

Catalogue no. 11-522-XIE

**Statistics Canada International  
Symposium Series - Proceedings**

**Symposium 2005 :  
Methodological Challenges for  
Future Information needs**



2005



**Statistics  
Canada**

**Statistique  
Canada**

**Canada**

## MODELING SURVEY CONTACT IN THE NATIONAL HEALTH INTERVIEW SURVEY (NHIS)

James M. Dahlhamer, Barbara J. Stussman, Catherine M. Simile, Beth Taylor<sup>1</sup>

### ABSTRACT

Using core survey, frame, and contact history data collected with the 2005 NHIS, a multi-purpose health survey conducted by the National Center for Health Statistics (NCHS), Centers for Disease Control and Prevention (CDC), a model of initial contact was developed and tested. Attempt-level, household-level, and social environmental measures were all found to influence initial contact with sample households. Among the key predictors were day and time of current contact attempt; mode of current contact attempt; days since last contact attempt; encountering an access impediment on a prior attempt; the presence of elderly adults; number of adults; region of residence; and urbanicity (metropolitan statistical area status). The implications for survey procedures and field operations are discussed.

KEY WORDS: survey contact; nonresponse; contact history data

### 1. INTRODUCTION

Like other government surveys, the National Health Interview Survey (NHIS) has witnessed a decline in household response rates over the past 10 to 15 years (from 95.5% in 1990 to 86.9% in 2004). One of the hypothesized reasons for this decline is the reduced accessibility of households, arising, in part, from increased physical control of access to housing units and a growth in household compositions in which no one is home for long periods of time (Groves and Couper, 1998). Consistent with this notion, noncontact rates in the NHIS have climbed from 1.5% in 1990 to 3.7% in 2004. And while refusal rates are comparatively higher (7.4% in 2004), obtaining an interview is a two-step process whereby an interviewer must first make contact with a sample household. If acceptable rates and quality of response are to be achieved, interviewers need to be highly efficient in contacting sample households so as to leave ample time for converting reluctant respondents.

This paper explores factors related to initial contact with NHIS sample households. Using core survey, frame, and contact history data collected with the first three quarters of the 2005 NHIS, a three-component model of initial contact is developed and tested. In addition to key social-environmental and household-level measures, analyses explore the impact of attempt-level influences, particularly those under the control of interviewers (e.g., day and time of contact attempt, the use of reconnaissance strategies, lag time between attempts). The goal is to identify attributes of difficult-to-contact households, and the strategies for improving accessibility, so that survey procedures may be adjusted to improve the efficiency of field operations.

### 2. SURVEY CONTACT

According to Groves and Couper (1998), contactability is a function of two factors: the accessible at-home patterns of households and the number and timing of contact attempts. The accessible at-home patterns of households are influenced by physical impediments to gaining access, socio-demographic and lifestyle attributes of householders, and the larger social environment in which sample households reside. In face-to-face surveys, physical impediments include gated communities, buzzer entries, and door persons, all of which have been found to increase the effort required to make contact with households (Callens and Croux, 2004; Groves and Couper, 1998). Among the various socio-demographic and lifestyle indicators found to reduce accessibility were single-person households

---

<sup>1</sup> James M. Dahlhamer ([fzd2@cdc.gov](mailto:fzd2@cdc.gov)), Barbara J. Stussman ([bjs6@cdc.gov](mailto:bjs6@cdc.gov)), Catherine M. Simile ([cus4@cdc.gov](mailto:cus4@cdc.gov)), and Beth Taylor ([bft8@cdc.gov](mailto:bft8@cdc.gov)), National Center for Health Statistics, 3311 Toledo Road, Hyattsville, Maryland, USA, 20782.

(Groves and Couper, 1998; Purdon et al., 1999) and households where all adults were employed (Callens and Croux, 2004). Conversely, households with elderly persons and young children were found to be more accessible (Groves and Couper, 1998; Purdon et al., 1999). Finally, among the common social-environmental influences explored in the literature are measures of urbanicity and population density. In general, more highly urbanized areas were associated with lower contact rates (Callens and Croux, 2004; Groves and Couper, 1998), although this is partially explained by these areas being characterized by less accessible household types (multi-unit structures, single-person households, renter-occupied units), longer commute times, and greater entertainment options (Groves and Couper, 1998).

A growing number of studies have focused on the second factor of the Groves and Couper model, especially how the timing of contact attempts is related to contactability. Results for face-to-face surveys (largely bivariate analyses) reveal that weekday evening hours are highly productive for initial contact (Bates, 2003; Dahlhamer et al., 2005; Purdon et al., 1999). Weekends as a whole have been found to improve contactability, although few studies have distinguished between morning, afternoon, and evening hours (Bates, 2003; Dahlhamer et al., 2005; Groves and Couper, 1998). Conversely, weekday mornings and afternoons are particularly poor times for making initial contact with sample units (Bates, 2003; Dahlhamer et al., 2005; Groves and Couper, 1998; Purdon et al., 1999). What are often missing from these analyses, however, are other attempt-level interviewer strategies for enhancing the odds of contact. For example, interviewers report “staking-out” households and talking to neighbors in efforts to identify likely at-home patterns. Other decisions, such as how long to wait between attempts, are also under the control of interviewers. We attempt to address this void in the literature by incorporating some of these overlooked measures in our model of initial contact.

### 3. METHODS

The analyses presented here rely on data collected during the first three quarters of the 2005 NHIS. The NHIS is an annual survey of the health of the civilian, noninstitutionalized household population of the United States conducted by the National Center for Health Statistics (NCHS), Centers for Disease Control and Prevention (CDC). Data are collected by trained interviewers with the U. S. Census Bureau using computer assisted personal interviewing (CAPI). The attempt-level, contact history data are collected via the automated Contact History Instrument (CHI). Interviewers use CHI to record information on each contact attempt, including the outcome of each attempt and any contact or cooperation-based strategies employed.

Analyses were restricted to eligible and “screened out” households. Screened-out households were those flagged as part of the oversampling of black and Hispanic populations, but were found, after collecting household roster information, to contain neither a black nor a Hispanic household member. While screened-out households are not included in calculations of final response rates, interviews must contact them to screen them out of the sample. A total of 117,102 contact attempts were recorded for 35,633 eligible and screened-out households for the time period under analysis.<sup>2</sup>

To assess the likelihood of making initial contact at a particular contact attempt, and allow for the inclusion of time-dependent independent variables (e.g., day and time of attempt), discrete-time logistic regressions were performed (see Allison, 1982).<sup>3</sup> With discrete-time logistic regression, the probability that initial contact will be made at time (or attempt)  $t$  given that no contact occurred before time (or attempt)  $t$  (often referred to as a discrete-time hazard) is modeled as a function of a set of independent variables. To perform the analyses, each household’s contact history was broken down into a set of discrete time units (contact attempts) that were treated as distinct records or observations in the dataset. By doing this, discrete-time logistic regression models can be fitted easily within any software package offering logistic regression. Essentially, binary logistic regression models predicting whether initial contact was or was not made at each contact attempt were estimated, and selected independent variables were allowed to vary over time (or attempts).

---

<sup>2</sup> Weeks one through nine of quarter one are not included in the analyses due to a programming error with the Contact History Instrument (CHI).

<sup>3</sup> Discrete-time logistic regression is one of several techniques that comprise a class of statistical methods known as survival or event-history analysis

Independent variables for this analysis were broken into three components: attempt-level measures, social environmental measures, and household-level measures (see Table 1 for variable coding). Attempt-level measures included number of previous attempts, day and time of current/previous attempt, mode of current attempt, days since last attempt, whether or not the interviewer used a reconnaissance strategy (e.g., “stake-outs,” checked with neighbors) on the prior attempt, whether or not the interviewer left a note or appointment card on the prior attempt, and whether or not the interviewer encountered an access impediment (e.g. locked gate, buzzer entry) on a prior attempt. Social environmental measures included U.S. Census Bureau region of residence and metropolitan statistical area (MSA) status (a measure of population density defined by the U. S. Census Bureau). Finally, household-level measures included whether or not someone in the household was over the age of 65, whether or not someone in the household was under the age of 5, and the number of adults. The dependent variable, initial contact, was coded 0 if initial contact was not made with a sample unit member during the current attempt or 1 if initial contact was made with a sample unit member during the attempt. Contacts with non-sample unit members, such as baby-sitters, were coded 0. Only contact attempts up through first or initial contact were included in the analysis.

All analyses were performed in SUDAAN (version 9.0, Research Triangle Institute, Inc., Research Triangle Park, NC). Since no attempt is being made to generalize findings beyond the NHIS, analyses presented in Table 1 are unweighted. However, characteristics of the clustered sample design were utilized to produce appropriate standard errors.

#### 4. RESULTS

To include the household-level measures, analyses were performed with a subset of eligible and screened-out households that provided roster or demographic data. To guard against possible selection biases, a model incorporating just the attempt-level and social environmental measures was applied to all eligible/screened-out households (“all households”) and the subset of households providing roster data (“households with roster data”). The results (first two columns of findings in Table 1) were then compared. Only one minor difference was observed involving the previous day and time of attempt measure (see Table 1). The results presented below, therefore, are from the full model incorporating household-level measures (third column of findings in Table 1). A total of 65,933 attempts were included in this analysis, with 34,577 (52.5%) resulting in noncontact and 31,351 (47.5%) resulting in initial contact.

Controlling for a variety of attempt-level, household-level, and social environmental measures, a strong impact of current day and time of attempt was observed (see Table 1). Compared to weekday afternoon attempts, attempts during weekend mornings, weekend afternoons, evening hours for Sunday through Thursday, and Friday or Saturday evening all increased the odds of initial contact at the current attempt. Conversely, attempts during weekday mornings reduced the odds of initial contact. While smaller in magnitude, some significant effects of prior day and time of attempt also emerged. Compared to previous weekday afternoon attempts, previous attempts during weekend mornings and Friday-Saturday evenings reduced the odds of initial contact at the current attempt, suggesting that failed attempts during otherwise lucrative time slots are likely indicative of overly difficult to reach households. As for number of previous attempts, a nine percent reduction in the odds of initial contact occurred for each additional contact attempt.

Surprisingly, current attempts by telephone (versus in-person attempts) increased the odds of initial contact, a finding warranting further investigation. Waiting longer after a failed attempt also increased the odds of initial contact at the current attempt. While waiting 1-3 days had no effect, compared to same day next attempts, waiting four or more days increased the odds of initial contact by 55%. This likely reflects the acquisition of information by interviewers that household members would be away for the allotted time between attempts.

As for other interviewer strategies, leaving notes and/or appointment cards or engaging in reconnaissance, such as talking to neighbors and “staking-out” households, was not associated with the odds of making initial contact. And finally, as anticipated, encountering an access impediment on a prior attempt reduced the odds of initial contact.

**Table 1. Results from Discrete-Time Logistic Regressions Predicting Whether or Not Initial Contact Was Made at the Current Contact Attempt**

Independent Variables	Models Excluding Household-Level Measures		
	All Households (n=77,188 attempts)	Households with Roster Data (n=65,928 attempts)	Households with Roster Data (n=65,928 attempts)
	Odds Ratio	Odds Ratio	Odds Ratio
<b><i>Attempt-Level Measures</i></b>			
<b>First Attempt</b>			
Yes (versus no)	1.43**	1.53**	1.50**
<b>Day and Time of Current Attempt</b>			
Saturday-Sunday morning	1.36**	1.38**	1.43**
Monday-Friday morning	0.91**	0.91*	0.91*
Saturday-Sunday afternoon	1.39**	1.40**	1.44**
Monday-Friday afternoon <sup>1</sup>	1.00	1.00	1.00
Sunday-Thursday evening	1.48**	1.53**	1.58**
Friday-Saturday evening	1.29**	1.32**	1.36**
<b>Mode of Current Attempt</b>			
Telephone (versus in-person)	1.63**	1.65**	1.65**
<b>Day and Time of Last Attempt</b>			
Saturday-Sunday morning	0.81**	0.81**	0.81**
Monday-Friday morning	1.02	1.01	1.01
Saturday-Sunday afternoon	0.91	0.93	0.94
Monday-Friday afternoon <sup>1</sup>	1.00	1.00	1.00
Sunday-Thursday evening	0.90**	0.94	0.94
Friday-Saturday evening	0.78**	0.80**	0.79**
<b>Number of Prior Attempts</b>	0.87**	0.90**	0.91**
<b>Days Since Last Attempt</b>			
Same day <sup>1</sup>	1.00	1.00	1.00
1 day	1.01	1.00	1.01
2-3 days	1.05	1.09	1.10
4+ days	1.40**	1.53**	1.55**
<b>Interviewer Used Reconnaissance Strategy on Last Attempt</b>			
Yes (versus no)	0.99	1.03	1.06
<b>Interviewer Left Note/Appt. Card on Last Attempt</b>			
Yes (versus no)	1.02	1.04	1.05
<b>Encountered an Access Impediment on a Prior Attempt</b>			
Yes (versus no)	0.80**	0.82**	0.83**
<b><i>Social Environmental Measures</i></b>			
<b>Region of Residence</b>			
Northeast	0.87**	0.88*	0.87**
Midwest	0.86**	0.82**	0.84**
South	0.94	0.94	0.95
West <sup>1</sup>	1.00	1.00	1.00

**Table 1. continued**

Independent Variables	Models Excluding Household-Level Measures		
	All Households (n=77,188 attempts)	Households with Roster Data (n=65,933 attempts)	Households with Roster Data (n=65,933 attempts)
	Odds Ratio	Odds Ratio	Odds Ratio
<b>MSA Status</b>			
MSA, central city	0.70**	0.73**	0.76**
MSA, non-central city	0.82**	0.85**	0.85**
Non-MSA <sup>1</sup>	1.00	1.00	1.00
<b>Household-Level Measures</b>			
<b>Someone in Household Over 65</b>			
Yes (versus no)			1.75**
<b>Someone in Household Under 5</b>			
Yes (versus no)			1.25**
<b>Number of Adults</b>			
1 adult <sup>1</sup>			1.00
2 adults			1.38**
3+ adults			1.63**
<i>Intercept</i>	0.88**	0.86**	0.59**

\* .01 &lt; p &lt; .05

\*\* p &lt; .01

<sup>1</sup> Reference category for the odds ratio (OR=1.00)

Both social environmental measures were significantly associated with initial contact (see Table 1). Compared to households in the West, households in the Northeast and Midwest were associated with lower odds of initial contact, while households in MSA, central cities and MSA, non-central cities had reduced odds of initial contact compared to households in non-MSA areas. Not surprisingly, NHIS households in more urbanized areas are less accessible.

As anticipated, households with two or three or more adults had higher odds of initial contact compared to single-adult households. In addition, households with persons over the age of 65 had odds of initial contact nearly two times higher than for households with all members 65 years of age or younger, a finding consistent with the larger literature. And finally, households with children under the age of five were easier to contact for the first time than households without young children.

## 5. DISCUSSION

Overall, we found support for all three components of our model. Of particular interest were the attempt-level findings, as we included measures tapping strategies, tactics, and decisions under the control of interviewers. We found strong support for the importance of day and time of attempt in making contact with sample households, net of other attempt-level, social environmental, and household-level influences. This was significant not only because most research to date had relied on bivariate analyses, but because a recent analysis showed that NHIS interviewers are making a significant number of attempts during lucrative weekday evening and weekend time slots, especially after initial attempts result in noncontact (Dahlhamer et al., 2005). Interestingly, while prior day and time of attempt also impacted initial contact, no significant interactions emerged between the two day and time measures in separate examinations. While this limits the potential number of day and time strategies for interviewers, it greatly reduces the complexity of the decision-making process. In sum, training protocols should introduce or continue to reinforce the importance of day and time of contact attempt as an important interviewer strategy.

The significant, positive effect of making attempts by telephone was highly surprising. Would initial attempts by telephone greatly reduce the number of attempts required to make first contact, and, in turn, significantly reduce field costs? Possibly, but contact by telephone still requires a face-to-face meeting to complete the interview. Furthermore, we know little at this point about the quality and outcomes of telephone contacts. First contacts by telephone may increase the likelihood of interim and final refusals. Clearly, further investigation is warranted.

That no significant effects of interviewer strategies such as checking with neighbors and leaving notes/appointment cards emerged is somewhat troubling. However, the current analytic approach may not be appropriate for determining the utility of such tactics. A better test would involve a subset of difficult-to-reach households. For example, what impact would the use (versus non-use) of such strategies have on the odds of initial contact among households where at least the first four attempts resulted in noncontact? Plans for such analyses are under way.

The findings for region of residence and MSA status highlight the importance of contextual or social environmental measures for shaping and understanding survey processes and outcomes. However, the aggregate measures we employed have minimal practical applications. For example, the New York regional office (U. S. Census Bureau) largely covers urbanized areas, so the finding for MSA status is meaningless from a planning and operational perspective. What are needed are analyses that incorporate more precise, social environmental measures at the neighborhood-level. Findings from such analyses could have direct benefit to regional offices in developing more targeted strategies for enhancing contact and cooperation. Geocoding efforts are currently underway so that tract-level data from the 2000 Census can be utilized.

Finally, we identified strong influences of household-level measures such as the number of adults and the presence of elderly adults. Consistent with recommendations of Groves and Couper (1998), we suggest the development of a protocol whereby sample listers record information on housing units, such as presence of wheelchair ramps or multiple cars in the driveway, indicative of household accessibility. Information of this sort could be fed into the Case Management system (system for accessing, managing, and transmitting cases) and used by interviewers to prioritize their weekly caseloads.

## REFERENCES

- Allison, P. (1982), "Discrete-Time Methods for the Analysis of Event Histories." in *Sociological Methodology*, San Francisco: Jossey-Bass, pp. 61-98.
- Bates, N. (2003), "Contact Histories in Personal Visit Surveys: The Survey of income and Program Participation (SIPP) Methods Panel," paper presented at the Annual Conference of the American Association for Public Opinion Research, Nashville, Tennessee.
- Callens, M. and C. Croux. (2004), "Contact and Cooperation in the Belgian Fertility and Family Survey," *Survey Methodology*, 30(1), pp. 115-124.
- Dahlhamer, J. M., C. M. Simile, B. J. Stussman, B. Taylor. (2005), "Determinants and Outcomes of Initial Contact in the National Health Interview Survey, 2004," paper presented at the 2005 Annual Conference of the American Association for Public Opinion Research, Miami Beach, Florida.
- Groves, R. M., and M. P. Couper. (1998), *Nonresponse in Household Interview Surveys*, New York: John Wiley & Sons, Inc.
- Purdon, S., P. Campanelli, and P. Sturgis. (1999), "Interviewers' Calling Strategies on Face-to-Face Interview Surveys," *Journal of Official Statistics*, 15(2), pp. 199-216.