NORC at the UNIVERSITY of CHICAGO

NORC WORKING PAPER SERIES

Efficiency Improvements in Multi-Mode ABS Studies: Modeling Initial Mode Decisions

WP-2014-002 | OCTOBER, 2014

PRESENTED BY:
NORC at the University of Chicago

AUTHORS:
Rebecca L. Reimer
Ashley Amaya
Felicia LeClere

Please send comments directly to principal author.

AUTHOR INFORMATION

Rebecca L. Reimer
NORC at the University of Chicago
55 E. Monroe St., 30th Floor
Chicago, IL 60603
Office: 312-357-3865
Fax: 312-759-4004
reimer-becky@norc.org

Ashley Amaya
University of Maryland
College Park, MD

Felicia LeClere
NORC at the University of Chicago
55 E. Monroe St., 30th Floor
Chicago, IL 60603
leclere-felicia@norc.org

Table of Contents

Abstract **1**

Introduction **2**

Methods **5**

Analysis Plan **7**

Measures and Methods 7

Analytic Strategy 10

Results **11**

Discussion **20**

References Cited **22**

Table of Tables

Table 1: Odds Ratios Indicating the Propensity for Cases with Address-to-Telephone Matches to Become 1) Resolved, 2) Resolved and Working, 3) Resolved, Working, and Contacted, and 4) Resolved, Working, Contacted, and Accurate during Phone Operations 12

Table 2: Odds Ratios Indicating the Propensity for Cases with Address-to-Telephone Matches and a Known First Call Outcome to Become Resolved, Working, Contacted, and Accurate during Phone Operations 16

Table 3: Potential Variation in Telephone Yields Based on Different Mode Decisions Using Predicted Probabilities from Model 4 (Resolved, Working, Contacted, and Accurate; Excluding First Call Outcomes) and Model 5 (Resolved, Working, Contacted, and Accurate; Including First Call Outcomes) 19

Table of Figures

Figure 1: Telephone Case Outcomes 6

Figure 2: Ideal and Actual Distributions of Predicted Probabilities of Becoming Resolved, Working, Contacted, and Accurate within the Telephone Mode 18

Abstract

Selecting the mode of data collection is one of the earliest and most important decisions to be made in planning for any survey. Many factors play a role in selecting the best mode or modes to meet particular project objectives, such as population coverage and nonresponse, speed of completion, and cost. With the decline in response rates and rise of survey costs, more surveys are adopting multi-mode designs. This study investigates a method of using telephone match information and sample line flags to choose an optimal mode for beginning data collection in a multi-mode survey. In this study, addresses from a subset of cases from the Racial and Ethnic Approaches to Community Health Across the U.S. Risk Factor Survey (REACH U.S.), a multi-mode ABS survey, are matched to telephone numbers by four vendors. Address and match characteristics are used to build logistic regression models predicting sample productivity in the telephone mode. These models are designed to enhance project and sample planning and steer initial mode decisions toward those that maximize yields in the telephone mode (while sending cases unlikely to be productive to other modes). Useful predictors include building type, own/rent status, vacancy flag, metropolitan statistical area designation, the identity of vendors returning matches, and the number of vendors returning identical and discrepant matches. Addition of the first dial outcome also substantially improved the fit of the model. We discuss how these models can be customized to meet the needs of particular organizations and projects.

Introduction

Selecting the mode of data collection is one of the earliest and most important decisions to be made in planning for any survey. Many factors play a role in selecting the best mode or modes to meet particular project objectives, such as population coverage and nonresponse, speed of completion, and cost (de Leeuw, 2008). With the decline in response rates and rise of survey costs (Curtin, Presser, & Singer, 2005; de Leeuw & de Heer, 2002), more surveys are adopting multi-mode designs. As noted by de Leeuw (2005; 2008), multi-mode surveys offer opportunities to balance the shortcomings of one mode against the advantages of another, enabling researchers to strike the right balance between coverage, quality, speed, and cost. These advantages are not guaranteed, however, with a multi-mode design. The researcher must choose appropriate data collection modes that fit the population of interest, budget parameters, and timelines. Given the possibility that administering the same questions in different modes will adversely affect the consistency of measurement, the impact of a multi-mode approach on measurement error must also be evaluated. Finally, business rules must be created to determine when each mode is offered and to which respondents. This final step is the focus of this paper.

Three strategies have emerged from the literature on how cases should flow through a multi-mode design. First, a pure cost focus (de Leeuw, 2005) suggests that the least expensive mode will be the choice for all sample lines initially. Nonresponders to this initial mode may be followed up via progressively more expensive modes. While the specific goal of this approach is to minimize data collection costs overall, it is often difficult to achieve in practice. Each sample line fielded in any mode incurs some cost. If a given mode yields few responses, it is no longer inexpensive. For example, if a sample line costs $2 per line to field in a web survey with only a 5 percent response rate, a total sample of 1,000 lines will yield a cost of $40 per completed interview. It is possible a telephone interview, which has higher costs but also higher response rates, could achieve the same cost per interview. In a multi-mode design, the 950 web nonrespondents will also incur additional costs as they are subsequently fielded in other modes. Fielding the least expensive mode first also assumes that no additional information is available about each sample line that may help optimize a starting mode for a sample line, cluster, or stratum.

The second strategy for choosing a starting mode in a multi-mode design is to let the respondent decide. Either respondents can be asked about their mode choice directly or all can be offered simultaneously. Panel members who were asked their preferred mode of response and are offered that mode achieved higher response rates than respondents who were not offered their preferred mode (Hoffer, Grigorian, & Fecso, 2007; Olson, Smyth, & Wood, 2012; Selfa & Sederstrom, 2006). This technique is only feasible in panel surveys where there is previous contact with the respondent. For cross-sectional surveys, the decision can be left to the respondent by offering multiple modes concurrently. This approach, however, has been documented to reduce response rates as many respondents are overwhelmed by the decision and either delay or avoid the choice (Medway & Fulton, 2012; Messer & Dillman, 2011; Millar & Dillman, 2011; Smyth, Dillman, Christian, & O'Neill, 2010).

The third strategy is to use an adaptive or responsive design to select the starting mode and subsequent modes. Originally described by Groves and Heeringa (2006), this approach is based on statistical models using paradata and response data to determine when and how a sample line should proceed through data collection. It has been used most successfully in computer-assisted telephone interviewing (CATI) administration where predictive models can be used to determine the best time to call a sample line (Brick, Allen, Cunningham, & Maklan, 1996; Durrant, D'Arrigo, & Steele, 2011). Other studies on mail surveys have used paradata available from the sampling frame and ancillary data regarding the demographic makeup of the geography to predict nonresponse to the U.S. Census (Bruce & Robinson, 2009). To date, however, little research has been conducted using sampling frame data and paradata to choose a primary data collection protocol or make an initial mode choice.

In this study, we suggest a fourth strategy in which sample line information and paradata are used prior to the start of data collection to guide the choice of the optimal starting data collection mode for a multi-mode survey. No single model will fit the need of every survey as the available inputs and the relationship of those inputs to survey performance will vary by geography, frame, data collection modes, and target population. The optimal mode to contact a respondent, moreover, may not be the optimal mode to elicit a response (Olson, Smyth, & Wood, 2012). Researchers may opt to build models to predict the probability of contact and response within each mode. For example, if topic salience is high, the researcher may give more credence to a contact model while still evaluating different response propensities.

This paper focuses on one example of how to use paradata to determine a priori which sample lines should initially be fielded by telephone. For the purposes of this example, we limit our focus to making contact with a potential respondent. We use data from a multi-mode (mail and telephone) address-based sample (ABS) design that uses the U.S. Postal Service's Delivery Sequence File (DSF). While the DSF is generated and updated for the purposes of delivering mail, it includes the actual mailing addresses, as well as a number of other variables, such as address type (e.g., flags for college addresses, seasonal addresses, vacancy) and building type (single unit vs. multi-unit). Users of the DSF often attempt to append additional information to the frame such as telephone numbers and demographic information at the individual or Census tract level. In many cases, additional demographic information such as race/ethnicity and education of householders can also be appended to the addresses by commercial vendors that use proprietary algorithms to predict the demographic makeup of these households. Commercial vendors also offer telephone matching services that can provide matches for a substantial percentage of addresses sampled from the DSF, although the match rate can vary based on sample characteristics such as geographic location or type of housing unit (Amaya, Skalland, & Wooten, 2010).

To date, decisions about the initial data collection mode in ABS surveys have been unsophisticated, with all cases assigned the same mode sequence. Some researchers elect to send all matched cases to the telephone first, switching incomplete cases to a second mode after time has passed or a maximum number of call attempts have been made (e.g., Bailey, Grabowski, & Link, 2010; Link, Daily, Shuttles, Bourquin, & Yancey, 2009). Given the potentially low yield of the telephone numbers, one may argue this is inefficient and opt for a mail first approach such as that used by the National Household Education Surveys (Brick, Williams, & Montaquila, 2011). The telephone remains useful, however. It results in a shorter data collection timeline. It also allows for complex skip logic and higher data quality than a mailed questionnaire. Response rates are also often higher as interviewers can build rapport and answer questions that the respondent may have. To maximize the potential of a complex mode start to an ABS design, we build a model to identify the likelihood that sample lines will result in telephone contact with a potential respondent. This information then can be used on a case-by-case basis to assign the starting mode for the sample line.

The key question for the current study is whether sample characteristics and telephone match characteristics will be useful in building a comprehensive model for predicting whether sample lines with matched telephone numbers will have working telephones and whether contact will be made with the sampled address when dialed in CATI. This predictive model will have three main benefits. First, we expect that it will simplify future sample planning by increasing the accuracy of yield projections for CATI cases. Second, we expect to improve on the common practice in multi-mode ABS studies of sending all cases to the same starting mode by being able to predict which mode will be most productive for a given sample line. We create models that minimize ad hoc choices and allow decisions about sample size and mode earlier in project planning. This should help avoid major recalibrations in sample needs and mode shifts after the study is already in the field. Our third goal is to improve the value of the initial dials where projects have some flexibility in their phone system operations to incorporate the outcomes of initial calls to each matched telephone number. We will examine the additional impact of the outcome of the first dial on our predictive model of working and accurate contact telephone numbers. While these goals are specific to a multi-mode ABS design, they are generalizable to other multi-mode surveys that use paradata and sampling frame information to make more informed decisions earlier in the survey process to improve data quality and efficiency.

Methods

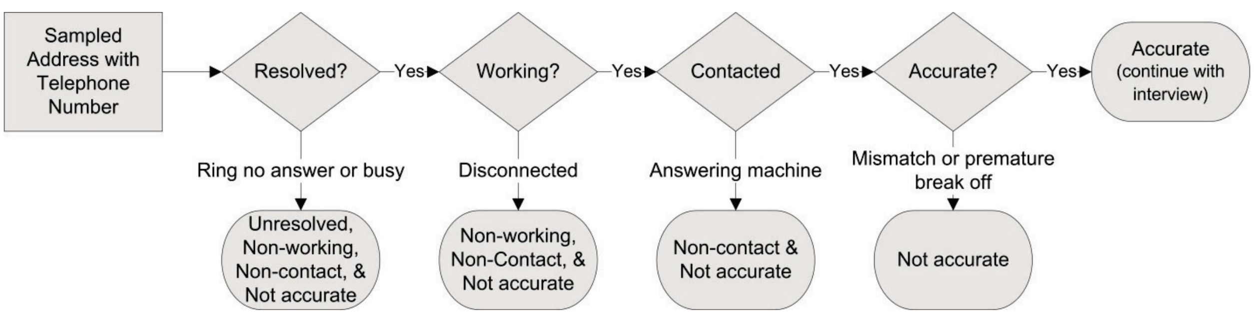
Data for this study come from a subset of cases from Phase 3 of the Racial and Ethnic Approaches to Community Health Across the U.S. Risk Factor Survey (REACH U.S.), conducted in 2010 and 2011. REACH U.S. is a key piece of the Centers for Disease Control and Prevention's efforts to eliminate racial and ethnic health disparities. The survey focuses on Black, Hispanic, Asian (including Native Hawaiian and Other Pacific Islander), and Native American populations. In each of 28 REACH U.S. communities, NORC conducts the annual risk factor survey addressing breast and cervical cancer, cardiovascular disease, diabetes, adult immunizations, and hepatitis B. The communities are located within 17 states, and range in size from a small neighborhood to an entire state (Oklahoma), depending on the target population. They are primarily located in urban areas, but several communities include suburban and/or rural areas (e.g., for Native American populations in Oklahoma and Michigan, African American populations in West Virginia and South Carolina). REACH U.S. uses a multi-mode ABS design that includes telephone and mail modes, with a limited amount of in-person interviewing. All cases with addresses that could be matched to telephone numbers were first attempted via CATI and moved to the mail mode later, if necessary. Cases that could not be matched to telephone numbers were sent directly to the mail mode.

We randomly selected approximately 1,000 cases from each of 26 REACH U.S. communities, for a total of 25,849 sample lines. Two additional REACH U.S. communities were excluded for design reasons. We sent the addresses for these cases to four vendors, including Marketing Systems Group (MSG), SSI (formerly known as Survey Sampling International), Relevate, and Valassis (formerly known as ADVO). All four vendors returned telephone matches for the addresses we provided, with varying formats. MSG's methods differ from other vendors, and they offer two categories for matches: "exact" matches where the street address and any unit number that is present both match, and "inexact" matches, which are less precise. The latter may include addresses where it is unclear whether the unit number matches (e.g., 123 Main St. vs. 123 Main St., Apt #2) or directional qualifiers vary (e.g., 123 Main St. vs. 123 North Main St.). Previous research (Amaya et al., 2010) suggests that inexact matches are of lower quality; thus, we excluded them from our request to MSG for this study. Along with telephone matches, MSG also provided us with a number of demographic variables for each household. Finally, we obtained several address designation variables from the DSF (via Valassis), including single unit versus multi-unit flag, vacancy flag, etc.

After we obtained phone matches from all vendors, we prioritized the numbers for each case such that only the number believed to be "best" would be dialed in the CATI system. In cases where only one vendor provided a match or where multiple vendors provided the same telephone number, that number was designated as "best." In cases where conflicting numbers were obtained, we selected the number provided by the majority of vendors, if possible (e.g., if two vendors returned one phone number and a third vendor returned a different number, we selected the first). If conflicting numbers were obtained from equal numbers of vendors, we selected the number provided by the vendor listed earliest in the following prioritized list: MSG, SSI, and then Relevate. (Note that telephone matches from Valassis were obtained after data collection was already underway and were not loaded into our CATI system. These numbers were dialed for some cases if one of the other vendors had provided the same telephone number.) All cases that did not have telephone matches were moved to the mail mode and excluded from this analysis.

Cases with telephone matches were called using standard REACH U.S. procedures during the data collection period from December 2, 2010, to July 31, 2011. Cases progressed through several stages before reaching the screener and extended interview. There are four key stages (displayed in Figure 1) that must be achieved before the sample line can move forward and an interview can be conducted. Telephone numbers must be: 1) resolved, 2) working, 3) contacted, and 4) accurate. These stages are described below.

Figure 1: Telephone Case Outcomes



First, we attempted to resolve whether a telephone number was working. If we were able to make a clear determination that a number was working, the telephone number was categorized as *resolved*. In cases where this could not be accomplished because all call results included non-contacts such as ring no answers and busy signals, the phone numbers were categorized as *unresolved*. Phone numbers where two call results indicated that the line was disconnected or reached a fax/modem were categorized as *non-working*. Those where we were able to connect to an individual or an answering machine were categorized as *working*. Working telephone numbers were identified as a *contact* if any dial resulted in human contact and *non-contact* if a dial only reached an answering machine, privacy manager, or voicemail.

Finally, we assessed whether contacted numbers were accurate. After reaching a respondent and confirming that the respondent was an adult, we attempted to determine the accuracy of the telephone-to-address matching process. It is important to note that survey research organizations differ in their protocols on the verification of the accuracy of telephone matches for ABS samples. In most circumstances, we choose to confirm with respondents that the telephone number dialed matches the sampled address. We maintain this practice to ensure that every household reached has a measureable probability of selection as defined by the sampling plan and that we are indeed surveying within the intended geographic areas. For this phase of data collection, the standard accuracy protocol was selectively applied based on accuracy rates from previous years of data collection. Several communities had high accuracy rates in the previous two years; thus, the verification process was eliminated. This offered the advantage of avoiding break-offs that frequently occur at the verification question, while maintaining a high level of confidence in the match accuracy. In some communities, low accuracy rates from previous years suggested that verification for all addresses in the community was necessary. Finally, for the remaining communities, previous analysis showed that addresses in single unit buildings had high accuracy rates while those in multi-unit buildings were lower; thus, the verification protocol was used only for cases located within multi-unit buildings.

When verifying accuracy, interviewers told each respondent that an advance letter had been mailed to the case address and then asked if it was still the correct address for the household. In cases where the respondent indicated that the address was incorrect, the interview was terminated, the telephone number was finalized as a mismatch, and the case was moved to the mail. If the street address matched but the respondent lived in a different unit of a multi-unit building than the unit number listed in our sample file, the case was treated as a mismatch. The *not accurate* category includes both mismatched cases and those where the accuracy protocol could not be completed because the respondent terminated the interview prior to answering the question.

Analysis Plan

Measures and Methods

With the goal of improving the productivity of telephone data collection, we estimate logistic regression models to address different objectives. While the ultimate goal of data collection is to complete full interviews in CATI, we model outcomes that occur earlier in the data collection process. Surveys vary greatly in their content, length, and screening procedures; thus, screener completion or interview completion rates introduce factors specific to the survey that may not be relevant to other projects or organizations. In multi-mode surveys, many transitions occur in the life of a particular case, making it difficult to disentangle determinants of the starting mode from subsequent modes. We focus on the quality of the telephone match, which determines the success of CATI as a starting data collection mode. Therefore, we define four binary dependent variables that refer to the status of each telephone line based on the final CATI disposition for each case: 1) *resolved*, 2) *resolved and working*, 3) *resolved, working, and contacted*, and 4) *resolved, working, contacted, and accurate* (RWCA).

Telephone numbers achieve resolution when a clear determination is made about whether the number is working. The only cases considered unresolved are those with call histories consisting solely of non-connected outcomes. All other cases that are dialed in CATI are classified as resolved. Attaining the status of resolved is the first step in successful telephone interviewing. This outcome is the dependent variable for Model 1.

Cases are classified as resolved and working when they are both resolved and confirmed as reaching a working telephone line. This outcome is the dependent variable for Model 2. Survey researchers have little control over whether a matched number is resolved and working, but the quality of the telephone match can be assessed using these criteria. We also include these models as points of comparison.

Cases are classified as resolved, working, and contacted when at least one dial to a resolved, working number results in contact with a respondent. This status is the desired case outcome for organizations and projects that do not require verification of the accuracy of telephone matches and should be most useful to them. It is the dependent variable for Model 3.

Cases are classified as achieving the fourth outcome when contact with a working number occurs and the accuracy of the telephone match is confirmed according to the accuracy protocol described above. Cases that do not achieve this status include both those where the respondent indicates that the address match is incorrect (mismatches) and cases where we are unable to complete the accuracy protocol due to non-contact or break-off by the respondent. The primary focus of this study is the fourth outcome because measurable sampling rates require match confirmation. It is the dependent variable for Models 4 and 5.

We use the same set of match characteristics and sample characteristics for all models. Model 5 includes these as well as variables classifying the outcome of the first dial in CATI. We have divided the models' independent variables into three categories: characteristics regarding the telephone match, housing unit variables from the DSF frame, and demographic flag variables from MSG. Telephone match variables include: 1) whether there was a telephone match from MSG or only from one or more of the other vendors, 2) the number of vendors that returned the same telephone number, and 3) whether any vendors returned conflicting phone number matches. The characteristics of housing unit are from the DSF address file. The DSF includes an address type, which indicates whether the address is a single unit building, located within a multi-unit building, or has a status of other/unknown. This designation is one of several fields updated by mail carriers for the purposes of streamlining mail delivery. The DSF also includes a vacancy flag for some addresses. Additionally, the county-level information included in the DSF has been geocoded to generate an indicator of metropolitan statistical area using the Office of Management and Budget statistical area definitions (2009). These federal guidelines are used based on population density include metropolitan, micropolitan, and rural.

Finally, demographic characteristics of sampled households come from the sample vendor MSG in the form of address flags. These characteristics are estimated from proprietary statistical models using a variety of data sources such as Census demographics, public listings, and marketing subscription lists. Demographic flags provided by MSG include: 1) whether the address is occupied by owners or renters, 2) the race/ethnicity of the household members, 3) the sex of the head of household, 4) the number of adults in the household, 5) a flag to indicate if there are children in the household, 6) the age of the head of household, 7) annual household income, 8) whether the head of household is married or single, and 9) the highest educational attainment in the household. Information is not available for all cases. We include the cases with missing demographic flags, as there is likely an inverse correlation between the lack of information on the sample line and the probability of reaching a respondent from the sampled address.

Results from Models 3 and 4 will inform initial mode decisions early in project planning before data collection. Model 5 incorporates additional variables to assess the predictive utility of the outcome of first dial attempts to cases with telephone matches. Model 5, therefore, includes all cases that have a telephone match that did not finalize or achieve accurate status on their first call. This information can be operationalized in two ways (Groves & Heeringa, 2006). Researchers could open data collection with a single dial attempt on all cases with a telephone match, followed by additional modeling. Interrupting data collection after a single dial is not likely to be feasible in all situations but might provide additional insight for those organizations with flexibility. Alternatively, researchers could develop responsive designs that anticipate the possible outcomes of first dials and implement the business rules accordingly.

Analytic Strategy

Our main goal with this analysis is to use model estimates to build a predictive tool for survey planning. First, we assess the impact of each predictor on the probability that the telephone number is resolved, working, and can be used to contact someone at an accurate address. This will help us to evaluate which predictors may be most useful to acquire during project planning. Second, we use the predicted probabilities from the planning models to assess how survey efficiency might be improved. As noted, we are interested in maximizing the efficiency of telephone operations by making initial mode decisions by sample line. There are two ways to approach this goal. The simpler approach is to use information discovered about each sample characteristic to develop rules of thumb applied to address types in aggregate. For example, if we find that telephone match quality in multi-unit buildings is very low, all multi-unit cases with telephone matches can be allocated to the mail mode first and all other matches to the telephone mode first.

Conversely, we pursue a more complex approach using predicted probabilities from our models to predict success in the telephone mode for matched cases. To simulate the effect of making different initial mode decisions, we use model parameters to generate an individual predicted probability of achieving a RWCA outcome for each sample line. Those cases with lower predicted probabilities have a combination of characteristics that make them less likely to be contacted and for the address to be accurate, while those with higher predicted probabilities suggest that they are more likely to be successful in CATI. Using data from cases on REACH U.S., we then simulate the consequences of making different mode decisions for cases with lower probabilities of success in CATI. For example, if these model parameters were available during the sample planning phase and only the cases in the upper quartile of the predicted probability of being RWCA went to the telephone and the remainder to mail, how could this affect overall survey efficiency? Similarly, what would be the impact of alternative thresholds on efficiency?

Using these thresholds to make mode decisions with a strong predictive model, we should observe improved efficiency. Specifically, if cases expected to have low CATI success rates are sent directly to the mail, time and money spent on unproductive calls can be saved. Analogously, CATI success rates should be higher for the subset of cases sent to the telephone mode than when all matched cases begin in the telephone mode. Success is measured as the yield of cases that are resolved, working, contacted, and accurate among all dialed cases. We use estimates from both the RWCA model (Model 4) and the first dial model (Model 5) to generate these simulations.

Results

Table 1 includes results from our models in which telephone numbers matched to addresses will be classified as resolved, working, contacted, and accurate within the telephone mode. All predictors in the models (except age) use "unknown" (e.g., those without values available) as the reference category unless otherwise specified. Age had a value for all sample lines. Models 1 through 4 include all cases that had at least one telephone match (N=13,719, or 53 percent of the 25,849 cases in the study). While the results indicate that one-third of the predictor variables are statistically significant, the model fit is only fair as demonstrated by the pseudo R^2 values. The fit does improve from Models 1 and 3 to Models 2 and 4. It is important to note that the direction of the odds ratios is different for several variables across the models. We expect this given that a resolved telephone number can include a wide variety of outcomes, including many non-productive calls such as disconnected numbers. Nearly all of the results for Model 3 are similar to those for Model 4.

Table 1: Odds Ratios Indicating the Propensity for Cases with Address-to-Telephone Matches to Become 1) Resolved, 2) Resolved and Working, 3) Resolved, Working, and Contacted, and 4) Resolved, Working, Contacted, and Accurate during Phone Operations

	Model 1:		Model 2:		Model 3:		Model 4:	
	Resolved		Resolved & working		Resolved, working, & contacted		Resolved, working, contacted, & accurate	
	(N=13,719)		(N=13,719)		(N=13,719)		(N=13,719)	
	Pseudo R^2 = 0.04		Pseudo R^2 = 0.12		Pseudo R^2 = 0.07		Pseudo R^2 = 0.11	
Characteristic	**OR**	**95% CI**	**OR**	**95% CI**	**OR**	**95% CI**	**OR**	**95% CI**
Telephone Match Variables								
MSG Match	1.25	0.98 - 1.59	0.71***	0.60 - 0.84	0.76***	0.64 - 0.89	0.75***	0.63 - 0.89
Count of Vendors with Same Match (Range: 1-4)	0.79***	0.74 - 0.83	1.58***	1.52 - 1.64	1.36***	1.31 - 1.41	1.38***	1.33 - 1.44
Match Conflict	0.76**	0.64 - 0.91	0.87*	0.77 - 0.98	0.93	0.82 - 1.06	0.86*	0.75 - 0.98
DSF Frame Variables								
Address Type								
Single unit	1.14***	0.96 - 1.35	0.71***	0.63 - 0.81	0.82***	0.73 - 0.92	0.75	0.66 - 0.85
Multi-unit	0.57***	0.47 - 0.68	1.03	0.90 - 1.18	0.99	0.87 - 1.13	0.57***	0.49 - 0.65
MSA (Reference=Metro)								
Micro	0.88	0.74 - 1.05	0.82	0.72 - 0.93	1.10	0.97 - 1.24	1.18**	1.05 - 1.34
Rural	0.60***	0.46 - 0.78	0.85	0.68 - 1.05	1.10	0.90 - 1.35	0.84*	0.68 - 1.04
Vacant	1.03	0.69 - 1.53	0.52***	0.39 - 0.69	0.60***	0.45 - 0.81	0.54***	0.38 - 0.76
MSG Demographic Variables								
Own/Rent Status								
Own	0.96	0.77 - 1.18	1.31**	1.12 - 1.52	1.10	0.95 - 1.28	1.16*	1.00 - 1.36
Rent	1.16*	0.94 - 1.44	1.21	1.04 - 1.42	1.12	0.96 - 1.30	1.08	0.92 - 1.27
Race/Ethnicity								
Black/African American	1.48***	1.00 - 2.20	1.20	0.90 - 1.61	1.20*	0.91 - 1.57	1.28**	0.96 - 1.70
Asian	0.60***	0.40 - 0.90	0.96*	0.70 - 1.30	1.03	0.77 - 1.38	1.18	0.88 - 1.60
Native Hawaiian/Pacific Islander	1.15	0.78 - 1.70	1.19	0.89 - 1.59	1.09	0.83 - 1.44	1.05	0.79 - 1.39
Hispanic	1.44	0.80 - 2.57	1.40	0.93 - 2.10	0.98	0.67 - 1.42	1.11	0.76 - 1.62
Other	1.16	0.78 - 1.72	1.17	0.87 - 1.56	1.13	0.86 - 1.49	1.05	0.79 - 1.39

	Model 1:		Model 2:		Model 3:		Model 4:	
	Resolved		Resolved & working		Resolved, working, & contacted		Resolved, working, contacted, & accurate	
	(N=13,719)		(N=13,719)		(N=13,719)		(N=13,719)	
	Pseudo R^2 = 0.04		Pseudo R^2 = 0.12		Pseudo R^2 = 0.07		Pseudo R^2 = 0.11	
Sex								
Male	1.35	1.09 - 1.68	1.21*	1.01 - 1.44	1.10	0.92 - 1.31	1.05	0.87 - 1.26
Female	1.49***	1.20 - 1.84	1.13	0.95 - 1.35	1.01	0.85 - 1.20	0.94	0.79 - 1.13
Number of Adults in the Household								
1	0.26*	0.06 - 1.15	0.53*	0.24 - 1.17	0.63*	0.29 - 1.37	0.63	0.28 - 1.41
2	0.29	0.07 - 1.28	0.59	0.27 - 1.30	0.75	0.34 - 1.62	0.72	0.32 - 1.64
3 or more	0.31	0.07 - 1.37	0.63	0.29 - 1.39	0.79	0.36 - 1.71	0.77	0.34 - 1.75
Has Children in the Household	1.16	1.00 - 1.36	0.91	0.82 - 1.01	1.05	0.95 - 1.16	1.04	0.93 - 1.15
Age of Head of Household (Reference=65 and older)								
18-24	0.26*	0.45 - 0.94	0.60	0.45 - 0.81	0.61	0.46 - 0.80	0.55	0.41 - 0.74
25-34	0.29	0.77 - 1.05	0.55***	0.48 - 0.62	0.58***	0.52 - 0.65	0.55***	0.49 - 0.62
35-64	0.31***	0.98 - 1.34	0.76	0.67 - 0.85	0.67	0.60 - 0.74	0.64	0.58 - 0.72
Annual Household Income								
Less than $20,000	1.66	0.39 - 7.11	1.38	0.64 - 2.94	1.58*	0.75 - 3.35	2.37**	1.07 - 5.24
$20,000-$34,999	1.63	0.38 - 6.98	1.26	0.59 - 2.68	1.39	0.66 - 2.94	2.00	0.91 - 4.44
$35,000-$54,999	1.61	0.38 - 6.91	1.32	0.62 - 2.83	1.40	0.66 - 2.96	2.00	0.90 - 4.43
$55,000 and up	2.05	0.48 - 8.81	1.49	0.69 - 3.19	1.32	0.62 - 2.80	1.79	0.81 - 3.98
Marital Status								
Married	1.05	0.88 - 1.25	0.96	0.84 - 1.09	0.86	0.76 - 0.97	0.86	0.76 - 0.97
Single	1.00	0.85 - 1.18	0.86*	0.76 - 0.97	0.80**	0.71 - 0.90	0.84	0.74 - 0.95
Education								
Less than high school	0.94	0.63 - 1.41	0.65**	0.48 - 0.88	0.79	0.59 - 1.04	0.84	0.63 - 1.13
High school diploma/GED	1.00	0.67 - 1.50	0.68*	0.51 - 0.92	0.70*	0.53 - 0.93	0.79	0.59 - 1.05
Some college	1.01	0.68 - 1.52	0.71	0.53 - 0.96	0.74	0.55 - 0.98	0.82	0.61 - 1.09
College degree or higher	1.26*	0.83 - 1.90	0.83	0.61 - 1.13	0.72	0.54 - 0.96	0.83	0.62 - 1.12

* p<0.05; ** p<0.01; *** p<0.001

The results from Model 4 (RWCA) are of most interest here. Sample addresses that match to a telephone number from MSG have lower odds of becoming RWCA than those with matches only through other vendors ($p<0.001$). While MSG provides a very large quantity of matches, there may be more variability in the quality they provide even among "exact" matches. The other three vendors return fewer matches but may use more conservative methods to yield higher proportions of contacted and accurate numbers. The odds of a telephone matched case being RWCA ($p<0.001$) increases as the number of vendors return the same telephone number in the matching process. By contrast, cases with match conflicts have lower odds of success than those without conflicts ($p<0.05$).

Previous work done by Amaya et al. (2010) suggests that telephone numbers matched to addresses in multi-unit buildings have lower odds of becoming RWCA than cases without a building type match ($p<0.001$). In the case of building type, "unknowns" include rural route addresses and P.O. boxes, which tend to have lower match rates but higher match quality. Sample lines in single unit buildings are also less likely to be RWCA compared to "unknown" addresses, but the differences are not significant. Cases in metropolitan areas have slightly lower odds of becoming RWCA than those in micropolitan areas ($p<0.01$) and have slightly higher odds than those in rural areas ($p<0.05$). Addresses flagged as vacant have lower odds of becoming RWCA than those that are not flagged ($p<0.001$). Many of the cases in the study with the vacancy flag did not have telephone matches and are thus excluded from these models. Less than 2 percent of cases included here have the vacancy flag.

Addresses where the housing units have any own/rent designation have slightly higher odds of being resolved, working, contacted, and accurate than unknown units, although only those flagged as owned are significantly different ($p<0.05$). The unknown category includes a more transient population than homeowners or long-term renters. Among the racial/ethnic groups, only addresses flagged as housing Blacks/African Americans are significantly more likely than the unknown group to have RWCA phone numbers ($p<0.001$). Although, all addresses flagged with some racial/ethnic category have higher odds than the unknown group. This finding may seem counterintuitive at first glance since minority groups are less likely to live in landline households and have listed telephone numbers (Blumberg & Luke, 2012). Model 4, however, is limited to addresses with a matched phone number, which is likely to be a landline, and is not representative of the general population or entire sample. While a lower proportion of Blacks may have a matched telephone number than is observed in other subgroups, those with a matched telephone number are more likely to result in a RWCA outcome. All of the younger age groups had higher odds than the households with a person age 65 years and older group, but only those in the age group 25-34 were significantly lower ($p<0.001$). Similarly, all addresses with any income indicator had higher odds than those addresses missing an income indicator, but only those flagged with an annual income less than $20,000 were significantly different from the unknown category ($p<0.01$).

None of the other demographic variables—including sex of head of household, number of adults living in the household, the presence of children within the household, marital status, and highest educational attainment—were significantly related to the odds that the matched phone number would be RWCA in Model 4. It is important to remember that these variables are attached to the sample line addresses by the vendor MSG and are generated through predictive algorithms. Little information is available about the accuracy of the sample flags in predicting household or household head characteristics.

Estimates from Model 5 are shown in Table 2. This model includes all cases that received telephone matches and required follow-up calls after the first dial (N=12,605). It excludes cases where respondents completed the accuracy protocol on the first dial and cases that were finalized for a variety of reasons on the first dial. This type of outcome is useful for projects where there is the flexibility to conduct additional modeling and to make mode change decisions after a single dial to all matched cases. The results show that cases where any human contact has been made that did not result in a refusal on the first call are much more likely to be recontacted and classified as accurate on future calls. This is not surprising since three of the four criteria (resolved, working, and contacted) have already been accomplished on the first dial. The difference between human contact and non-working outcomes is statistically significant ($p<0.001$), and the difference between human contact and answering machine outcomes is marginally significant ($p=0.05$). Several other variables are no longer significant, including MSG match, vacancy, owner occupancy, and if the head of household is 25-34 years old. While the magnitude of the odds ratio for single unit addresses did not change (0.75 and 0.76 for Model 4 and Model 5, respectively), the value is significant in Model 5 ($p<0.01$). All variables continued to trend in the same direction. The addition of the first call outcome results is a substantial improvement in model fit; the pseudo R^2 value is 0.11 for Model 4 and rises to 0.23 for Model 5. While the model fit remains relatively poor, it is still true that it provides significantly more information than assigning a case to the telephone simply by the presence of a matching phone number.

Table 2: Odds Ratios Indicating the Propensity for Cases with Address-to-Telephone Matches and a Known First Call Outcome to Become Resolved, Working, Contacted, and Accurate during Phone Operations

	Model 5:	
	RWCA	
	(N=12,605)	
	Pseudo R^2 = 0.23	
Characteristic	**OR**	**95% CI**
Telephone Match Variables		
MSG Match	0.84	0.69 - 1.02
Count of Vendors with Same Match (Range: 1-4)	1.18***	1.13 - 1.24
Match Conflict	0.83*	0.71 - 0.96
DSF Frame Variables		
Address Type		
Single unit	0.76**	0.66 - 0.87
Multi-unit	0.40***	0.35 - 0.47
MSA (Reference=Metro)		
Micro	1.27***	1.10 - 1.46
Rural	0.78**	0.61 - 1.00
Vacant	0.72	0.48 - 1.06
MSG Demographic Variables		
Own/Rent Status		
Own	1.05	0.88 - 1.26
Rent	0.99	0.83 - 1.19
Race/Ethnicity		
Black/African American	1.36**	0.99 - 1.88
Asian	1.15	0.81 - 1.61
Native Hawaiian/Pacific Islander	1.02	0.74 - 1.42
Hispanic	1.21	0.79 - 1.86
Other	1.14	0.82 - 1.57
Sex		
Male	1.08	0.88 - 1.33
Female	0.97	0.79 - 1.19
Number of Adults in the Household		
1	0.60	0.24 - 1.48
2	0.65	0.26 - 1.60
3 or more	0.69	0.28 - 1.72
Has Children in the Household	1.09	0.97 - 1.22

	Model 5: RWCA (N=12,605) Pseudo R^2 = 0.23	
Characteristic	**OR**	**95% CI**
Age of Head of Household (Reference=65 and older)		
18-24	0.57*	0.41 - 0.80
25-34	0.68	0.60 - 0.77
35-64	0.79	0.70 - 0.89
Annual Household Income		
Less than $20,000	2.61**	1.08 - 6.33
$20,000-$34,999	2.27	0.94 - 5.50
$35,000-$54,999	2.30	0.95 - 5.57
$55,000 and up	2.00	0.82 - 5.57
Marital Status		
Married	0.92	0.80 - 1.06
Single	0.91	0.79 - 1.05
Education		
Less than high school	0.87	0.62 - 1.22
High school diploma/GED	0.83	0.60 - 1.16
Some college	0.84	0.60 - 1.18
College degree or higher	0.82	0.59 - 1.15
Data Collection Operation Variables		
Outcome of First Call (Reference=Human Contact[a])		
No answer	0.31	0.27 - 0.36
Answering machine	0.08*	0.07 - 0.09
Non-working[b]	0.30***	0.26 - 0.35
Refusal[c]	0.29	0.25 - 0.33

* p<0.05; ** p<0.01; *** p<0.001

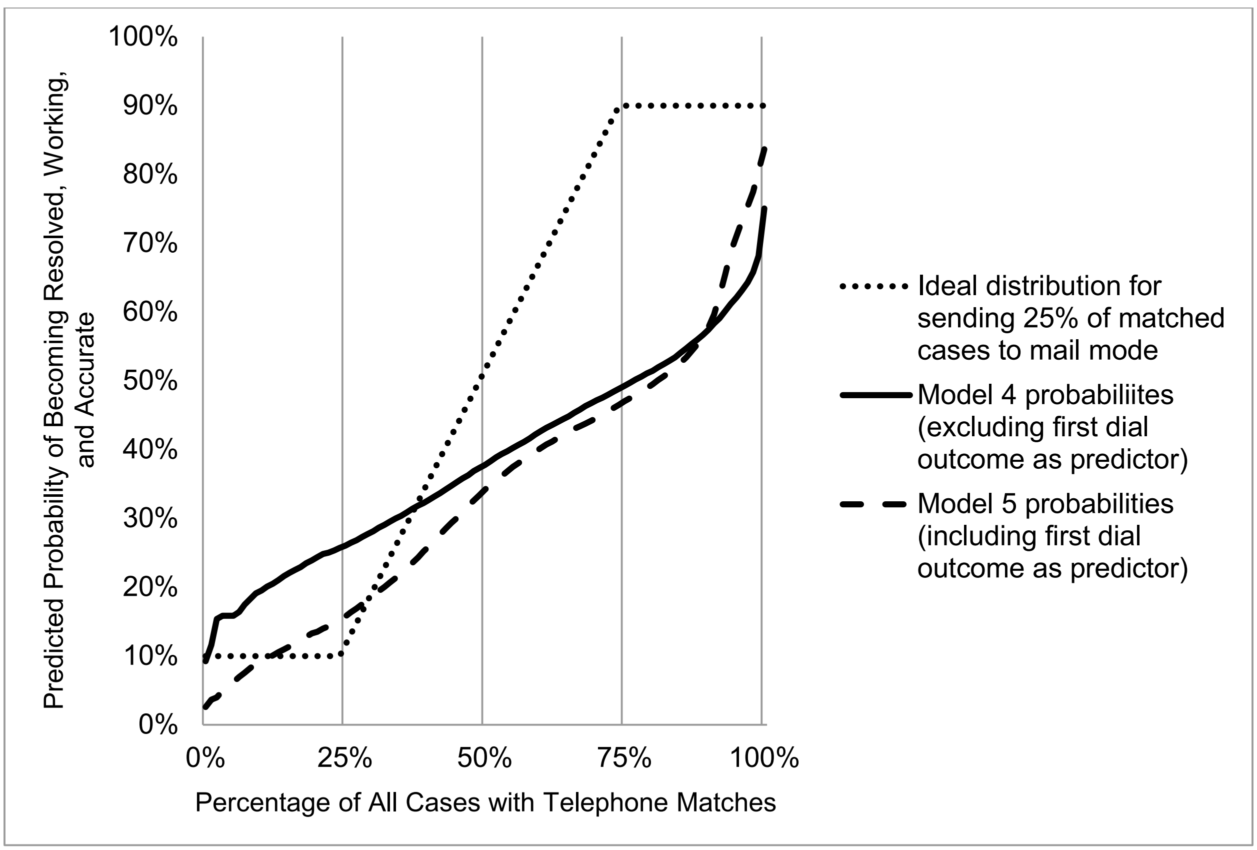
[a] This category includes cases that had human contact, including call backs, appointments, and cases flagged for supervisor review. Cases that progressed far enough through the screener to verify match accuracy on the first call were excluded from this model.

[b] This category includes cases with disconnected numbers, changed numbers, fax/modems, GPS/Onstar outcomes, technical problems, and other non-contacts with unresolved outcomes.

[c] This category includes all types of refusals, including cases where the respondent hung up during the introduction.

In the next phase of the analysis, we investigate whether this method for making mode choices can improve CATI operations and, thus, survey planning. Using model parameters from Models 4 and 5, we simulate how different choices in mode allocation will affect survey efficiency by restricting the use of the telephone mode to those cases that are most likely to be successful. First, using the estimated model parameters, we investigate cases that are unlikely to be resolved, working, contacted, and accurate via the telephone mode. We then investigate the potential impact of moving these cases to the mail mode early in the data collection process to minimize unproductive dialing. To do this, we generate predicted probabilities for each case, for both Model 4 and Model 5. The distribution of the estimated probabilities is in Figure 2. To assess the effect of our responsive design approach to initial mode choice, we propose using the distribution of the predictive probabilities of sample lines being RWCA to establish thresholds for assigning cases to initial modes. The vertical lines in Figure 2 represent our selected thresholds, which are quartiles. We use these thresholds to simulate the efficiency consequences of sending some matched cases immediately to the mail mode without dialing or after making a single dial. We are interested specifically in the yield rate that would be achieved at each threshold level. The yield rate is defined as the proportion of cases starting in the telephone mode that become resolved, working, contacted, and accurate. There are certainly trade-offs to be considered as implementation of mode decisions based on these thresholds may improve overall efficiency within the telephone mode but may also result in the need for more intense efforts in other modes. These downstream consequences, however, are beyond the scope of this analysis.

Figure 2: Ideal and Actual Distributions of Predicted Probabilities of Becoming Resolved, Working, Contacted, and Accurate within the Telephone Mode



The simulated yield rates of resolved, working, contacted, and accurate cases using the thresholds described above are shown in Table 3. With both models, higher thresholds for allocating a case to the telephone are associated with higher yields. Restricting the use of the telephone mode to only those cases with higher probabilities of success in CATI would increase the efficiency of telephone operations. For Model 4, the original yield is 38.2 percent when all cases are included. If a quarter of the cases with the lowest predicted probability of success were sent directly to mail rather than beginning in CATI, the yield rate for the remaining three-quarters of cases would rise to 44.1 percent. If the lowest two or three quartiles of cases were sent to mail, the yield rate for the remaining cases fielded in the telephone mode would be 50.3 percent or 55.9 percent, respectively.

Table 3: Potential Variation in Telephone Yields Based on Different Mode Decisions Using Predicted Probabilities from Model 4 (Resolved, Working, Contacted, and Accurate; Excluding First Call Outcomes) and Model 5 (Resolved, Working, Contacted, and Accurate; Including First Call Outcomes)

	Cases that would be initially fielded in telephone mode		Cases that would become resolved, working, contacted & accurate		Yield rate of resolved, working, contacted, & accurate cases
Characteristic	**N**	**% Retained**	**N**	**% Retained**	
Model 4:					
All matched cases	13,719	100.00%	5,245	100.00%	38.2%
Threshold 1	10,313	75.17%	4,553	86.81%	44.1%
Threshold 2	6,856	49.97%	3,448	65.74%	50.3%
Threshold 3	3,505	25.55%	1,958	37.33%	55.9%
Model 5:					
All matched cases not screened or finalized on first call	12,605	100.00%	4,240	100.00%	33.6%
Threshold 1	9,464	75.08%	4,183	98.66%	44.2%
Threshold 2	6,284	49.85%	3,052	71.98%	48.6%
Threshold 3	3,183	25.25%	1,824	43.02%	57.3%

With estimated parameters from Model 5, where decisions are made after the first dial to all matched cases, the yield increases are similar. These yields are calculated based on the cases that require a second dial. Note that 8.1 percent of all cases dialed became RWCA on the first dial and are excluded from the yield calculations. Model 5 does not include cases that progressed past the accuracy protocol and cases that were finalized for other reasons (e.g., mismatches, businesses) on the first dial, since predictions of future success would not be necessary. The yield of RWCA cases among those requiring a second dial is 33.6 percent. If the quartile of cases with the lowest probabilities of becoming RWCA were sent to the mail mode after a single dial in CATI, the yield rate would become 44.2 percent. Sending the lowest two or three quartiles of cases to the mail mode after the first dial would lead to yield rates of 48.6 percent and 57.3 percent, respectively.

Discussion

The results demonstrate that pre-survey modeling can be used to improve the CATI efficiency of multi-mode ABS surveys such as REACH U.S. Characteristics of the sample and of the telephone matches returned by commercial vendors can be used to predict, with some degree of certainty, how productive cases will be when fielded in CATI. The outcome of a single call to each matched telephone number can be used to improve this type of model substantially for projects where this strategy is feasible. Use of these types of models may be more efficient than previous methods of selecting modes, which usually includes sending all cases with telephone matches to the telephone mode. Models such as these should allow researchers to make more careful, informed decisions regarding sample plans and mode sequences.

It is clear that the outcome of the first dial can be very valuable for predicting the final CATI outcome. The choice to field all cases in CATI for at least one call attempt before making mode decisions has important consequences for logistical planning and project costs, however. For new studies, researchers are unlikely to know how easily interviewers will be able to make contact with households on the first call. This is likely to make precise planning about the number of interviewing hours and mailings required difficult, especially for studies with short data collection periods. This decision will be intrinsically different for each survey design and will depend on the total number of cases, the speed of data collection, and the cost of each data collection mode. No universal recommendation can be made about the true predictive value of the first dial.

The variables measuring the overlap between telephone matches from different vendors (number of vendors returning the same match and presence/absence of any conflicting numbers) are also valuable predictors. It is important to note that cost may be a concern for these variables. While some of the other predictor variables (such as the vacancy flags) used for the models described here are easily available because they are included in the DSF, the use of multiple vendors can be quite expensive. We included four vendors here for the sake of evaluating their abilities. The use of multiple phone matching vendors is, again, a decision to be made based on the efficiency gains weighed against costs. The availability of alternative contact information will also vary by frame, country, and population of interest. This analysis, however, underscores the variability among vendors and their abilities to provide accurate telephone matching.

There were several limitations of the present study that can be addressed in future work. First, REACH U.S. surveys a specialized group of communities that are not representative of the U.S. population. This sample is located primarily in metropolitan areas and focuses on particular racial and ethnic groups. Some of the telephone matching patterns observed here may vary between demographic groups, areas of different population densities, or specific geographic regions. While this might mean that models constructed for different sample types could show different relationships between variables, we believe that the overall utility of building this type of model should remain the same.

A second limitation relates to our methods of verifying the accuracy of telephone matches among working numbers. As noted above, the accuracy of matches ensures that every household has a known probability of selection and that the households are within the desired geographic areas. We used a modified version of the accuracy protocol, however, in some communities that had very high accuracy rates in previous years of REACH U.S. data collection; thus, some of the households did not undergo explicit accuracy verification. While this strategy was justified given project objectives of screening households efficiently and eliminating break-offs, it is a limiting factor for this study.

Finally, we limited the scope of this paper to predicting the efficiency of CATI. More complex analyses will include both mail and CATI and, therefore, can help predict which mode will have the highest predicted probability of success. It was not a feasible analysis here as cases that were accessible in CATI were never moved to mail and no telephone matched cases were initially fielded in mail.

The types of models shown here are flexible and can be tailored to refine survey planning. They can also be used as starting points for overall cost-efficiency models including multiple modes. This approach can be used to help determine thresholds for the proportion of telephone matched cases that are initially fielded in CATI versus other modes. Finally, this strategy can also be adapted to identify conditions under which it is appropriate to move cases from their initial mode to another and to predict what proportion of cases will change modes during data collection. We plan to explore many of these potential uses and investigate how mode decisions may affect demographic distributions and estimates of key variables.

References Cited

Amaya, A., Skalland, B., & Wooten, K. (2010). What's in a match. *Survey Practice*, December. Available from: www.surveypractice.org.

Bailey, J. T., Grabowski, G., & Link, M. W. (2010). Your home was specially selected: Using address based sampling as a recruitment technique. *Proceedings of the American Statistical Association, Section on Survey Research Methods* (pp. 5938-5948). Alexandria, VA: American Statistical Association.

Blumberg, S. J., & Luke, J. V. (2012). Wireless substitution: Early release of estimates from the National Health Interview Survey, January-June 2012. National Center for Health Statistics. December 2012. Available from: http://www.cdc.gov/nchs/nhis.htm.

Brick, J. M., Allen, B., Cunningham, P., & Maklan, D. (1996). Outcomes of a calling protocol in a telephone survey. *Proceedings of the American Statistical Association, Section on Survey Research Methods* (pp. 142-149). Washington, DC: American Statistical Association.

Brick, J. M., Williams, D., & Montaquila, J. M. (2011). Address-based sampling for subpopulation surveys. *Public Opinion Quarterly,* 75(3): 409-428.

Bruce, A., & Robinson, J. G. (2009). Tract level planning database with Census 2000 data. Available from: http://www.census.gov/2010census/partners/pdf/TractLevelCensus2000Apr_2_09.pdf.

Curtin, R., Presser, S., & Singer, E. (2005). Changes in telephone survey non-response over the past quarter century. *Public Opinion Quarterly,* 69(1): 87-98.

De Leeuw, E. D., & de Heer. W. (2002). Trends in household survey nonresponse: A longitudinal and international comparison. In R. M. Groves, D. A. Dillman, J. L. Eltinge, & R. J. A. Little (Eds.), *Survey Nonresponse* (pp. 41-54). New York: Wiley.

de Leeuw, E. D. (2008). Choosing the method of data collection. In E. D. de Leeuw, J. J. Hox, & D. A. Dillman (Eds.), *International Handbook of Survey Methodology* (pp. 113-135). New York: Lawrence Erlbaum/Psychology Press, Taylor and Francis Group.

de Leeuw, E. D. (2005). To mix or not to mix data collection modes in surveys. *Journal of Official Statistics,* 21(2): 233-255.

Durrant, G. B., D'Arrigo, J., & Steele, F. A. (2011). Using field process data to predict best times of contact conditioning on household and interviewer influences. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 174(4): 1-21.

Groves, R. M., & Heeringa, S. (2006). Responsive design for household surveys: Tools for actively controlling survey errors and costs. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 169(3): 439-457.

Hoffer, T. B., Grigorian, K., & Fecso, R. (2007). *Assessing the effectiveness of using panel respondent mode preference data*. Paper presented at the 2007 Joint Statistical Meetings, Salt Lake City, Utah.

Link, M. W., Daily, G., Shuttles, C. D., Bourquin, H. C., & Yancey, L. T. (2009). Addressing the cell phone-only problem: Cell phone sampling versus address based sampling. *Survey Practice*, February. Available from: www.surveypractice.org.

Medway, R., & Fulton, J. (2012). When more gets you less: A meta-analysis of the effect of concurrent web options on mail survey response rates. *Public Opinion Quarterly*, 76(4): 733-746.

Messer, B., & Dillman, D. (2011). Surveying the general public over the internet using address-based sampling and mail contact procedures. *Public Opinion Quarterly*, 75(3): 429-457.

Millar, M., & Dillman, D. (2011). Improving response to web and mixed-mode surveys. *Public Opinion Quarterly*, 75(2): 249-269.

Office of Management and Budget. (2009). Update of statistical area definitions and guidance on their uses. *OMB Bulletin,* no. 10-02. December. Available from: http://www.whitehouse.gov/sites/default/files/omb/assets/bulletins/b10-02.pdf.

Olson, K., Smyth, J. D., & Wood, H. M. (2012). Does giving people their preferred survey mode actually increase survey participation rates? An experimental examination. *Public Opinion Quarterly*, 76(4): 611-635.

Selfa, L., & Sederstrom, S. (2006). *Respondent mode preference in a multi-mode survey*. Paper presented at the 61st Annual Conference of the American Association for Public Opinion Research, Montreal, Quebec.

Smyth, J., Dillman, D., Christian, L., & O'Neill, A. (2010). Using the internet to survey small towns and communities: Limitations and possibilities in the early 21st century. *American Behavioral Scientist*, 53(9): 1423-1428.