House Effects in a Household Transportation Telephone Survey*

J. Neil Russell and Jonaki Bose
U.S. Department of Transportation, Bureau of Transportation Statistics, Washington, DC 20590

Abstract

Interviewer effects have been studied extensively in the literature while fewer studies have focused on "house effects", or the effects of multiple survey firms collecting data on the quality of survey data. The limited number of house effect studies may stem from fewer situations of multiple survey firms collecting data during the same time period and from the same population of inference. The Bureau of Transportation Statistics (BTS) sponsored the Omnibus Household Survey (OHS), which provides various transportation measures. Two commercial survey firms contracted by BTS collected OHS data, and this provided a unique opportunity to study this phenomenon.

For the OHS, each survey firm collected data from a random national sample during the same time period, used the same computer assisted telephone interview (CATI) instrument, interviewer manual and training procedures, and procedures for data collection quality control. However, each firm had its own personnel policies, management, and unique physical facilities; each hired, trained, and supervised its own interviewers.

This paper provides an overview of the survey and its data collection procedures, and then compares national estimates of sociodemographic and transportation-related measures by each of the two survey firms to help determine whether differences exist that could suggest the presence of a survey-firm (house) effect. Using six months of data, bi-variate analysis showed no sociodemographic or transportation measure was significantly different across survey firms. This may suggest a lack of house effects in the data. Future research is proposed to better understand this phenomenon.

Keywords: House Effects; Bureau of Transportation Statistics; Omnibus Household Survey

Background

Bureau of Transportation Statistics (BTS) initiated the monthly Omnibus Household Survey (OHS) in August 2000. Difficulties in survey administration placed the survey on hiatus for three months in 2001 so BTS could recontract for OHS administration services. BTS set aside the OHS contract for "8a" contractors. Only small, minority owned contractors were allowed to bid on conducting the survey. One concern BTS had about small survey research firms was the possible lack of overall infrastructure and resources necessary to conduct the survey.

In May 2001, BTS awarded the OHS contract to a small survey research firm who teamed with two other contractors to help complete the work. One subcontractor would help the main contractor with interviewing, and another subcontractor would clean the data, produce the weights and data documentation. The main contractor did not have sufficient staff or facilities to accomplish all the interviewing in house. With two different data collection firms conducting interviewing, BTS was concerned about survey organization or house effects contributing to non-sampling errors in the OHS estimates.

Because the OHS used two survey firms to collect data at the same time using independently drawn national samples, this provided a rare opportunity to estimate any possible survey organization effects for a set of demographic and transportation related variables.

Overview of the Omnibus Household Survey

The purpose of the OHS was to monitor expectations of and satisfaction with the transportation system and to gather event, issue, and transportation mode-specific information. The unique features of the OHS were the customer satisfaction component and the ability to provide final data on a range of questions on a continuing basis.

Questionnaire content included a core set of demographic questions to determine respondents' age, gender, and geographic area. Also, each month's survey questionnaire contained three other types of questions:

- Core set of transportation questions - These questions, which remained the same from month to month, asked respondents about their use of different modes of transportation and their perceptions and experiences using these modes.
- Department of Transportation (DOT) strategic goal questions - On a rotating basis, questions addressing the DOT goals of safety, mobility, human and natural environment, and security were asked. For example, questions on the environment were asked three times a year.
- Other DOT agency questions - These questions addressed specific issues of immediate interest to other agencies in DOT. For example, the National Highway Traffic Safety Administration asked opinion and behavioral questions about headlight

* Paper presented at the 2004 AAPOR annual meetings, Phoenix, Arizona.

glare and tire pressure measurement. These kinds of questions might be included only once or might be asked for several consecutive months.

OHS data were collected every month from approximately 1,000 U.S. households using a random digit dial (RDD) telephone methodology. One adult from each household was randomly chosen using the next birthday method. Two different contractors collected the data over a 10-day period each month.

BTS provided the contractor with the survey questionnaire. Then, one of the contractors programmed the computer assisted telephone interviewing (CATI) instrument. The draft survey questionnaire was reviewed each month by a panel of experts selected by the survey contractor and drawn from the statistical and transportation communities. Cognitive interviewing was used to highlight potential questionnaire design problems, and the CATI instrument was pretested to ensure that it was operating as designed.

Once the data collection period was complete, the data were cleaned based on the skip patterns and range checks. Then, national level weights were developed for making estimates and data documentation was produced. BTS received the final data and documentation 7 days after the data collection was completed. BTS would review the data and report any errors. The contractor would fix major errors before BTS disseminated the data on its website.

Data Collection Procedures and Non-sampling Error

BTS developed uniform data collection procedures to minimize potential house effect bias. First, each interviewing house had a national sample to interview. Other procedures were controlled so that each house had the same:

- 10 days to collect the data,
- CATI software,
- CATI instrument,
- Interviewer training curriculum,
- Interviewer training manuals,
- Data quality control procedures, and
- Interviewer monitoring strategies.

However, there were likely differences between the interviewing houses. BTS assumed there were differences between the houses in:

- Personnel policies,
- Style of management,
- Interviewer hiring practices, and
- Interviewer supervision practices.

BTS accumulated anecdotal evidence of interviewer and facility differences between the interviewing houses. Each month, BTS staff would travel to each contractor's data collection site. This always occurred prior to data collection. BTS staff observed interviewer training and then participated in monitoring live interviews. From these observations, BTS discerned the following differences between each of the survey organizations:

- Interviewer retention rates,
- Interviewer sociodemographics,
- Interviewer experience and professionalism,
- Interviewer supervisor experience and professionalism,
- Interviewer rewards for achievement,
- Interviewing facilities and infrastructure, and
- Organizational culture.

Even though each survey organization operated under a common set of survey administration procedures and had national samples, BTS was concerned that survey estimates obtained may differ over and above differences due to pure sampling error. This can occur because of non-sampling errors, both variable and systematic, introduced by the differences in interviewing staff, management practices, and the total interviewing environment.

Total error in a given survey statistic includes a component due to the data collection organization. If there is a survey organization effect, a substantial contribution to total error of survey estimates may occur (Cohen and Potter 1990).

Response Rates

Large response rate differences between survey firms are another potential source of house effect bias. Hence, response rates for the OHS were a concern, not only for non-response bias, but also in terms of mitigating any house effect bias in the estimates.

The OHS contract required the contractor to maintain a minimum response rate. Based on this, BTS asked the contractor to send daily reports on response rates during the data collection period and a final report at the end of collection. The daily response rate reports were formatted so response rates could be compared between survey firms. This was advantageous for BTS for each survey firm worked to ensure that their response rates were competitive with the other firm. As a result, there was little difference, month to month, in response rates across survey firms.

Related Research

Survey research findings on survey organization (house) effects have not been consistent. For example, Goldfield et al (1977), Smith (1978, 1982), Cohen (1986), and Mariolis and Graber (2003) all found significant organization effects.

However, Cohen (1982) and Cohen and Potter (1990) did not find substantial survey organization effects.

The lack of univocal findings on house effects did not help BTS make a decision on how to award the OHS contract. That is, if house effects are universal, BTS would have considered contract award to an organization that could complete all interviews without subcontracting out any of the work. By doing this, BTS would reduce a source of non-sampling error in the OHS estimates. However, there was no clear evidence for universal house effects. Hence, BTS undertook a post hoc analysis to ascertain whether house effects in the OHS contributed to any bias in survey estimates or to total survey error.

Methodology

We used the final sample weights and SPSS to conduct initial bi-variate analysis (chi-square) and proportional t-tests to ascertain differences in demographic, transportation, and sample distribution variables by survey organization. We chose an alpha level of 0.05 for tests of significance, but then used a Bonferroni adjustment that set the significance level for the statistical tests at 0.002. This adjustment was used because in multiple comparison tests, some could be significant due to chance.

For any significant difference found in SPSS, we then conducted additional statistical tests in SUDAAN to make a final determination of significant differences. For tests in SUDAAN, the estimates were weighted and the complex sample design was specified.

To ensure that each data collection organization actually had national samples, we compared the percentage of completed interviews by survey firm for Metropolitan Statistical Area (MSA) status and Census division.

The sociodemographic variables tested for house effects were:

- Disability status,
- Age,
- Gender,
- Ethnicity,
- Race,
- Education level,
- Income level, and
- Number of adults in the household.

The transportation mode variables in the OHS indicated whether a respondent used a particular mode of transportation in the past month. Transportation mode variables tested were:

- Personal vehicle,
- Car or vanpool,
- Public transit,
- Intercity bus,
- Intercity train,
- Commercial airline,
- Private aircraft,
- Motorcycle,
- Bicycle,
- Walking,
- Recreational boat,
- Watercraft,
- Commercial boat,
- Cruise ship, and
- Commuter status

Six months of OHS data were analyzed. We randomly chose one month per quarter over the life cycle of the survey. The following months of data were used:

- July 2001,
- November 2001,
- February 2002,
- June 2002,
- October 2002, and
- April 2003.

Results

Analysis of the sample distribution by survey firm showed no statistical difference. That is, there was no difference in how the numbers of completed interviews were distributed across MSA statuses and Census divisions for each month. This finding confirms the fact that the contractor carried out the sample design as specified by BTS for each month. Because each survey organization administered a national sample, any house effect found could not be explained by sampling alone.

The distribution of sociodemographics across all variables by survey firm showed no statistical difference for each month. For example, each survey firm interviewed similar proportions of males and females each month. Because each survey firm completed a balanced number of interviews by sociodemographics, any house effect found could not be explained for this reason.

We then tested fifteen different modes of transportation usage for each of six months. This equated to conducting 90 statistical tests looking for differences in reported transportation mode usage by survey firm. Out of these 90 tests, only two transportation modes showed a statistical difference in SPSS. In June 2002 data, the percent of persons who commuted was different across survey firms. For October 2002 data, the percent of persons who used a recreational boat was also different across the survey firms. However, both of these differences were not significant when the tests were conducted in SUDAAN.

Conclusion

There is an apparent lack of a survey organization (house) effect in the OHS estimates, which is supported by the results of the statistical tests. We expected to see house effects in the earlier months of data compared to the later months. We hypothesized that as the survey and procedures were new to the contractors, there would likely be more variation in the estimates in the early months due to house effects. However, after the contractors had established universal survey procedures and practices, and the survey administration became more robust, house effects would be minimized over time. We did not find a trend for house effects over time.

There are some limitations to this study. First, not all months of OHS data were tested. It is possible that house effects could have been found in other months of data. Second, we did not test all possible variables. For example, there were questions we did not test which were opinion or attitudinal based. It is also possible that opinion or attitude questions would have shown some level of house effect. Lastly, we only tested fact-based questions that were not sensitive in nature. We could not test highly sensitive questions because the OHS did not have any. Future research can address these limitations.

The results of this study should not be construed to indicate that house effects do not exist. Rather, the results of the study indicate that two similarly qualified survey firms can collect comparable survey data when a common survey administration methodology is practiced and there are uniform response rates across survey firms.

References

Cohen, SB (1982) "Estimated data collection organization effect in the National Medical Care Expenditure Survey," *American Statistician*, 36:337-341.

Cohen, SB (1986) "Data collection organization in the National Medical Care Utilization and Expenditure Survey," *Journal of Economic and Social Measurement*, 14:367-378.

Cohen, SB and Potter, DEB (1990) "Data collection effects in the National Medical Expenditure Survey," *Journal of Official Statistics*, 6:275-293.

Goldfield, ED, Turner, AG, Cowan, CD, and Scott, JC (1977) "Privacy and confidentiality as factors in survey response," *Proceedings of the Social Statistics Section of the American Statistical Association*, Alexandria, VA.

Mariolis, P and Graber, J (2003) House effects in a telephone survey. Paper presented at the AAPOR annual meetings.

Smith, TW (1978) "In search of house effects: A comparison of responses to various questions by different survey organizations," *Public Opinion Quarterly*, 42:443-463.

Smith, TW (1982) "House effects and the reproducibility of survey measurements: A comparison of the 1980 General Social Survey and the 1980 American National Election Study," *Public Opinion Quarterly*, 46:54-68.