NONRESPONSE IN THE NATIONAL HOUSEHOLD TRANSPORTATON SURVEY (NHTS)

DAVID CANTOR, GARY SHAPIRO, LI WAN CHEN,
G. HUSSAIN CHOUDHRY, MARK FREEDMAN

Westat, 1650 Research Blvd., Rockville, Maryland 20850 – Fax. 301/294-2034

Corresponding author: Gary Shapiro, 301/517-8028, garyshapiro@westat.com
David Cantor, 301/294-2080, davidcantor@westat.com
Li wan Chen, 301/610-5586, liwanchen@westat.com
G. Hussain Choudhry, 240/453-2783, ghussainchoudhry@westat.com
Mark Freedman, 301/294-2857, markfreedman@westat.com

ABSTRACT

This paper examines correlates of nonresponse for the 2001-2002 National Household Travel Survey (NHTS) that Westat conducted for the U.S. Department of Transportation. The paper compares nonrespondents at any stage of the survey to respondents at all stages, and also makes comparisons separately for the screener interviewing stage and for the extended interviewing stage. We used demographic, socio-economic and geographic characteristics for these comparisons. We compared the screener response rates using exchange level characteristics, and we compared the response rates at the extended survey level using the person and household characteristics. The results in the paper have implications for the methods used to conduct future national travel surveys using random digit dialing methodology. Most characteristics that are examined show wide ranges in response rates, implying that the potential exists for high bias due to nonresponse. However, weighting adjustments were made by homogeneous nonresponse cells defined by some of these characteristics, so that much of the potential bias may have been eliminated by the weighting methodology. However, such adjustments have adverse impact on the variances. The analysis consists mainly of direct comparisons of response rates across dimensions for a number of characteristics. Among the characteristics for which nonresponse rates greatly differ by category are home value, race/ethnicity, level of education, number of adults in the household, and dwelling type and size. For example, the response rate is 50.5 percent in telephone exchanges with low percentages of Hispanics and it is 37.6 percent in exchanges with high percentages of Hispanics.

This paper also assesses whether there are better ways to schedule the screening calls for the NHTS that decrease the number of calls that result in a household contact, i.e. increase the efficiency of the calling, e.g., best time to call. Overall, the scheduling algorithm is found to be efficient. One possibility for spreading out the calls across calling periods is to adopt a procedure that sets priorities for filling time slices within the first *four* calls instead of during the first *seven* calls (Cunningham, et al., 2003).

1. INTRODUCTION

Nonresponse is a serious concern in all household surveys. It was particularly difficult to obtain high response rates in the National Household Transportation Survey for several reasons: Initial household contacts were through random digit dialing, individual extended interviews were required from all adults in the household in a specified 6-day period, and there was a need for household respondents to keep a travel diary prior to the extended interviews.

Sections 2 and 3 relate to different issues regarding nonresponse in this survey. Section 2 compares respondents and nonrespondents (both for the screener interview and the extended interview) on a number of different dimensions, using several sources of data. Section 3 assesses whether there are better ways to schedule screening calls for the survey, by explicitly accounting for the calling period at previous time periods or whether there are other characteristics associated with the sample telephone number that can be used to prioritize the times that calls are made.

2. CHARACTERISTICS OF NONRESPONDENTS

2.1 Type of Analysis

This section examines nonresponse using data from several different sources. A telephone number can be classified as a respondent, a nonrespondent, ineligible (business or nonworking), or one with unknown eligibility (ring, no answer and answering machines) at the screener interview. The ineligible telephone numbers are not of interest for the travel survey. If the telephone number is a respondent at the screener interview, then the extended interview was attempted for eligible members of the household. The household and the persons identified in the screener interview can be classified as respondents or nonrespondents in the extended interview.

We compare respondents and nonrespondents for three categories:

1. Screener interview respondents vs. screener nonrespondents (section 2.4)
2. Extended interview respondents vs. extended interview nonrespondents (section 2.3)
3. Extended interview respondents vs. all nonrespondents (to either the screener interview or the extended interview) (section 2.2)

Note that only the concept of “usable households” is examined in this paper. A “usable household” is a household for which at least 50 percent of the adults in the household completed the extended interview. A household is a respondent for the extended interview if it is a usable household. A person is a respondent for the extended interview if the person belongs to a usable household and also completed the extended interview.

There are three sources of data used to examine nonresponse. Telephone exchange level data is produced by a Westat vendor that provided the frame for the random digit dial sampling. The vendor used Decennial Census data at the zip code level to estimate the proportion of households using that telephone exchange that have various demographic characteristics.

The second source of data is telephone level data that was provided by a second vendor. This is for characteristics of the household assigned to the telephone number rather than characteristics of the entire telephone exchange. This is potentially more useful data for examining nonresponse bias, but suffers from several drawbacks. First, the data is generally for current (2003) persons and households assigned to a telephone number, whereas the survey was conducted in 2001-2002 when there could have been a completely different household at that telephone number. Second, some or all data are not available for some households. Third, the information may not always be correct – we are not aware of any measures of the data quality. The final source of data, available for all households that completed the screener

interview (regardless of whether it was a respondent or usable household in the extended interview), includes all demographic data collected in the screener interview.

The paper provides results and discussion from the comparisons of response rates across a wide range of characteristics, most with indications of statistical significance level. Statistical significance was determined in different ways, depending on the nature of the comparison. In most cases, a t-test was done to determine the significance of the response rate for a specific cell compared to the overall mean response rate. Chen et al (2004) provides detailed statistical analysis for more characteristics, with specific standard errors and results of logistic regression significance tests.

2.2 Comparisons for Nonresponse at Either Screener or Extended Interviews

This section compares final respondents (to the extended interview) and nonrespondents at any level (either the screener or extended interview). This is the fundamental issue of most interest, as it is the overall indicator of potential nonresponse bias. Conceivably, screener nonrespondents and extended nonrespondents might tend to have opposite characteristics, such that there were big differences when examining only screener nonresponse and when examining only extended interview nonresponse, but small differences when examining nonresponse at either level.

2.2.1 *Characteristics with Substantial Differences*

Most of the examined characteristics show large differences between respondents and nonrespondents at any level. There were twelve characteristics that clearly show substantial differences: home value, race/ethnicity, region, time zone, whether address is known, number of adults per household, dwelling type, number of dwellings in the building, religion, language, whether a household contains married persons, and whether a household contains working women.

As an example, consider the following four tables pertaining to race, ethnic origin and language. All tables show that Hispanics and Spanish language households have higher nonresponse rates. Table 2.2-1 shows that the average proportion of respondent households who are Hispanic, based on telephone exchange data, is 9.1 percent, while the average proportion of nonrespondent households who are Hispanic is 12.2 percent. Table 2.2-2 shows that the response rate for telephone exchanges with low percent Hispanic is 50.5 percent, while the response rate for telephone exchanges with high percent Hispanic is only 37.6 percent. Table 2.2-3 shows that households identified as Hispanic ethnic origin have a low response rate of 34.1 percent. Table 2.2-4 shows that Spanish language households have a much lower response rate (31.3 percent) than English language households (50.9 percent).

TABLE 2.2-1 Comparison of Extended Interview Respondents and Nonrespondents by Race/Ethnicity (Exchange Level Data)

Characteristics	Significance	Respondents	Nonrespondents at Any Level
Percent Asians		23.2	23.2
Percent Black	**	10.0	13.1
Percent Hispanic	**	9.1	12.2
Percent White	**	76.8	69.6

*$p < 0.05$.

**$p < 0.01$.

TABLE 2.2-2 Extended Interview Household Response Rates by Race Distribution (Exchange Level Data)

Race	Significance	Low	Medium	High
Percent Asian		44.2	46.7	43.5
Percent Black	**	52.6	44.7	38.8
Percent Hispanic	**	50.5	46.4	37.6
Percent White	**	35.1	46.0	54.2

*p < 0.05.

**p < 0.01.

TABLE 2.2-3 Household Response Rate by Ethnic Origin (Telephone Level Data From Vendor)

Ethnic Origin	Response Rate
Asian (non-Oriental)	38.9%
Southern European	40.4
French	55.2**
German	56.2**
Hispanic	34.1**
Italian	45.1
Jewish	49.8**
Miscellaneous	54.7*
European	53.4**
Asian	31.2**
Polynesian	29.8*
Arabic/Middle Eastern	33.3**
Scottish/Irish	52.7**
Unmatched, uncodeable	43.0
African American	35.3**
Overall	48.0

*p < 0.05.

**p < 0.01.

TABLE 2.2-4 Household Response Rate by Language (Telephone Level Data From Vendor)

Response rate	English	Spanish	Other	Unknown	Overall
Response rate	50.9**	31.3**	37.0*	43.0**	48.0

*p < 0.05.
**p < 0.01

As another example of a characteristic with large differences, consider number of adults in a household. Table 2.2-5 shows that households with 1 adult have a much lower response rate (43.7 percent) than households with two or more adults.

TABLE 2.2-5 Household Response Rate by Number of Adults in Household (Telephone Level Data From Vendor)

Number of Adults	Response Rate
1 = 1 Adult	43.7%**
2 = 2 Adults	53.1**
3 = 3 Adults	53.2**
4 = 4 Adults	55.5**
5 = 5 Adults	51.6*
6 = Greater than 5 adults	49.0
Default is blank(s)	38.3**
Overall	48.0

*p < 0.05.

**p < 0.01.

2.2.2 Characteristics for Which Data are Inconsistent or with Small Differences

There were three characteristics for which some data show substantial differences between respondents and nonrespondents and some data sources show little differences. The three characteristics are educational attainment, owner- vs. renter-occupied, and age. For example, the average owner-occupied rate, based on telephone exchange data, is 67.4 percent for respondent households and 64.1 percent for nonrespondent households, a relatively minor difference. However, Table 2.2-6 shows a response rate of 37.1 percent in exchanges with low percentages for owner occupied and 49.8 percent in exchanges with high percentages for owner occupied, a rather large difference.

TABLE 2.2-6 Household Response Rate by Owner/Renter Occupied (Exchange Level Data)

Household	Low	Medium	High
Percent owner occupied **	37.1	47.1	49.8
Percent renters **	49.8	47.1	37.0

*p < 0.05.

**p < 0.01.

Income is the one characteristic for which all the data indicates that response rates do not vary substantially.

2.3 Comparisons for Nonresponse at the Extended Level

The comparison of respondents at the extended level to nonrespondents at the extended level, among those who completed the screener interview, is particularly interesting because it allows direct comparisons between respondents and nonrespondents for characteristics collected in the screener interview. This is the best source of data that is used in this paper.

2.3.1 Characteristic with Substantial Differences

Most of the examined characteristics show large differences between respondents and nonrespondents. The nine characteristics for which there are large differences are owner occupied, race/ethnicity, with a job or with more than one job, household size, number of drivers, number of vehicles, age, number of workers/work status, and number of drivers/driving status.

For example, Table 2.3-1 shows that 6.0 percent of respondents to the extended interview belong to the 18-24 age groups while 11.7 percent of nonrespondents belong to this age group.

TABLE 2.3-1 Age Comparison of Extended Interview Respondent and Nonrespondent Persons (Screener Interview Level Data)

Characteristic	Extended Interview Respondent persons (*n*=60,282)	Extended Interview Nonrespondent Persons (*n*=37,032)
Percent age 0–17	24.2	28.3
Percent age 18–24	6.0	11.7
Percent age 25–34	11.3	15.1
Percent age 35–44	15.3	16.5
Percent age 45–54	16.0	13.4
Percent age 55–64	11.4	7.2
Percent age > 64	15.7	7.9

Note: Exact significance testing was not done for this table.

As a second example, Table 2.3-2 shows that households with one or two drivers have much higher response rates than households with no driver or with three or more drivers. Also, drivers have a much higher response rate (64.1 percent) than non-drivers (53.4 percent), based on screener interview data.

TABLE 2.3-2 Extended interview Household Response Rates by Number of Drivers (Screener Interview Data)

Household Characteristics	Response Rate
Households with no driver	64.7**
Households with one driver	70.1**
Households with two drivers	74.6**
Households with 3 or more drivers	60.1**

*p < 0.05.

**p < 0.01.

As a further example, Table 2.3-3, using the best data source (screener interviews), shows that households with 2 vehicles have a relatively high response rate of 73.4 percent, while households with no vehicles have a relatively low response rate of 63.9 percent. Table 2.3-4, using data from a vendor source, shows smaller though statistically significant differences by number of vehicles. On the other hand, the average number of cars per household is about the same for both screener respondents and nonrespondents.

TABLE 2.3-3 Extended Interview Household Response Rates by Numbers of Vehicles (Screener Interview Data)

Household Characteristics	Response Rate 1
Households with no vehicles	63.9*
Households with 1 vehicle	70.1
Households with 2 vehicles	73.4**
Households with 3 vehicles	69.9
Households with 4 or more vehicles	67.5

*p < 0.05.
**p < 0.01.

TABLE 2.3-4 Household Response Rates by Number of Known Owned Cars (Telephone Level Data from a Vendor)

Response rate	One	Two	Three or more	Zero or unknown	Overall
	70.9%**	69.8%**	69.0%	63.5%**	66.5%

*p < 0.05.
**p < 0.01.

2.3.2 Characteristics for Which Data Are Inconsistent or with Small Differences

There was no indication of large differences between respondents and nonrespondents for three characteristics (number of bicycles, number of cellular phones, and gender).

2.4 Comparisons for Screener Nonresponse

This section compares respondents and nonrespondents to the screener interview, with no regard as to whether screener respondents also responded to the extended interview.

2.4.1 *Characteristics with Substantial Differences*

Here again, most of the examined characteristics show large differences between respondents and nonrespondents. There were twelve characteristics with large differences: median home value, race/ethnicity, language, region, time zone, whether address is known, number of adults, length of residence, number of dwellings in the building, religion, marital status in the household, and home value. For home value, the median is lower for respondents, and the response rate is low for households with home value over $400,000, but there is little variation in response rates among other value categories. Table 2.4-1 shows that the Midwest Region has a relatively high screener response rate of 69.7 percent.

TABLE 2.4-1 Household Response Rates by Census Region at the Screener Interview (Exchange Level Data)

Region	Northeast	Midwest	South	West
	60.7%**	69.7%**	63.0%**	62.8%**

*p < 0.05.
**p < 0.01.

2.4.2 Characteristics for which data are inconsistent or with small differences

There are three characteristics (age, income, and percent owner-occupied) for which the data are inconsistent, and one characteristic, education, for which there is no evidence of differences between respondents and nonrespondents.

3. PREDICTING FIRST CONTACT RATES

An increasingly important component of non-response for random digit dial (RDD) surveys is the ability to make contact with eligible households (Curtin et al., 2003). Barriers to making contact with the household have increased because of the need for lines for computers (Pierkarski, 2002) and call screening devices (Tuckel and O'Neil, 2001; Murray et al, 2003).

The analyses described below assess whether there are better ways to schedule the screening calls for the NHTS that either increase the number of households that are contacted or increase the efficiency of the calling. The primary question of interest is whether there are characteristics of the sampled telephone numbers that suggest different scheduling strategies. To address this question, two analyses were conducted:

1. Should the scheduling algorithm explicitly account for the time periods during which the previous calls were attempted?

2. Are there area or other characteristics associated with the sampled telephone number that can assist in prioritizing the times calls are attempted?

In the next section, we describe the data-set used to address each of these questions.

3.1 Data-Set

The analyses described below examine the timing of calls made on the 2001-2002 NHTS. The data-set was initially composed of all call attempts made to complete the screening interview for the national RDD sample. The state supplements were excluded from theses analyses. These data were subset to include those telephones that were eventually identified as residential. Deleted from the data-set were non-working numbers and those that were identified as a business numbers. These two types of numbers typically require only one call to determine residential status and do not generally pose a scheduling issue. Also deleted from the data-set were numbers where no-one ever answered the telephone. Given that over 20 calls were made to these numbers, it is highly unlikely that changes in the scheduling routine would have much effect on their outcomes. The numbers that remain in the data-set are where a human eventually picked up the telephone and talked to an interviewer.

The data-set analyzed has a total of 56,645 cases where a first attempt was made. These data were not weighted by any selection probabilities because the RDD sample is an equal probability sample. Therefore, the analysis estimates significance using formulas assuming simple random sampling.

3.2 Calling Patterns by Timing of Call

As noted in the introduction, the proliferation of telephone numbers has led to increased efforts to identify and contact residential households. Nonetheless, it is still the case that the vast majority of the units are contacted within the first 2-3 call attempts. For the NHTS more than half of all households were contacted on the first call, with approximately 80 percent being contacted by the third attempt. By the seventh attempt, about 95 percent of households had been contacted

Stokes and Greenberg (1987) found that the rate of success depended on the previous calling period. Calling period refers to specific day-time the call was made. For example, if a call has already been placed during the day, there is an increased chance of a successful contact if the next call is made during the evening. The opposite was also the case. Among those calls that are made during the day, it is most likely to be a contact for those that had a previous call during the evening or weekend. The logic is that if it didn't work the first time at one period, then it may be because the target population is not at home during those hours. Trying a different time period may hit upon the period of occupancy.

The NHTS scheduling algorithm is based on this principle. Once a call has been made during the day, for example, the algorithm will specify other time periods that are eligible to be called next based on the time slices that have not been filled. However, it is not completely deterministic. The time slices used in the algorithm overlap and cover multiple time periods within each of the three major calling periods (i.e., day, evening and weekend). Depending on the workload and the number of interviewers making calls, it is possible for consecutive calls to be made during the same calling period. It is after multiple calls that the algorithm guarantees that calls are attempted across all relevant time periods the appropriate number of times.

Table 3-1 provides the success rate in contacting respondents at the second attempt by the calling period of the first attempt. These do not show a strong interaction between the first and second calling period, as would be predicted by Stokes and Greenberg (1987). The rate of success at the second contact is highest in the evening for two of the three first attempt calling periods (day = 48.9 percent and weekend = 48.8 percent). It is virtually the same as the weekend time when the previous call was made in the evening (with success rate equal to 39.3 percent). The least successful period is during the day across all previous calling periods (25.3, 27.2, and 31.4). The interaction between the previous and current calling period is significant primarily because of the relatively low contact rate when both calls were made during the day (25.3).

TABLE 3-1. Percent of First Contacts at Second Attempt by Current and Previous Calling Period (Number of Cases in Denominator in Parenthesis)

Calling Period For First Attempt	Calling Period For Second Attempt			Total
	Day	Evening	Weekend	
Day	25.3	48.9	40.1	32.3
	(4423)	(1123)	(1654)	(7200)
Evening	27.2	39.3	40.3	32.1
	(7856)	(1866)	(3044)	(12866)
Weekend	31.4	48.8	43.2	41.2
	(1276)	(443)	(4376)	(6095)
Total	27.0	43.3	41.7	34.3
	(13555)	(3532)	(9074)	(26161)

This analysis was extended to the third attempt with similar results (data not shown). The interaction between current and previous calling period is significant. However, it is not a strong relationship.

Overall, therefore, these data are more consistent with the analysis by Brick et al (1996), which did not find a strong relationship between first contact rates and previous calling periods. While the interaction between calling periods is statistically significant, it is due primarily to one or two changes in

rates, rather than a general pattern. Furthermore, this interaction seems to weaken at higher order attempts. The reduction in the strength of the interaction, based on the call history may be a function of the scheduling algorithm and the flow of the staffing on the NHTS.

While the NHTS algorithm does not explicitly give priorities based on call history, the logic strongly pushes the scheduling in this direction. For example, among the 3,303 third attempts where only a day call had previously been made, virtually none were again made during the day (12 out of 3303). A similar pattern is evident for attempts that had been preceded by 2 calls in the evening. However, a higher number of calls were made in the evening when two calls had previously been made in that period (426/1192=35%). Given this, it may be worth changing the algorithm to insure a greater spread of cases across the calling periods at this stage in the process.

3.3 Predicting Success of Contact with Household and Area Characteristics

Another possibility for improving the contact rate is to use the characteristics that are associated with the RDD sample frame. Developing priorities based on these characteristics would follow the same logic as using the timing of previous calls. If variables can be found that interact with the calling period, then it might be beneficial to use this information to create priorities when calling households that have not been contacted. Stokes and Greenberg (1987) found not only previous outcome to be related to optimal calling period, but also the number of previous calls, the outcome of the previous call and the time since the last previous call.

The characteristics that are known about the sampled telephone numbers for the NHTS are somewhat limited. One source of characteristics that are associated with the telephone number is the Census data. The supplier of the RDD numbers links the area code-exchange of each number to a zip code. These are, in turn, linked to Census data. For the present analysis, we included those listed in Table 3-2, which include demographic characteristics of the area (e.g., age, race, ethnicity, etc.), socio-economic characteristics of the area (e.g., housing values, education, income, etc.) and administrative data (matching to an address, whether a pre-notification letter sent prior to contact was returned as undeliverable).

In order to test for the significance of these variables, a series of logistic regressions were estimated that tested whether there was a significant interaction between the calling period and the characteristic of interest for the first four attempts to contact the household. A significant interaction indicates that it may be worth considering prioritizing the cases according to that characteristic.

3.3.1 Predicting First Contact Rates at the First Attempt

Table 3-2 provides the results of predicting contact rates at the first attempt. Four variables significantly interacted with the calling period at the .05 level or less --- the percent of persons age 65 and older ($p<.0136$), the percent Hispanic ($p<.0039$), the percent white ($p<.0034$) and the percent that own a home ($p<.0352$).

TABLE 3-2. Tests of Interactions of Characteristics and Calling Period Predicting First Contact Rates at the First Attempt

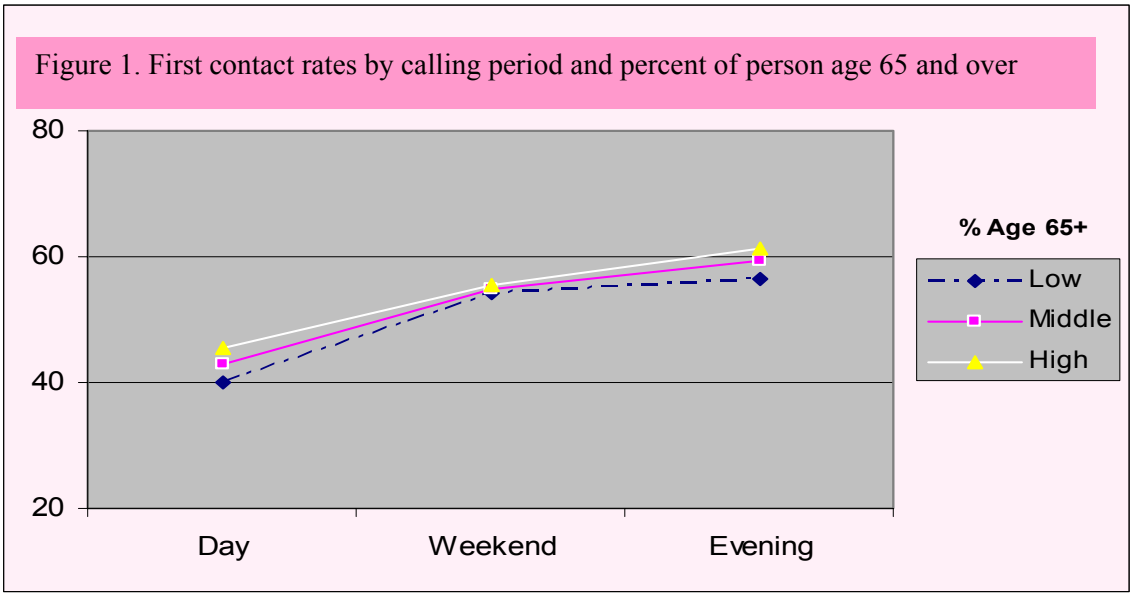
Characteristic	Degrees of Freedom	Chi-square	P-value
% Age 65 and over	2	8.6	.0136
% Black	2	2.57	.2769
% Hispanic	2	11.08	.0039
% White	2	11.36	.0034
% Own home	2	6.69	.0352
Log median income	2	2.76	.2516
Log median home value	2	1.10	.5766
% With income 75,000 or more	2	1.69	.4293
% Graduated college	2	3.31	.1911
Telephone number matched to address	2	4.32	.1152
Pre-notification letter returned	2	2.77	.2509

+ Logistic regressions included calling period (day, evening, weekend), characteristic shown above and interaction between the calling period and the characteristic.

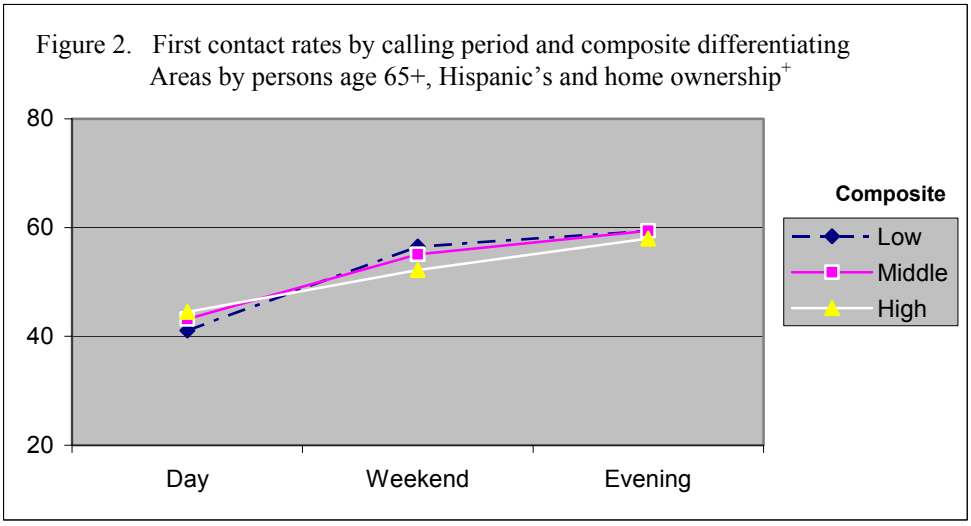
While these effects are highly significant, none of the estimated parameters are extremely large (data not shown). The significance may be partly a function of the large sample sizes available for the analysis. To get a concrete idea of the magnitude of the significant effects, the first contact rates were plotted for different levels of each characteristic. An example of this is shown in Figure 1. This plots the contact rates on the vertical axis, with a line representing a different level of the characteristic of interest. Lines that are not parallel indicate a significant interaction. Figure 1 plots the percent of the population that is more than 65 years old against the contact rate. One line is for the areas in the upper 25 percent of the distribution for the proportion of persons age 65 and over (HIGH). Another line is for the areas in the middle of the distribution of persons 65 and over (25^{th} to 75^{th} percentile; MIDDLE) and the third line is for the lower quartile (LOW). As can be seen, the lines between the Day and Weekend are not parallel. The line representing areas with high elderly populations is not as steep as the line with low proportions of elderly. A similar pattern occurs when looking at the lines between weekend and evening.

To increase the predictive power of the model, all of these variables were combined into a single logistic regression which included all of the main effects of the calling period and each variable, as well as the 2-way interactions of each characteristic and the calling period. As one might expect, not all of the 2-way interactions were significant in this combined model. The characteristics are not independent of one another and this reduces their significance in the combined model. The best fitting model is the one that includes the percent age 65+, the percent Hispanic and the percent of homeowners.[1]

[1] The significance of the 2-way interaction between homeownership and the calling period is not statistically significant. However, when this interaction is removed from the model, the overall goodness of fit is increased by a statistically significant amount.



To assess the strength of the combined relationships in this equation, a graph was generated that displayed the first contact rates by the combined levels of these three independent variables (Figure 2). The "HIGH" line represents areas that we would expect to have a smaller difference between day and weekend calls. It was created by choosing areas that have two of the following three characteristics: 1) upper 25th percentile for persons age 65 and over, 2) upper 25th percentile for percent Hispanic and 3) lower 25th percentile of percent of homeowners. The "LOW" line represents areas that we would expect to have a large difference between day and weekend calls. It was created by choosing areas that have two of the following three characteristics: 1) lower 25th percentile for persons age 65 and over, 2) lower 25th percentile for percent Hispanic and 3) upper 25th percentile for homeowners. The MIDDLE line represents all the other areas. As with the bi-variate results, the effects are not large. The slopes across the different periods are different, with the HIGH line crossing the other two. But the general ranking of the contact rates remains relatively unchanged. The differences across these areas, therefore, seem to be one of degree, rather than highly significant shifts.



+ High = High elderly, high Hispanic, low homeowners
Low = Low elderly, low Hispanic, high homeowners
Middle = All others

3.3.2 Predicting Contact Rates at the Second, Third and Fourth Attempts

Contact rates at the second, third and fourth attempts were predicted using the same set of independent variables used for predicting contact at the first attempt. The models included the calling periods for previous attempts, the particular characteristics of interest, and the interaction between the two most recent calling periods. The results found that for the second and fourth attempts, no variables were statistically significant. Several variables were significant for the third attempt. These included the percent age 65+, percent Hispanic and median home value. The directions of the effects for the third attempt are similar to those for the first attempt. For the third attempt, areas with a high percentage of persons 65+ are relatively more likely to be contacted in the evenings. Areas with a high Hispanic population were more likely to be contacted during the day. Persons living in areas with high median home values were more likely to be contacted during the day relative to the evening. The magnitudes of these effects are not large, however (similar to those discussed for the first attempt).

4. SUMMARY AND DISCUSSION

This report has examined response rates in the 2001-2002 NHTS across a number of different characteristics and from three different sources, and has also examined the methods used to schedule the initial screening calls.

Comparisons of response rates have been made:

(1) Between screener interview respondents and nonrespondents;
(2) Between extended interview respondents and nonrespondents; and
(3) Between final respondents (to the extended interview) and nonrespondents at either the screener or extended interview.

For each of the three comparisons, most of the examined characteristics show large differences between respondents and nonrespondents. The overall conclusion from these results is that there is a potential for substantial bias across a wide array of characteristics of importance to the NHTS.

It is difficult to draw many general conclusions from the results reported in this paper. However, extended interview nonresponse rates tend to be higher for low socio-economic groups and in lower socio-economic telephone exchange. This is indicated by high nonresponse rates associated with low median years of education, low percent college graduates, low percent owner occupied, low median income, high percent African American, high percent Hispanic, and households without any workers. Conversely, the screener interview nonresponse rates tend to be *lower* in low socio-economic telephone exchanges, although the data is somewhat mixed. Screener nonresponse rates are relatively low for telephone exchanges with low median income, low median home value, low percent of college graduates, and low percentages of households with income between $50,000 and $75,000. On the other hand, screener nonresponse rates are relatively high for telephone exchanges with higher percentages of renter occupied, high percentage of African American, and high percentage of Hispanic.

This report only looked at a limited set of characteristics, few directly related to the crucial travel data of NHTS. Thus, no direct conclusions are possible with respect to nonresponse bias on trip and travel estimates. Also, the report looked only at characteristics of the sample. We do not account for bias-reduction resulting from the nonresponse adjustment factors and raking applied in the weighting process. Thus, this report can only assess the *potential* for bias due to nonresponse, not the actual nonresponse bias for any survey estimates. During weighting, we applied the nonresponse adjustment for both screener interview nonresponse and extended interview nonresponse within homogeneous nonresponse adjustment cells to minimize the biases due to nonresponse. The homogeneous nonresponse adjustment cells were

determined with the Chi-squared Automatic Interaction Detector (CHAID) software that is based on a statistical technique for segmentation, or tree growing, developed by Kass (1980).

Although the nonresponse bias can be reduced by applying nonresponse adjustment within homogeneous nonresponse adjustment cells, there are variance implications due to differential nonresponse adjustment factors. The differential nonresponse adjustment introduces large variations in the sampling weights, which results in increased variances of the survey estimates. Those cells with extremely large nonresponse adjustment factors are generally collapsed with other cells to avoid large increases in the variances of the survey estimates. This would result in less bias reduction, but there is a trade-off between bias and variance, and the threshold value for the nonresponse adjustment factors is determined to minimize the mean square errors of the estimates.

The goal of the analysis in Section 3 was to assess whether there were alternative methods that might be used to increase the efficiency in resolving the residency status of sampled telephone numbers. By increasing efficiency, it might be possible to reduce interviewer hours. Theoretically, it could also reduce the number of telephone numbers with unknown eligibility status. Both scenarios would have positive effects on the response rate. Increasing efficiency allows for resources to be freed up to follow-up nonrespondents that have already been identified. Reducing the numbers with an unknown eligibility status would directly reduce the denominator of the response rate.

The results of this analysis suggest that the scheduling algorithm used for the NHTS was efficient. The vast majority of the residential telephone numbers were contacted within the first three call attempts (80%). All but five percent were contacted within the first 7 call attempts. The analysis does suggest that this process might be made more efficient by reducing the number of call attempts that are initially tried during the same calling period.

One possibility for spreading out the calls across calling periods is to adopt a procedure that was recently reported by Cunningham, et al (2003). This paper describes an experiment that sets priorities for filling time slices within the first four calls. This contrasts the algorithm used for the NHTS which filled time slices during the first 7 calls. Cunningham, et al (2003) found that the four call method resulted in a more even distribution across day, evening and weekend periods over a shorter period of time. That is, the four-call strategy had fewer call histories that had only tried one particular calling period after the first four calls were made. This strategy leads to a more efficient resolution of residential status during the first 7 attempts. The study did not find big differences in the response rates across the procedures, either with respect to the number of refusals or the proportion eventually classified as having an unknown residential status.

An alternative procedure to that used on the NHTS would be to set more explicit priorities on a particular call, based on the calling periods that have already been tried. This process is more deterministic than that used for the NHTS and is closer to what is described by Stokes and Greenberg (1987).

The above analysis found a few variables that were correlated with the calling period for the first and third attempts. No variables were statistically significant for the second and fourth attempts. The percent age 65 and over, as well as the percent Hispanic were significant for both the first and third call attempts. Areas with high median home values were positively related to contacting persons during the day.

Taken at face value, this suggests that calling efficiency might be improved if the scheduling algorithm accounted for these variables. However, given that the effects did not consistently predict contact rates across all attempts and that the significant effects were not large, we are skeptical that accounting for these variables would significantly improve upon the scheduling algorithm used for the NHTS. Given the many other priorities associated with scheduling calls and the relative payoff in slightly

increased efficiency, we would not recommend implementing a more deterministic scheduling logic based on these variables.

REFERENCES

Brick, M., Allen, B., Cunningham, P. and D. Maklan (1996) "Outcomes of a calling protocol in a telephone survey." Pp. 142 – 149 in Proceedings of the Section on Survey Research Methods of the American Statistical Association, Alexandria, VA.

Chen, L, Choudhry, G. H., Shapiro, G., and Freedman, M. (2004) "2001 National Household Travel Survey-Nonresponse Analysis" Report prepared for the Department of Transportation by Westat.

Cunningham, P., Martin, D. and J. M. Brick (2003) "An experiment in call scheduling." Paper presented at the 2003 Annual Meeting of the American Association for Public Opinion Research, Nashville, TN.

Curtin, R., Presser, S. and E. Singer (2003) "Recent response rate changes on the Michigan Survey of Consumer Attitudes" Paper presented at the 2003 Annual Meeting of the American Association for Public Opinion Research, Nashville, TN.

Murray, et al., (2003) "Impact of Changes in the Telephone Environment on RDD Telephone Surveys" Paper presented at the 2003 Annual Meeting Of the American Association for Public Opinion Research, Nashville, TN.

Piekarski, L. (2002) "Challenges to Telephone Sampling" Paper presented at the 2002 Annual Meeting of the American Association for Public Opinion Research, May 15-18, St. Pete Beach, FLA.

Stokes, S.L., and B.S. Greenberg (1990) "A priority system to improve callback success in telephone surveys" pp. 742-747, Proceedings of the Section on Survey Research Methods of the American Statistical Association, Alexandria, VA.

Tuckel, P. and H.W. O'Neill (2001) "The vanishing respondent in telephone surveys." Paper presented at the Annual Meetings of the American Association for Public Opinion Research, May 17-20, Montreal, Quebec.