show lower- and higher-than-average response rates; and whether the use of this information to allocate differential incentive amounts can increase overall response rates, reduce nonresponse bias, and/or reduce costs. Data from the first phase of the project suggest that a simple logistic regression model, employing data on the address-based sampling frame and from publicly available sources, can successfully differentiate between likelier-than-average responders (requiring smaller incentives) and less-likely-than-average responders (requiring higher incentives) (Jackson et al 2016).

Preliminary data from the second phase of the aforementioned project show that using a tailored approach to assign incentives to groups with different response propensities resulted in slightly lower responses from the groups receiving no incentive or a $2 incentive when compared to those that received the $5 treatment. The group receiving the $10 incentive had a slightly higher response rate when compared to the $5 treatment group. These early findings support the use of the tailored approach; however, the authors had not completed data collection and some additional aspects of the research are outstanding (McPhee et al 2016).

1.1 Respondent Incentives for the Survey of Consumer Finances (SCF)

The Survey of Consumer Finances (SCF) is a triennial study funded by the Board of Governors of the Federal Reserve System (FRB). It employs a dual-frame sample, including an area probability sample of U.S. households and a list sample that supplies an oversample of wealthy Americans. The SCF is regarded as a premier source of data on U.S. household finances. NORC at the University of Chicago (NORC) has conducted the SCF on behalf of the FRB since 1992.

In late 2014 NORC partnered with the FRB to conduct an experiment to determine if offering larger monetary incentives would encourage early participation among households in higher-income Census tracts within Miami, Los Angeles and New York. These cities were selected because of their large concentrations of wealthy households, which have historically proven slow to respond to the SCF. Three hundred randomly selected addresses in each city were assigned to one of three incentive groups and received an initial offer of $50, $100, and $150. Each of the three initial incentive groups was divided into two second phase treatments; for some, there was no increase to the initial incentive offer. For the others, the incentive offer increased to $75, $150, or $250 (depending on the initial amount) if a household initially refused to participate. We found that by offering increased incentives, we reached a higher percentages of male, college-educated, wealthy respondents during a short period. We took this as evidence supporting the use of higher incentives on the 2016 SCF as a means to shorten the data collection period.

For the 2016 SCF, it was decided that all households in the area probability sample and the lower three income strata of the list sample would receive an initial offer of $75 in an advance letter, along with a $5 cash prepaid incentive. In the second quarter of data collection, we began offering increased incentives to households that refused or have been otherwise nonresponsive, using a set of statistical models described in the following sections.

2. Current Study

For the 2016 SCF, researchers at NORC and the FRB are exploring the use of propensity modeling to inform respondent incentive escalation. In this paper we describe our methods for a) developing an "increase score" measuring the need for a given household to be offered an increased incentive, b) incorporating the increase score in an algorithm – that is, a predetermined set of rules in a problem-solving operation – to identify households that present the most challenges for data collection, and c) offering escalated incentives to a selected subset of households.

During the summer of 2016 we drew on data from two sources: real-time, household-level case management data from the first few months of data collection for the 2016 SCF and the publically-available Census Planning Database. We used these data to develop the initial increase score. That score became the main ingredient in an algorithm that allowed us to quickly identify households to receive an escalated incentive for the 2016 SCF. Case selections are communicated to field interviewers in the case management system and increased incentive offers are made via several modes: verbally and/or in-person by the interviewer, or in a letter. We are currently monitoring call histories to evaluate the real-time efficacy of the escalated incentives. At the close of the 2016 data collection period, we will assess the overall utility of the increase score in predicting the probability of interview completion.

2.1 Research Questions

In this paper we investigate three research questions.

R1a: How well does response propensity at the tract/block group level for another survey help predict response propensity for the 2013 SCF?

In the design and early implementation stages of any new survey it is useful to draw on the experiences of other surveys completed with a comparable population. This assumes that there are characteristics of households that are clustered geographically and which may influence a given household's probability of participating in a survey. In our case, the 'other' survey is the American Community Survey and 2010 Census. We will test whether response propensity for the ACS and 2010 Census may be transferable to the 2013 SCF.

R1b: How well does case management data from the SCF help predict response propensity for the 2013 SCF?

Because NORC has conducted the SCF triennially for the FRB since 1992, a wealth of project-specific data are available to expand upon estimates derived from other surveys like the ACS and 2010 Census. We will evaluate the extent to which case management data from the 2013 SCF predict survey completion for the 2013 SCF.

R2: How can the influential factors revealed through R1a and R1b be used to identify households that should be considered for incentive escalation early in the 2016 SCF field period?

In addition to the published research and 2014 SCF experiment summarized in the Background section above, past experience conducting the SCF has shown that increased incentive offers play an effective role (though certainly not the only role!) in gaining cooperation from respondents and converting refusals. We aim to leverage the findings regarding response propensity from our first and second research to questions in isolating households that are best candidates to receive an escalated incentive. Our hope is that doing so will increase the probability of completing data collection for the 2016 SCF within the calendar year.

3. Data and Methodology

We have integrated data from three sources for this research. First, we drew from the Census Planning Database, a publically-available compilation of housing, demographic, socioeconomic, and census operational data. These data are used to inform survey and census planning. The database includes select decennial Census and select American Community Survey (ACS) estimates at the tract and block level. Of primary interest to us is the variable low_response_score in the 2015 Planning Database Block Group Data, which predicts the likelihood that a given Census block group will produce a low mail return rate. Higher scores indicate increased potential for enumeration challenges.

Our second and third data sources come from the 2013 and 2016 SCF, respectively. These data are available only to authorized members of the SCF research team at the FRB and NORC. From the 2013 SCF we use household-level case management data that aggregate the outcomes of each contact attempt. Included are measures such as the total number of unsuccessful contacts, total number of contacts with and without refusals, the interviewer's assessment of the probability of survey completion, and the final survey outcome. From the 2016, we again incorporate household-level case management data; however, as the survey is in progress at the time of this writing, we base our analysis on the first 11 weeks of data collection for the 2016 SCF.

To investigate our first and second research questions regarding the efficacy of response propensity for another survey and case management data from the 2013 SCF in predicting 2013 SCF survey completion, we used a combination of geographic mapping and multivariate statistical techniques. These are described further in the Findings section. To investigate our second research question, we developed an algorithm – that is, a step-by-step process for solving a problem. In this case our problem is the identification of households that are likely to present challenges for survey completion. Our aim is to offer these households increased monetary incentives early in the 2016 SCF field period.

Beginning in week 9 of the 2016 SCF data collection period, a universe of 5,731 cases were considered for incentive escalation. During week 11, we excluded cases that were a) complete, b) finalized as being out of scope or a final refusal, c) had been first contacted within fewer than 30 days, and/or d) belonged to the upper four wealth strata of the list sample. We ended up applying the algorithm to 3,982 cases.

4. Findings

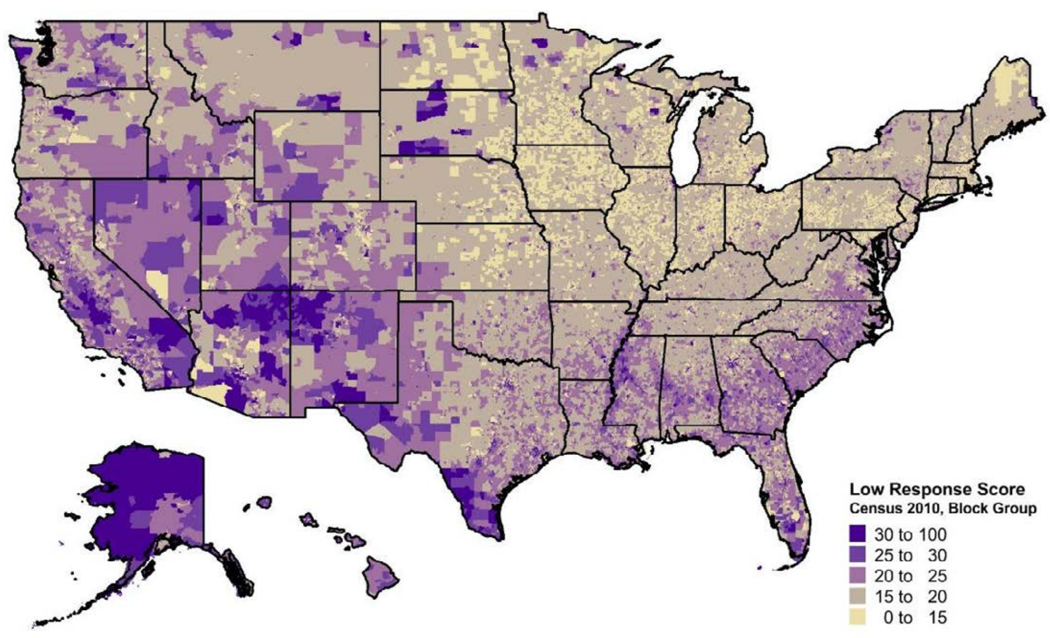
In this section we present key findings for each research question.

R1a: How well does response propensity at the tract/block group level for another survey (in this case, the ACS/2010 Census) help predict response propensity for the 2013 SCF?

As an initial exploratory step, we wanted to see which parts of the country have been problematic for the U.S. Census Bureau when conducting the decennial Census and American Community Survey. We hypothesized that there would be overlap between the Census' trouble spots and those encountered for the SCF.

Figure 1 contains a map of the Census' Low_response_score by block group across the U.S. Again, this score predicts the probability that a given block group will yield a low mail return rate, with higher scores indicating increased enumeration challenges. We are using darker shading to indicate areas that are predicted by the Census low response score to be more challenging for survey response. We find that like the Census, NORC has historically encountered difficulty reaching respondents in coastal regions of southern California and Florida – places with both pockets of extreme wealth and non-primary residences (i.e. second homes) – when conducting the SCF.

Figure 1. Census Low Response Score by Block Group (2010)



Data Source: Census 2010. Included in Census 2015 Planning Database Block Group Data.

We later removed low_response_grp from the initial increase score to test the null hypothesis that there is no relationship between low_response_score and incr1_score (see the findings for R2 below for more information about the components of incr1_score). The null hypothesis was rejected. After removing low_response_score from the formula to create the initial increase score, there remained a positive relationship between low_response_score and the adjusted increase score, *incr1a_score* (Pearson's r=.07, P<.001). However, it was considerably weaker than the relationship between low_response_score and incr1_score (Pearson's r=.39, P<.001). This suggests that there is some alignment between the geographic areas identified by the Census as having low response propensity and our working algorithm identifying cases for incentive escalation. The alignment is magnified, expectedly, if we include low_response_score as a moderately-weighted independent variable in the increase score formula. In summary, we found that the Census low response score did provide explanatory information above and beyond our other independent variables.

R1b: How well does case management data from the SCF help predict response propensity for the 2013 SCF?

Would the predictive power of the Census low response score hold up if we controlled for other factors? We constructed a series of multivariate logistic regression models to isolate the unique impacts of various contact history measures on the probability of interview completion for the 2013 SCF, with 1 meaning completed and 0 meaning not completed. We included all households in the 2013 SCF area probability sample that were still eligible to be interviewed in week 11 of the 2013 field period.

Among the independent variables, we constructed measures indicating the number of each of the following as of week 11 in the data collection period: unsuccessful contacts, contacts with refusals, and contacts without refusals. Other measures captured the total number of contacts after week 16, the interviewer's assessment of the probability of interview completion (adjusted so that 1=low, 2=mid, and 3=high), and whether the household received an increased incentive (0=no, 1=yes). Finally, we included a quantile adjustment of the Census low response score.

The regressions revealed three key findings. First, escalation made low-likelihood respondents (based on the interviewer's assessment) more likely to complete the survey. Second, escalations had a greater effect on getting completions from respondents living in areas with high Census low-response-scores – that is, in tough areas. Finally, refusals delayed and/or prevented completion, but escalations helped offset this.

R2: How can the R1 influential factors be used to identify households that should be considered for incentive escalation early in the 2016 SCF field period?

Building on the findings from our first and second research question, we developed a formula to construct a score summarizing each household's "escalation need;" that is, the degree to which the household should be considered for incentive escalation. The components of the initial increase score are listed below.

> *Num_refs*: This is a count of refusals as of week 11.
>
> *Noluck*: This variable counts the number of unsuccessful contact attempts through week 11.
>
> *Barrier*: This dummy variable indicates whether an interviewer has encountered a gated community, locked building, language barrier, etc. while trying to make contact with a given household, with 0=no and 1=yes.

Phase2: This dummy variable indicates whether the household has entered phase 2 of the SCF contacting strategy (0=no, 1=yes). In this phase the interviewer focuses on connecting with the respondent and using gaining cooperation techniques to persuade him/her to participate. Having reached this phase is taken as evidence that efforts to introduce the study to the sampled household have already been made.

Four_attempts: Another dummy variable, this indicates whether at least four call entries have been recorded (0=no, 1=yes).

low_response_grp: This is a quartile ranking of the Census Low Response Score, with values ranging from 0 to 3. The higher the value, the more likely to present challenges for survey data collection.

FI_transfer: This dummy variable indicates whether call entries have been entered by more than one interviewer (0=no, 1=yes).

Stagnant: If, for a given household, more than 50 days have elapsed since the last record of call was entered, this dummy variable will be set to 1. Otherwise it is set to 0.

FI_rpt_tough: This dummy variable indicates whether, at the last applicable contact attempt, the interviewer assessed the household as being tough/unlikely to complete (0=no, 1=yes).

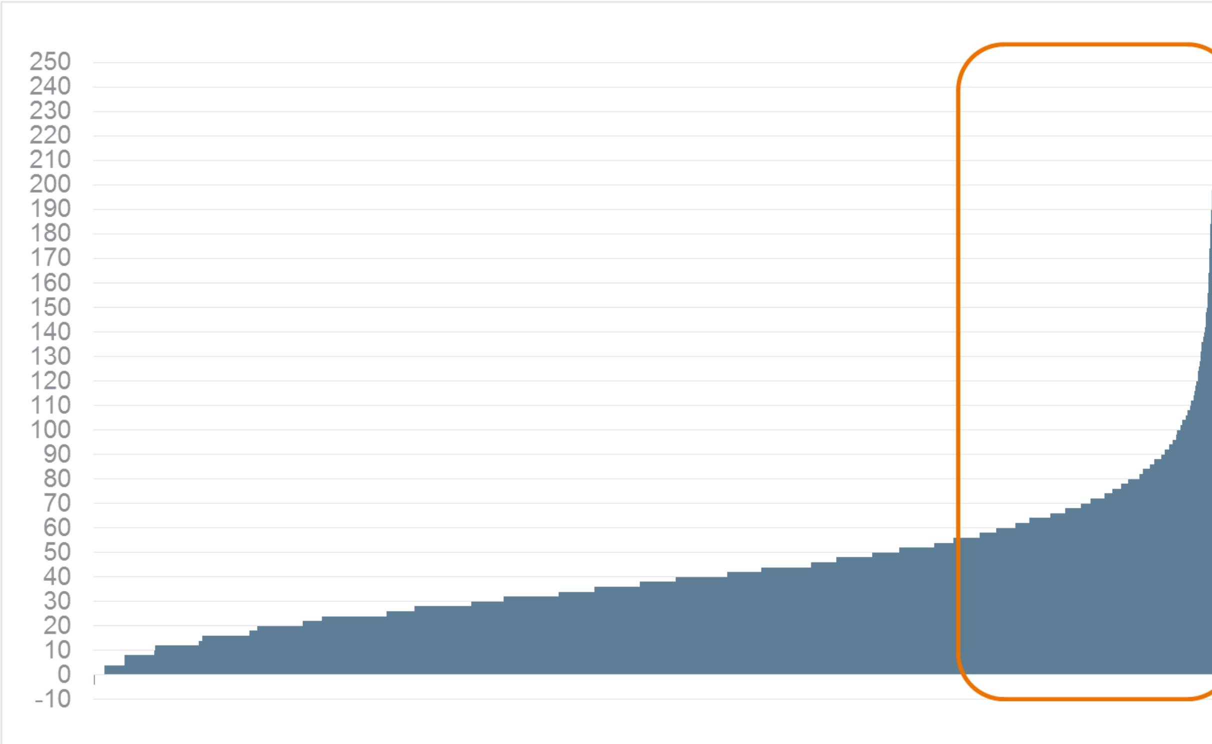
As stated earlier, our dependent variable can be interpreted as the "escalation need score" for the 2016 SCF. The variable name is *incr1_score*. It is the sum of each independent variable multiplied by a weight ranging from -4 to 10. We also constructed incr1_score_grp, a quartile ranking of the escalation need score (incr1_score), with values ranging from 0 to 3. Values equal to 3 indicate cases in the top quartile (75th percent) – that is, with the highest estimated escalation need. The formula for calculating this score is as follows:

incr1_score = (num_refs*10) + (noluck*8) + (barrier*8) + (phase2*4) + (four_attempts*4) +(low_response_grp*4) + (fi_transfer*-4) + (stagnant*4) + (fi_rpt_tough*8)

An important limitation should be acknowledged here. We know that there is multicollinearity (when two or more of the predictors in a regression model are moderately or highly correlated) between these variables. We accept this noise now because our primary interest is in identifying households with any combination of these characteristics and quickly offering them an increased incentive. We plan to refine this formula later in the 2016 field period, as we will need a more streamlined model for predicting survey response moving forward.

Figure 2 displays the distribution of the initial increase score – that is, our approximation of each household's need for an escalated incentive – among eligible households as of week 11. Values range from -4 (lowest need for escalation) to 248 (highest need for escalation).

Figure 2. Distribution of Initial Increase Score among Eligible Households as of Week 11 (2016)*



*Results are unweighted

After examining the distribution of the initial increase score, we decided to integrate a quasi-experimental design in the allocation of escalated incentives. The goal was to identify the households that had exhibited the greatest challenges for survey completion and quickly offer them an increased incentive. The quasi-experimental design is shown in Figure 3. 80% of 2016 SCF sampled households were randomly assigned to a Treatment group. From this group, we selected those households whose initial increase score fell in the top quartile. Interviewers were universally instructed to offer an escalated incentive of $150 to these households. A second group was classified as a control group and included 10% of sampled households. Households in this group were deemed ineligible for an increased incentive (though many would have received the base $75 incentive offered to all households in the area probability sample and the lower three income strata of the list sample). The final group, comprising 10% of the total sample, was classified as the “interviewer’s discretion” group (labeled in Figure 3 as “Control2”). All of these households were deemed eligible to receive an increased incentive; however, interviewers were instructed to work with their field managers to determine whether, on a case-by-case basis, a given household should be extended such an offer. If so, they were asked to use their discretion in determining how much should be offered (with a maximum of $150).

Figure 3. Initial 2016 SCF Incentive Escalation Plan

<table>
<tr><td>GROUP</td><td colspan="4">Treatment</td><td colspan="4">Control1</td><td colspan="4">Control2</td></tr>
<tr><td>INCENTIVE INCREASE NOTE IN CMS</td><td colspan="4">Increase incentive offer to $150</td><td colspan="4">Do not increase incentive offer</td><td colspan="4">Discuss incentive with your FM</td></tr>
<tr><td>INCENTIVE ACTION TAKEN BY FI</td><td colspan="2">Offered</td><td colspan="2">Not offered</td><td colspan="2">Offered*</td><td colspan="2">Not offered</td><td colspan="2">Offered</td><td colspan="2">Not offered</td></tr>
<tr><td>INTERVIEW OUTCOME</td><td>Complete</td><td>Not complete</td><td>Complete</td><td>Not complete</td><td>Complete</td><td>Not complete</td><td>Complete</td><td>Not complete</td><td>Complete</td><td>Not complete</td><td>Complete</td><td>Not complete</td></tr>
<tr><td>PERCENTAGE OF SAMPLE</td><td colspan="4">80%</td><td colspan="4">10%</td><td colspan="4">10%</td></tr>
</table>

*Should expect a small number due to human error and/or exceptional circumstances

For an initial incentive increase in July of 2016, we identified 1,010 households within the Treatment Group to receive an offer of $150. 557 households fell into the control group, and 504 were allocated to the interviewer discretion group. Incentive instructions were communicated to interviewers within the case management system. At the time of this writing, we are closely monitoring the potential impact of the increased incentives relative to the $75 base offer (and also relative to the absence of a monetary incentive) on the likelihood of survey participation for the 2016 SCF.

Discussion

The coming weeks will reveal whether the allocation of differential incentive amounts is effective in reducing data collection time and reducing nonresponse bias for the 2016 SCF. Early results have provided some support for expanding the use of escalated incentives to additional households, including those in the upper four wealth strata of the list sample. Even modest gains in our ability to predict the probability of survey completion and use this information to direct incentive allocation could improve the efficiency of data collection. In an era of declining response rates across surveys of U.S. households, this is a promising endeavor.

Our next step in this research is to revise the algorithm used to identify cases for incentive escalation based on the evolving realities of field data collection. Variables, weights, and inclusion criteria may be amended to achieve data collection goals and/or reflect new information about the challenges precluding survey completion. Subsequent incentive escalations are planned for the fall. At the close of the 2016 data collection period, we will assess the efficacy of incentive escalation on the overall 2016 response rate.

Disclaimer

The analysis, discussion, and conclusions set forth in this paper are those of the authors and do not indicate concurrence by colleagues, the Board of Governors of the Federal Reserve System, or NORC at the University of Chicago.

References

Haggerty, C., Nelson, S., Bachtell, K., Pedlow, S., Curtis, B., Joyce, A., Hsu, J., Schmeiser, M., and Moore, K. 2015. "Encouraging Early Participation in a Lengthy Survey That Collects Sensitive Personal Data: Do Large Monetary Incentives Make a Difference?" *2015 Proceedings of the American Statistical Association, Survey Research Methods Section* [CD ROM], Alexandria, VA: American Statistical Association.

Hsu, J., Schmeiser, M.D., Haggerty, C., and Nelson, S. 2015. "The Effect of Large Monetary Incentives on Survey Completion: Evidence from a Randomized Experiment with the Survey of Consumer Finances." Presented at the 70th annual meeting of the American Association of Public Opinion Research, Hollywood, Florida.

Lavrakas, P.J., McPhee, C. and Jackson, M. (2016). "Conceptual Background on Response Propensity Modeling (RPM) for Allocating Differential Survey Incentives:

Purpose, Rationale, and Operationalization." Presented at the 71st annual conference of the American Association for Public Opinion Research, Austin TX.

Jackson, M., Lavrakas, P.J., McPhee, C. and Masterton, M. (2016). "Developing and Validating a Response Propensity Model for the Efficient Allocation of Tailored Noncontingent Incentives." Presented at the 71st annual conference of the American Association for Public Opinion Research, Austin TX.

McPhee, C., Jackson, M., Lavrakas, P.J., and Masterton, M. (2016). "Can Response Propensity Modeling be Used to Improve Response Rates, Reduce Nonresponse Bias, and Reduce Cost through the Use of Tailored Incentives?" Presented at the 71st annual conference of the American Association for Public Opinion Research, Austin TX.