

# Modeling the optimal contact mode in the Current Population Survey November 2016

John Dixon<sup>1</sup>, Brian Meekins<sup>2</sup>, Polly Phipps<sup>1</sup>

<sup>1</sup>Bureau of Labor Statistics

<sup>2</sup>AARP

## Abstract

The Current Population Survey (CPS) is designed to measure labor force characteristics of the United States, as well as other population characteristics. Households are interviewed eight times; four consecutive months, followed by an eight month break, then four more consecutive months. Interviewers contact households in person for the first interview and subsequent interviews are often done by telephone. Because of the eight-month break, conducting the fifth interview in person is encouraged; however, 40 percent are conducted by telephone. This study uses paradata from the Contact History Instrument (CHI), and demographic and household characteristics to determine which households are most likely to be successful using telephone data collection in the fifth interview. A successful interview is one that collects sufficient labor force information. A logistic model was used to identify the respondent and household characteristics which predict the best mode for conducting the interview in the fifth interview. The model included interviewer effort, differences in regions, age, race, ethnicity, gender, labor force status of the previous respondent (from the 4<sup>th</sup> interview), and respondent concerns (from the previous 4 interviews) to explore the impact of mode on nonresponse. The best mode for a successful household interview will be modeled using a competing risk model where interviewer effort and likelihood of nonresponse are minimized. The results can be used by managers and interviewers to decide which mode should be used.

**Key Words:** survey, mode of data collection, number of contacts, nonresponse

## 1. Introduction

The cost of surveys to collect information from households has steadily risen over time. The motivation of this study is to improve the use of telephone to replace personal visits to reduce costs while maximizing data quality and response rates.

The CPS is the primary source of information on the labor force characteristics of the U.S. population.<sup>1</sup> Some characteristics of the survey design and administration include:

- The CPS consists of 8 separate interviews spread out over a 16 month period using a complex sample rotation design.
- The data collection period for the CPS is for 10 days.
- Months 1 and 5 (with an 8-month break in between) are designed to be in-person interviews. However, about 24 percent of month 1 and 44 percent of

---

<sup>1</sup> Details about the CPS can be found in Technical Paper 66 (<http://www.census.gov/prod/2006pubs/tp-66.pdf>).

month 5 personal visit cases are collected by field interviewers over the telephone.

- Months 2, 3, 4, and 6 through 8 are designed to be telephone interviews, either conducted by field interviewers in a decentralized manner (about 66 percent) or sent to a telephone call center (about 10 percent). The remaining 24 percent are conducted in person.

The interviewers use the Contact History Instrument (CHI) to record each attempted contact with the respondent. On the screen shown in Figure 1, the interviewers check all that apply to note respondent concerns. While the consistency of use probably varies between interviewers, it still provides valuable information about respondent reactions to the survey request. The concerns most often selected are “Not interested” or “Too busy”, but a wide range of concerns have been recorded by interviewers over the years. A similar screen is available for contactability issues.

Contact History Instrument v5.8.5 Created 08/25/2004

Forms Answer Navigate Options Help

CHI

**CONCERN / BEHAVIOR / RELUCTANCE**

Select the categories that describe respondent concerns, behaviors, or reluctance during this contact attempt.

Enter all that apply, separate with commas.

<input type="checkbox"/> 1. Not interested / Does not want to be bothered	<input type="checkbox"/> 12. Hostile or threatens FR
<input type="checkbox"/> 2. Too busy	<input type="checkbox"/> 13. Other household members tell respondent not to participate
<input type="checkbox"/> 3. Interview takes too much time	<input type="checkbox"/> 14. Talk only to specific household member
<input type="checkbox"/> 4. Breaks appointments (puts off FR indefinitely)	<input type="checkbox"/> 15. Family issues
<input type="checkbox"/> 5. Scheduling difficulties	<input type="checkbox"/> 16. Respondent requests same FR as last time
<input type="checkbox"/> 6. Survey is voluntary	<input type="checkbox"/> 17. Gave that information last time
<input type="checkbox"/> 7. Privacy concerns	<input type="checkbox"/> 18. Asked too many personal questions last time
<input type="checkbox"/> 8. Anti-government concerns	<input type="checkbox"/> 19. Too many interviews
<input type="checkbox"/> 9. Does not understand survey / Asks questions about the survey	<input type="checkbox"/> 20. Last interview took too long
<input type="checkbox"/> 10. Survey content does not apply (retired, healthy, no crimes to report)	<input type="checkbox"/> 21. Intends to quit survey
<input type="checkbox"/> 11. Hang-up / slams door on FR	<input type="checkbox"/> 22. No concerns
	<input type="checkbox"/> 23. Other - specify

**Figure 1. The Contact History Instrument (CHI) Respondent Concerns Screen**

## 2. Study Design and Findings

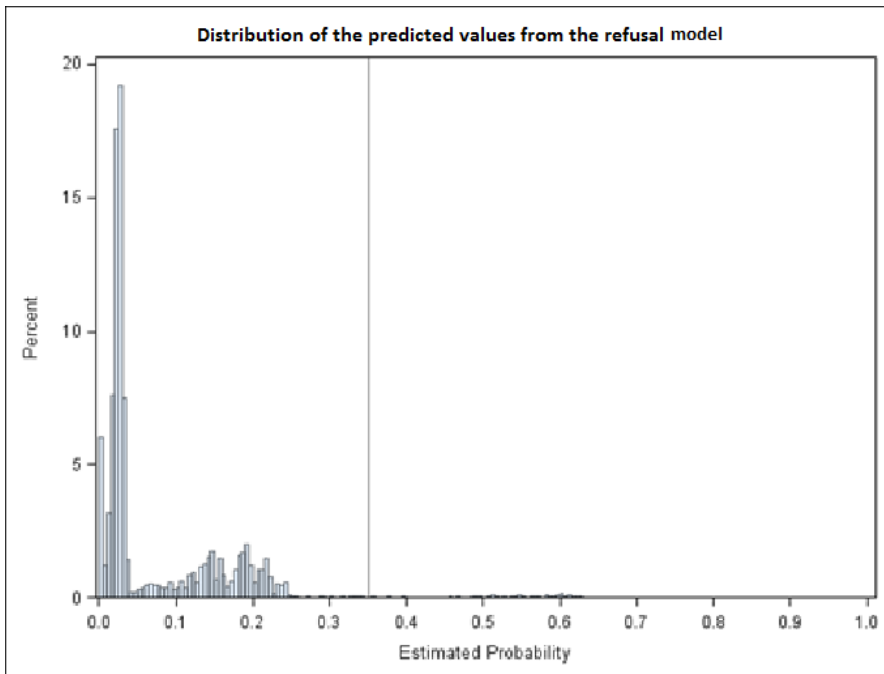
The focus of this study is whether the fifth interview (MIS 5) could be collected primarily by telephone, and if so to identify the characteristics of households best suited for that mode. To study this, we looked at households who had their first month in the survey between April and July 2013. We then used data for each subsequent month they are in the sample through July. Ultimately, 4,195 unique households were selected from the selected panels between May and July. The data were adapted from previous studies (Phipps et al., 2015 and Meekins and Phipps, 2014).

Two models were used in this study. The first model was used to provide propensity scores for nonresponse which were then used in the second model to study the effect of nonresponse on the mode choice. This model is described below. The second model has the mode of interview from the fifth interview as the dependent variable, and included CPS data from previous interviews. The demographic characteristics include Employment

status, Gender, Hispanic ethnicity, Age, and Race. Household characteristics include Tenure (Own/rent), Urban/Rural, Previous successful CATI interview, and Previous nonresponse.

### 2.1 Nonresponse measurement – Propensity scores

Logistic models were used to produce propensity scores for refusal and noncontact, using predictors from the CHI. Contact History Instrument (CHI) responses are used to categorize responders to the CPS as similar to nonresponders based on their contact history and reasons for not responding. Using CHI variables related to the attempts to contact the household we also identified difficult to contact respondents to represent noncontacts. Propensity scores are predicted values from a logistic model based on the CHI. These are used to estimate nonresponse bias.



**Figure 2:** Distribution of the refusal propensity scores.

This histogram shows the predicted values from the logistic regression. Most households are not like those who refused, since they didn't express any concerns. The reference line around .35 divides the sample into those who are most like the refusers and those who are not. The reference is chosen to select the same proportion as those who refused that month. So if for one month, 8% refused, 8% would be selected from the responders to represent the refusers. A similar procedure was carried out for noncontact, resulting in a group of respondents who will be used as proxies for nonrespondents, and another group who will be used as proxies for noncontacts.

### 2.2 Logistic models

To model a binary variable (successful (completed) telephone interview), we use;

$$\ln\left(\frac{\hat{p}}{(1-\hat{p})}\right) = b_0 + b_1X_1 + b_2X_2 + \dots + b_pX_p$$

Where  $\hat{p}$  is the expected probability that the outcome is observed (successful telephone interview);  $X_1$  through  $X_p$  are distinct independent variables; and  $b_0$  is the intercept, and  $b_1$  through  $b_p$  are the regression coefficients.

**Table 1: Logistic model predicting a successful telephone 5<sup>th</sup> interview**

Parameter	DF	Estimate	SE	Chi-square	Pr>	Odds ratio
Intercept	1	0.1611	0.0597	7.2825	0.0070	
Prev personal visit	1	-0.9981	0.0159	3946.088	<.0001	0.369
Prev phone visit	1	0.5547	0.0127	1922.542	<.0001	1.741
Education	1	0.1271	0.0108	137.761	<.0001	1.135
Own home	1	-0.1393	0.0134	107.847	<.0001	0.757
Hispanic	1	0.1721	0.0186	85.485	<.0001	1.411
Child at home	1	-0.1608	0.0282	32.584	<.0001	0.851
White	1	0.1413	0.0313	20.429	<.0001	1.152
Completed CATI	1	0.4540	0.1239	13.424	0.0002	1.575
Predicted refusal	1	-0.1025	0.0282	13.187	0.0003	0.903
Age	1	0.0353	0.0121	8.4378	0.0037	1.036
Predicted noncontact	1	0.0865	0.0335	6.6645	0.0098	1.090
Not working	1	-0.0296	0.0134	4.877	0.0272	0.943
Urban	1	-0.0208	0.0257	0.656	0.4180	0.979
Female	1	-0.0032	0.0121	0.070	0.7912	0.994

Table 1 shows the results of the logistic regression. The predictors of the mode of completion for the 5<sup>th</sup> interview (month in sample; MIS 5) are ordered by their strength of prediction (based on the Chi-square value). The previous successful mode was the strongest predictor of the successful mode in 5<sup>th</sup> interview. The odds ratio shows lower odds of a telephone interview if the previous interview was in person, and much higher odds if the previous interview was by telephone. Renters and households with children at home were less likely to complete the interview by telephone, while higher educated, Hispanic, White, and older respondents, and those who completed a CATI interview were more likely to complete by telephone. Those respondents similar to respondents who refused were less likely to be interviewed by telephone, while those who were like those who were difficult to contact were more likely to participate in a telephone interview. The respondents who had scores similar to those who refused may need more personal interaction to overcome their reluctance, while those who are difficult to contact may have scheduling issues, which may be best handled by a telephone interview. The higher educated tend to respond to surveys more, which may make them easier to interview by telephone. Hispanics may be easier to interview by phone if the interviewer has good report and can interview in the respondents preferred language. Households with young children at home tend to be home more, and so may be easier to contact. Households which had previously refused may need greater persuasion, possibly involving a personal visit. Once they were persuaded, it would be easiest for the interviewer to interview quickly before they changed their mind.

### **3.1 Suggestions**

Based on the study results, more telephone interviews should be attempted for respondents who are;

- Households which were difficult to contact, but were more cooperative.
- Higher educated,
- Renters,
- Hispanic, or White,
- Renters,
- Households which didn't have young children, and

### **3.2 Limitations**

The current study was not an experiment, but reflects the current procedures, so some effects may be due to survey field practices, rather than actual relationships between the respondent characteristics and successful telephone interviews. For example, young children at home may mean the household is likely to be easier to contact, making personal visits at the 5<sup>th</sup> interview easier, but it might be just as easy to conduct a telephone interview. Similarly, some groups (based on demographics or household characteristics) may have a higher proportion of nonrespondents, so the relationship of those groups might be masked by nonresponse in predicting successful telephone interviews.

### **3.3 Future research**

Possible topics for future research to evaluate mode success and cost reduction include:

- Efforts to find ways to reduce survey costs while maintaining the quality of estimates from the survey. Models to examine the trade-offs between the number of attempted contacts for different modes could help to reduce costs, although reducing contact attempts would increase nonresponse. Models to determine what type of households would be best to start by telephone contact could reduce travel costs. Currently the interviewers must choose which ones are most likely, and having some guidance may be welcome.
- Models to determine when in the interview process to change from telephone to personal visit could reduce cost while maintaining response rates. Currently the interviewers must decide when to attempt a telephone interview.
- Describing the characteristics of 1<sup>st</sup> interview telephone households could help guide more efficient data collection. While we know little about households before the first contact, there are neighborhood and housing characteristics which are available from the census planning database and other sources.
- Models to determine if there are different attrition patterns for telephone interviews compared to in person interviews could help maintain response rates while maintaining data quality.

## **References**

Phipps, P., Yang, D., Meekins, B., and Toth, D., "Investigating the Effect of Mode Assignment on the Response Rate in the Current Population Survey", Paper presented at the Joint Statistical Meetings, 2015.

Meekins, B. and Phipps, P., "In-Person or Telephone Collection: Mode Selection and Outcomes in the Current Population Survey", Presented at the American Association for Public Opinion Research conference, May 2014, Anaheim California.