OUTCOMES OF A CALLING PROTOCOL IN A TELEPHONE SURVEY

J. Michael Brick, Bruce Allen, Pat Cunningham, and David Maklan
J. Michael Brick, Westat, Inc., 1650 Research Blvd., Rockville, MD 20850

Key Words: Random digit dialing survey, CATI, call scheduling

1. Introduction

The technology of computer assisted telephone interviewing (CATI) has greatly enhanced the ability to conduct interviews in ways that are not feasible with paper and pencil interviews. These methods have enabled researchers to improve the quality of responses by introducing computerized skip patterns, customized interview wording, and real-time edit checks. As a result, survey developers have been able to create interviews that are both more specific to the respondent and of higher quality.

Another feature of CATI is the potential to improve methods of scheduling interviews to make the interviews more cost-efficient and reduce nonresponse bias in estimates from telephone surveys. These methods depend upon call scheduling capabilities that are truly feasible only in a CATI environment. Despite its importance, the literature on this application of CATI is not very extensive. This paper analyzes data with the goal of developing scheduling procedures to reduce data collection cost and nonresponse bias. The focus is on random digit dial (RDD) household surveys, but the results may also apply to some list samples.

Figure 1 presents a conceptual framework for the task of conducting a CATI survey, with inputs, processes under the control of the survey organization, and outcomes. The inputs are the telephone numbers themselves, the results of previous call attempts, and characteristics of the numbers, such as the geographic area the numbers cover. The process that can be manipulated to affect the outcomes are: the calling protocols or procedures used to schedule telephone calls, the workforce size and availability, the data collection period, and the characteristics of the interviewers. The ability to organize all of these processes effectively is an important determinant of survey costs and quality. The outcomes in the figure include completed and refused interviews, contacts that result in call-back attempts and noncontacts (unanswered).

While all of the processes that affect the outcome are of interest in conducting CATI surveys, this paper concentrates on the influence of calling protocols. Other factors, such as the number of call attempts, are not discussed. One of the first descriptions of calling protocols for CATI surveys was presented by Weeks (1988). Many of the protocols he described use information based on overall patterns of people being at home and do not take full advantage of the capabilities of the computer. Weeks, Kulka, and Pierson (1987) provided some data on times that people are home and respond to surveys that could be used in devising strategies for calling. Groves (1989) and Kulka and Weeks (1989) suggested the conditional probability of contacting a household might be dependent on distribution of previous call attempts and using this call history might improve the efficiency of the survey.

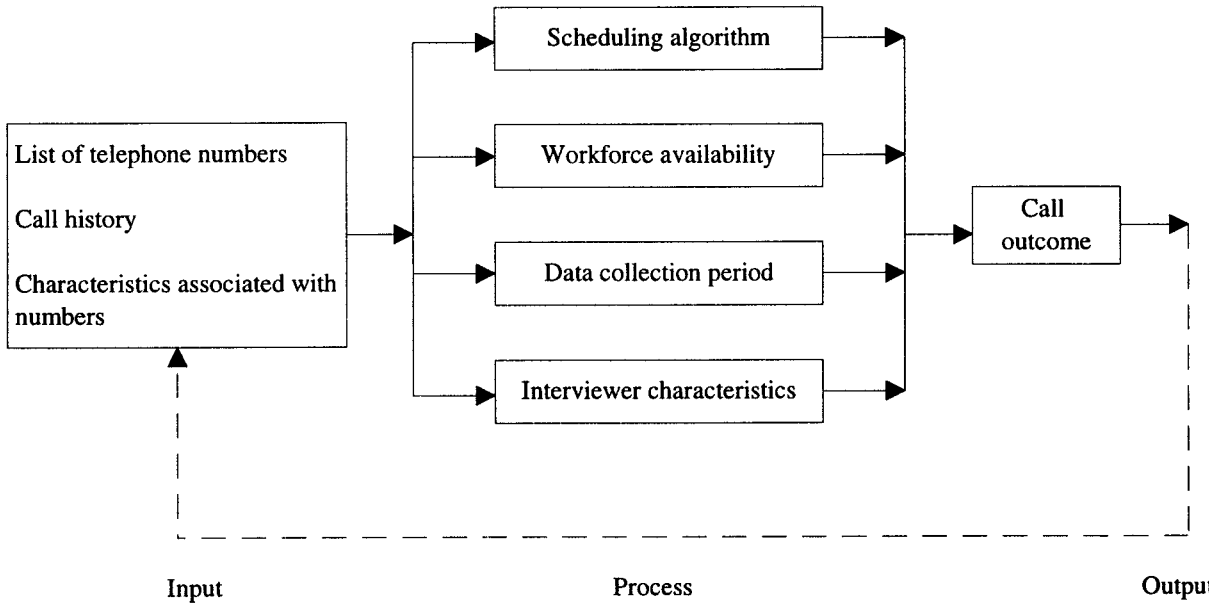
Greenberg and Stokes (1990) presented an optimal or priority system for call scheduling. They developed models of the probability of success based on the characteristics of previous calls and used a Markov decision process to try to minimize the expected number of calls. In another approach, Stokes and Greenberg (1990) used logistic regression to predict the probability of success of a call at a particular time and used this model to develop a ranking system for callbacks.

The next section describes the national survey used as the empirical basis for the analysis. Some of the analytic results are then presented and compared to results in the literature. Subsequently, logistic regression models are used to analyze the same data. The final section summarizes the results and suggests future research.

2. Data Source

The basis for the analysis of calling protocols in this study is a national, list-assisted RDD sample of telephone numbers conducted by Westat between January and May of 1995. The survey was the 1995 National Household Education Survey for the National Center for Education Statistics of the US Department of Education. The sample contained 59,703 residential telephone numbers. The size of this data file is large enough to permit a detailed analysis of some calling protocols. Analysis of all telephone numbers dialed instead of just the residential ones was considered, but current procedures used to eliminate a portion of the nonresidential telephone numbers limit the usefulness of this type of analysis.

Figure 1. Flow of Process of a Telephone Survey



The assignment of the telephone numbers to be dialed followed a specific calling protocol that was typical for a Westat study of this size and duration. It is important to realize that the numbers were not randomly assigned to the times available for calling and any analysis of the data must take this into consideration. This point is essential to understanding how the empirical results of this study may, or may not, be useful to planning and conducting other surveys.

The analysis is restricted to call attempts to establish the first contact with the household, *i.e.*, the attempts prior to first speaking with a person in the household. First contacts are obviously critical to the success of the survey. Most studies of calling protocols restrict attention to first contacts because once a first contact has been made, the approaches to soliciting participation in the survey depend on a number of other factors. These other factors include the sponsorship of the survey, the length of the interview, the type of data being requested, as well as a host of other factors that may be different from survey to survey. Furthermore, once a household has been contacted, additional contacts with the household are typically driven by appointments.

The first step of the analysis involved comparing the contact rates (the percentage of calls that were answered) of our first telephone call attempts to rates using the detailed categories of time of day and day of week reported by Weeks, Kulka, and Pierson (1987). The times of day are those of the respondent not the interviewer. The results were consistent with the overall patterns reported in that study, but the percent contacted were slightly different. Next, the calling times were collapsed into seven time periods which were defined so the contact rates were relatively homogeneous within the time period but different across the periods. The seven groups are: Monday to Friday from 9 am until 4 pm (M-F, 9-4), Monday to Friday from 4 pm until 7 pm (M-F, 4-7), Monday to Friday from 7 pm until 9 pm (M-F, 7-9), Saturday from 9 am until 4 pm (Sat, 9-4), Saturday from 4 pm until 9 pm (Sat, 4-9), Sunday from noon until 4 pm (Sun, N-4), and Sunday from 4 pm until 9 pm (Sun, 4-9).

As mentioned above, the telephone numbers were not randomly assigned to each available time period because this would be very inefficient. Table 1 shows the distribution of first and second calls made in each of the seven time periods. Since we are only considering calls needed to make first contact with the household, the number of first calls is 59,703 and the number of second calls is only 20,738. The distributions of times for first and second calls are dissimilar, with a much greater percentage of second calls made during the weekday mornings and early afternoons. This is a product of the calling protocol that attempted to spread the telephone attempts over daytime, evening, and weekend attempts.

Table 1. Percentage distribution of first and second calls*, by time period

Time period	First call	Second call
Weekday, 9-4	4.5	48.4
Weekday, 4-7	32.4	26.2
Weekday, 7-9	39.9	5.9
Saturday, 9-4	5.7	8.4
Saturday, 4-9	4.0	1.5
Sunday, N-4	0.7	6.8
Sunday, 4-9	12.8	2.8

* The number of first calls is 59,703, and the number of second calls where the first call was a noncontact is 20,738.

The vast majority of households (79%) were contacted in either the first or the second call attempt, with 65 percent contacted on the first call and 39 percent of the remaining numbers contacted on the second call. Because of the high percentage contact in these two calls and the sparseness of the data when third and subsequent call attempts are examined, all the analysis in this paper is for the data from the first two attempts.

3. Descriptive Analysis

Most of the literature on calling protocols concentrates on contacting households or completing interviews but ignores refusals. However, refusals are very important in a survey, impacting on both the response rates and the costs to complete the survey. For this reason, refusal outcomes are considered explicitly in most of the analyses that follow.

Figure 2 summarizes the results of first and second call outcomes to contact the household by the time of the calls. The high contact and completion rates for first calls made on weekday evenings is consistent with reports in the literature. The graphs also show the refusal rates are higher for this period, but both the high completion and refusal rates are largely a result of contacting more households in this period. When the completion and refusal rates for first and second attempts are computed conditioning on the number of contacted numbers, the rates are virtually identical. Another expected feature of the figure is that the rates are higher for first attempts than for second attempts. The rates for second attempts are nearly parallel to those of the first attempts, only lower.

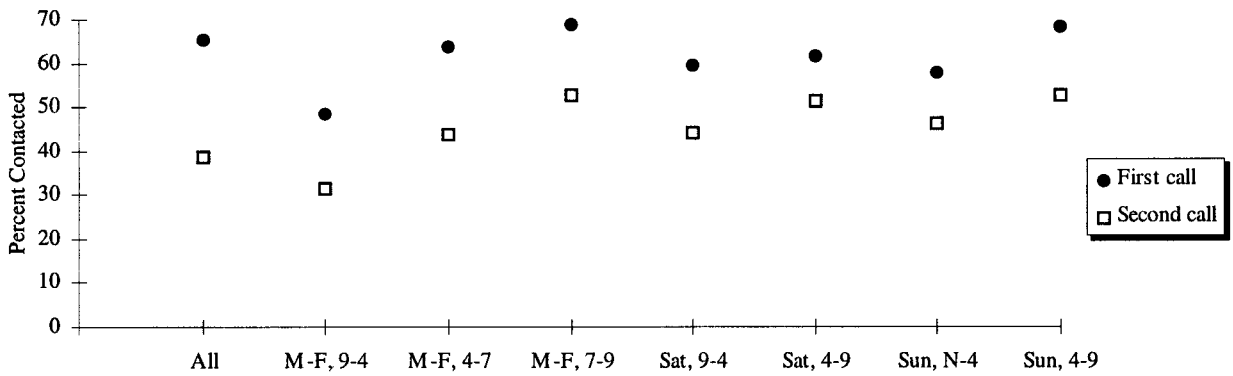
A very important issue in developing a calling protocol is the association between the time of the first call and the second. Previous studies suggested that the conditional probability of reaching a household on the second call was very dependent on the time of the first call attempt, with lower contact rates occurring when both first and second calls were made in the same time period.

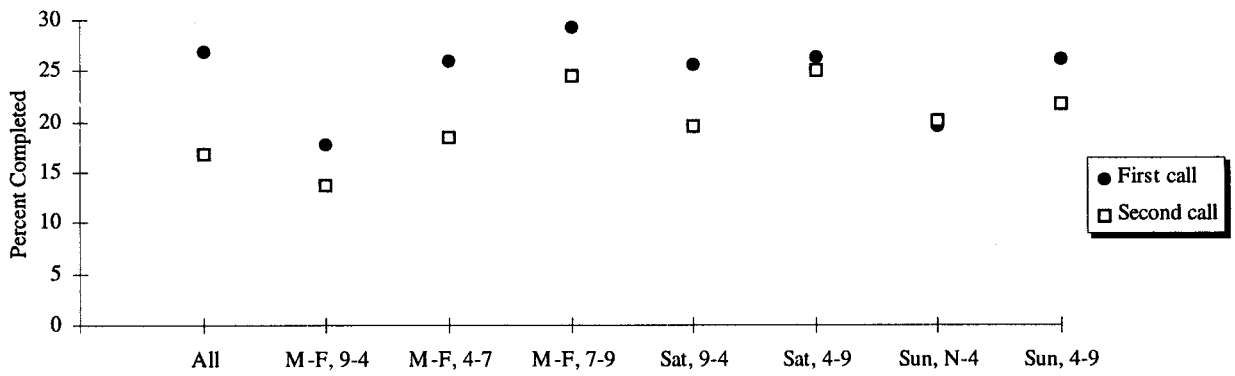
To examine this issue, Table 2 gives the distribution of the 20,738 second call contact rates by the time of the first call. If the hypothesis of lower contact rates for calls in the same time period is true, we would expect the diagonal elements in the table to be depressed relative to the overall percentages in the first column. This is not apparent from the table, although many of the cells are suppressed because of insufficient sample size.

We suspect that this study did not show a decrease in contact rates for the same time period because the results are based on a particular study and with its calling protocol. Other studies which demonstrated this pattern (Kulka and Weeks 1989, Stokes and Greenberg 1990) were also empirical studies with their own data collection requirements and calling protocols that might have been very different from that used in this study. For example, Stokes and Greenberg reported on a study that had to be completed in 14 days and required frequent callbacks, with several callbacks on the same day possible. In the study reported here, most of the first and second calls were made several days apart.

This lack of randomization in assigning the numbers to the time periods means that the results may not be very portable from one study to the next, unless the same types of conditions exist. In general, the calling protocols used in the study may have a great influence on the results. Because of this and the need to better understand the factors affecting the contact, completion, and refusal rates, the next section uses models to describe the results from this survey.

Figure 2. Percentage of First and Second Call Outcomes, by Time of Attempt





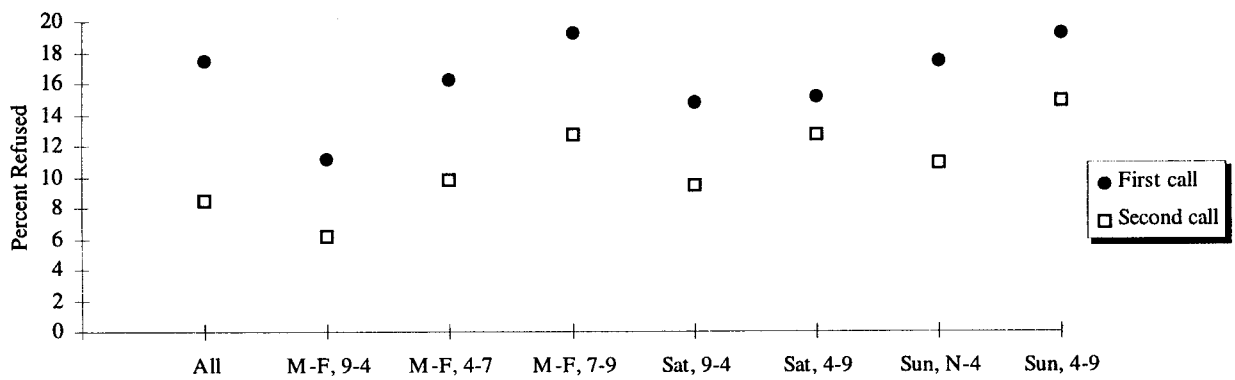


Table 2. Percent contacted on second call*, by time of first call

Time of second call	Percent contacted or second call	Time of first call						
		M-F, 9-4	M-F, 4-7	M-F, 7-9	Sat, 9-4	Sat, 4-9	Sun, N-4	Sun, 4-9
M-F, 9-4	31.3	30.0	28.8	33.1	29.2	41.1	35.8	34.9
M-F, 4-7	43.6	50.4	47.9	39.1	47.7	57.9	41.4	43.4
M-F, 7-9	52.8	-	51.4	56.2	50.0	-	-	50.8
Sat, 7-4	44.3	50.6	45.3	43.0	-	56.6	-	-
Sat, 4-9	51.2	-	50.0	48.0	-	-	-	-
Sun, N-4	46.3	-	-	-	45.4	49.4	-	-
Sun, 4-9	52.4	-	-	-	-	48.6	-	58.6

*The number of second call attempts was 20,738. If the number of calls in a cell is less than 30, the percent contacted is replaced by a "-".

4. Logistic Regression Analysis

Since a randomized design is not a practical alternative for a large scale national survey, logistic regression models are used to examine the relationship between the procedures used to assign the numbers and the outcomes of the calls. Three dependent variables are considered where the probability of success is defined as: contacting a household, completing the interview, and refusing an interview. The outcome of contacting a household includes completing and refusing an interview along with other contacts that required callbacks.

Logistic regression models of the form below were fitted to the data,

$$\log it(y_i) = \beta_0 + \beta_1 x_i \ ,$$

where y_i is the binary outcome variable (either contact, complete, or refusal), and x_i is a vector of predictor variables. A number of different predictor variables were considered in this analysis, including indicator variables for the time period for the call, variables related to this study's calling protocol, characteristics associated with the telephone number, and variables indicating the previous call history for a telephone number. We begin below with a simple model involving only the variables indicating when the call was made and then add other predictors.

Table 3 shows the estimated parameters of logistic regression models when the predictors are the 6 time periods (the M-F, 9-4 period is the reference cell) for first and second call attempts. The estimated parameters of the models are very similar for both first and second call attempts. The parameter estimates are also consistent with the results observed in Figure 2, with larger parameters for the weekday evenings. Furthermore, the magnitudes of the parameter estimates for predicting contact, complete, and refusal outcomes are very similar. This suggests that models that do well for one type of outcome may be adequate for the other types of outcomes.

Table 3. Estimated parameters of logistic regression models for first and second call attempts with time of call predictors

Predictors	Contacted		Completed		Refused	
	Call 1	Call 2	Call 1	Call 2	Call 1	Call 2
M-F, 4-7	.64	.53	.48	.35	.44	.50
M-F, 7-9	.86	.91	.65	.75	.64	.77
Sat, 9-4	.45	.57	.46	.43	.33	.51
Sat, 4-9	.54	.84	.50	.74	.35	.79
Sun, N-4	.39	.64	.12	.46	.52	.62
Sun, 4-9	.85	.88	.49	.57	.65	.98

The models in Table 3 use only the time of the call to predict the outcomes. We investigated several other predictors of the outcomes using variables related to the calling protocols. Only one of these variables had a significant effect, the time zone of the respondent, and this was only significant for first call attempts. The estimated parameters for this model in Table 4 show the probability of contacting and completing an interview in the Pacific and Eastern time zone are less than in the other time zones. Interactions between the time zone variables and the time of the calls were not statistically significant. The lack of interactions implies that the calling protocol does not have to assign numbers differently in the different time zones.

The next set of predictors that was introduced where the characteristics of the area that could be identified from the telephone number. Mohadjer (1988) describes how data from the decennial Census can be matched to the areas served by telephone exchanges. This matching is a simple and inexpensive by-product that can be obtained from commercial firms which produce list-assisted samples. The following 1990 Census data were considered as predictors in this analysis: median years of education (YRSED), log of median home value (HOM), log of median income (INC), and two indicator variables, one if the population in the area is 20 percent or more Black (%BLACK) and the other if the population is 20 percent or more Hispanic (%HISPANIC).

Table 5 shows the estimated parameters for a model with the time period, time zone, and demographic predictors that fits well for first call attempts. In this model, the time periods were collapsed so there are only three periods (weekdays before 4 pm, weekdays after 4 pm, and weekends). The demographic predictors are generally highly correlated with outcomes, with the magnitudes of the parameters consistent across the three outcomes except for INC where the contact and complete have different signs. The estimated parameters for areas with high concentration of Black and Hispanic populations are suggestive. They indicate households in these areas are harder to contact and have lower refusal rates. Of course, these outcomes are correlated but they do land some support to conjectures about schemes for scheduling telephone interviews in minority areas. When interactions between the time periods and the demographic and time zone predictors were examined, none of the interactions were statistically significant.

Table 4. Estimated parameters of logistic regression model for first call attempts with time of call and time zone predictors

Predictors	Contacted	Completed	Refused
M-F, 4-7	.60	.39	.41
M-F, 7-9	.78	.48	.59
Sat, 9-4	.39	.34	.29
Sat, 4-9	.46	.35	.30
Sun, N-4	.42	.18	.54
Sun, 4-9	.79	.36	.61
East	-.01*	-.10	.03*
Pacific	-.15	-.37	-.07*

*Indicates parameter is not significantly different from zero.

Table 5. Estimated parameters of logistic regression model for first call attempts with time of call, time zone, and demographic predictors

Predictors	Contacted	Completed	Refused
M-F, 4-9	.23	.11	.21
Sat, Sun	.09	.00*	.12
Pacific	-.16	-.18	-.10
INC	.15	-.40	.27
YRSED	-.19	-.04	-.09
%BLACK	-.05	-.19	-.13
%HISP	-.10	-.36	-.18

*Indicates the parameter is not significantly different from zero.

The last set of models was for second call attempts and predictor variables that indicated the time period of the first call were added to the other predictors already in the model. The three call history indicator variables were: first call attempted M-F, 9-7 (WKDY1), first call attempted M-F, 7-9 (EVE1), and first call attempted on Saturday or Sunday (WKND1). The estimated parameters for the fitted model are shown in Table 6. The only call history indicator variable that came into the model was whether or not the first call was made on the weekend (WKND1) and that even variable does not have a large impact on the predicted power of the model.

Interactions between the call history indicator variables and the time of the second call were tested and none of these was statistically significant. This differs from the results presented by Stokes and Greenberg (1990), where the call history interaction with the time of the call was a significant predictor of the outcome. As they noted in their report and we re-iterated above, these results may be dependent on the specific conditions of the calling protocol and not generalizable to other situations. For example, they found a significant main effect for the number of days between attempts that does not appear in our model because the lag was almost always greater than 2 days (the largest category in their study).

The findings of the models with call history do not imply that the time of the first attempt is irrelevant when scheduling the second attempt. Rather, they do suggest that within the framework of the calling protocol used in this study, the time of the first call did not have a large impact on the outcomes of the second calls. Different protocols, such as that used in the study reported by Stokes and Greenberg (1990) could result in different outcomes. Clearly, no one calling protocol can be developed that will be optimal for all telephone surveys. The calling protocols should take into account not only the optimal times for contacting and completing interviews, but also should incorporate factors such as the data collection period and the other factors mentioned in Figure 1.

Table 6. Estimated parameters of logistic regression model for second call attempts with time of call, time zone, demographic, and previous call history predictor

Predictors	Contacted	Completed	Refused
M-F, 7-9	.67	.57	.52
Sat, Sun	.41	.32	.38
WKND1	.15	-.05*	.15
INC	.09*	-.42	.43
YRSED	-.20	-.07	-.22
%BLACK	-.08	-.16	-.20
%HISP	-.11	-.39	-.14

*Indicates parameter is not significantly different from zero.

5. Concluding Remarks

The findings from this evaluation are consistent with much of the published literature on when the best times are for contacting households. However, there is an important relationship between the calling protocol used in the study and the outcomes that must be addressed if the results are to be applied to other surveys. If these results, or those from any other empirical studies of this kind, are to be applied to a future survey, the designers of the calling protocols should carefully consider which features of their survey might cause the outcomes to be different and modify their calling protocols to account for these features.

We are currently pursuing several areas of additional research that might shed some light on issues confronted in this study. We are planning to repeat much of the analysis reported here for a similar survey conducted in 1996 to determine if the estimated regression parameters are stable. In the analysis, we also plan to examine additional predictor variables that are specific to the telephone number (data for telephone numbers listed in the White Pages). Other areas of active research are the relationship between the final outcome being a refusal and the time of the first contact with the household and the extension of the analysis to additional calls.

The product of this research is not likely to result in the development of a calling protocol that is optimal for all CATI surveys. The survey conditions encountered are diverse and no one protocol will be adequate for all these conditions. However, this research can and should lead to a better understanding of the factors that must be considered and the methods that should be applied in a variety of circumstances. If this is successful, then calling protocols such as that suggested by Stokes and Greenberg (1990) could help to reduce data collection costs and improve response rates.

References

Greenberg, B.S., and Stokes, S.L. (1990) "Developing an Optimal Call Scheduling Strategy for a Telephone Survey," *Journal of Official Statistics*, 6 (4), 421-435.

Groves, Robert M. (1989) *Survey Errors and Costs.* Wiley and Sons, New York.

Kulka, R. A. and Weeks, M. F. (1989) "Toward the Development of Optimal Calling Protocols for Telephone Surveys: A Conditional Probabilities Approach," *Journal of Official Statistics*, 4 (4), 319-332.

Mohadjer, L. (1988) "Stratification of Prefix Areas for Sampling Rare Populations" Chapter 10 in *Telephone Survey Methodology*, eds. Groves, Biemer, Lyberg, Massey, Nicholls, and Waksberg, 161-174. Wiley and Sons, New York.

Stokes, S.L., and Greenberg, B.S. (1990) "A Priority System to Improve Callback Success in Telephone Surveys," *Proceedings of the Section on Survey Research Methods of the American Statistical Association,* 742-747.

Weeks, M. F. (1988) "Call Scheduling with CATI: Current Capabilities and Methods" Chapter 25 in *Telephone Survey Methodology*, eds. Groves, Biemer, Lyberg, Massey, Nicholls, and Waksberg, 403-420. Wiley and Sons, New York.

Weeks, M.F., Kulka, R.A., and Pierson S.A. (1987) "Optimal Call Scheduling for a Telephone Survey," *Public Opinion Quarterly* 51: 540-549.