# Are Late/Difficult Cases in Demographic Survey Interviews Worth the Effort? A Review of Several Federal Surveys

Detis Duhart, Bureau of Justice Statistics, Nancy Bates, Barbara T. Williams, and Gregg Diffendal, Bureau of the Census[1], and Pei Lu Chiu, National Center for Health Statistics

**Abstract**

Government statistical agencies currently expend a great deal of time and resources to keep survey nonresponse at or below current levels. This paper uses data from the American Housing Survey - National (AHS-N), the Current Population Survey (CPS), the National Crime Victimization Survey (NCVS) and the National Health Interview Survey (NHIS) to explore the characteristics of late or 'difficult' cases that comprise the last few percentage points of survey response rates. It also examines the household and person-level demographic characteristics of late and early cases. In addition, to determine if critical survey estimates are affected by late cases, we run the estimates with and without these data. Bates and Creighton (2000), in their examination of the NCVS and CPS, found that a portion of late cases are similar to nonrespondents. They also found the magnitude of the difference between estimates with and without these cases is usually small; however, in some cases the estimates are significantly different without the late cases. This study extends their research by examining results from several federal surveys.

## 1. Introduction

Federal statistical organizations strive to conduct household surveys with high response rates in a timely fashion. However, the task is becoming increasingly difficult in the wake of a less cooperative public, the popularity of access impediments (e.g. caller ID and answering machines), and other societal and environmental factors contributing to lower response rates. Considering this trend, it is becoming increasingly important to understand the impact that lower response rates may have on official surveys and whether it is worth the time and effort to pursue the last few cases.

Recent studies by Keeter et al. (2000) and Curtin et al. (2001) have called into question the generally held notion that the higher the nonresponse rates, the more biased the survey data. The Keeter (2000) study found surprisingly few differences between data gathered from identical phone questionnaires -- one using a less rigorous method that achieved a 36 percent response rate and the other a more rigorous method with a 60.6 percent response rate. The Curtin et al. study examined both cross-sectional and time series estimates from the Survey of Consumer Attitudes between 1979 and 1996. Again, somewhat surprisingly, the authors found that one measure, the Index of Consumer Sentiment, changed very little after the initial refusers and higher contact cases were removed from the datasets (thereby allowing comparisons of estimates based on lower response rates).

Our paper explores late interviews across four federal household surveys. We attempt to answer two basic questions surrounding these cases. First, are there demographic similarities among the late

[1]This paper reports the results of research an analysis undertaken by Census Bureau staff. It has undergone a Census Bureau review more limited in scope than that given to official Census Bureau publications. This report is released to inform interested parties of ongoing research and to encourage discussion of work in progress.

interview households across surveys and are they different from the earlier (and presumably easier) cases? Second, would critical survey estimates be impacted if these cases were not represented in the data? That is, what would be the result if we reduced the time and resources to go after these households, and consequently, allowed nonresponse to climb slightly? If we find that these interviews exhibit unique demographic traits, their absence would intrinsically seem to adversely bias certain estimates. For example, if difficult/late interviews tend to overrepresent younger persons, we might see a shift in employment rates or certain health statistics if they are no longer included.

The four surveys studied here include the May 1999 Current Population Survey (CPS) , two quarters of the 1999 National Crime Victimization Survey (NCVS) , the 1999 American Housing Survey - National (AHS-N), and the 1998 National Health Interview Survey (NHIS). Results from the CPS and NCVS draw from a previous paper by Bates and Creighton (2000). Newly documented findings from the NHIS (see Chiu, Riddick, and Hardy, 2001) and AHS (Williams, 2001) are summarized along with findings from the previous two analyses.

## 2. Survey Descriptions

The four surveys studied differ significantly in sponsors, subject matter, and design features. One commonality is that the U.S. Census Bureau acts as the data collection agent in all four surveys. The AHS-N is a longitudinal survey of housing units (both occupied and vacant) sponsored by the U.S. Department of Housing and Urban Development (HUD). The survey asks questions such as home ownership, housing/rent value, number of rooms, and utility costs. The survey typically interviews one adult resident per unit and interviews are conducted by both personal visit and telephone. Interviews are conducted biennially with approximately 46,600 households and the survey is in the field for 16 weeks. The response rate for the 1999 AHS-N was 90.8 percent.

The CPS provides monthly labor force estimates and annual data on poverty, income, and work experience. It is conducted monthly with approximately 50,000 households – interviews are carried out by personal visit and telephone. Most interviews are conducted during a one-week period (Sunday to Saturday) that includes the 19$^{th}$ of the month. Like the AHS-N, the interview is typically conducted with a single household respondent who provides information for other household members. In 1999, the annual average response rate for the CPS was 93 percent.

The NCVS is sponsored by the Bureau of Justice Statistics, is conducted semiannually, and is used to provide estimates of crime victimization. Interviews are conducted at 56,000 households and within a household each person aged 12 and over is interviewed. The survey uses both personal visit and telephone interviews. The annual average response rate in 1999 was 93.6 percent.

The NHIS is sponsored by the National Center for Health Statistics and collects data of general health information annually from the U.S. civilian household population. One adult is interviewed for household questions while other selected adults respond to specific sections of the questionnaire. Interviews are generally conducted by personal visit. In 1998, the household response rate was 90.0 percent. The survey is conducted over a two week span.

## 3. Operationalizing "Late/Difficult" Interviews

Each of the four surveys maintains different degrees of history information regarding number of contact attempts, number of contacts made, and outcome details of each interim contact. Unfortunately, in most cases this type of information is extremely limited or is not kept in electronic format along with the interview data. This, coupled with the fact that each survey varies in terms of field length, means that the definitions used to categorize a case as difficult or late also varied across the four surveys. Table 1 contains a summary of definitions used for each.

The AHS-N is somewhat unique in that it has a long field period (16 weeks from August to mid-November). Interviewers receive their entire assignment for the survey at one time. The instrument captures date of interview but does not collect detailed information on the number of attempts prior to making contact nor does it store interim outcomes of prior contacts (e.g., soft refusal, appointments set for a later date). Consequently, the AHS-N cases could only isolate late interviews based on date of interview. November 4$^{th}$ was the cutoff date used to define the last five percent of interviews – interviews completed after this date were labeled late interviews and essentially cover interviews completed during the last two weeks of data collection. We cannot assume, however, that these were necessarily "difficult" or required an extraordinary number of contacts. However, anecdotal comments from interviewing staff along with monthly cost figures, suggest that interviews conducted during the last month are the most costly, presumably because they required a greater number of trips due to inability to contact or required refusal conversions. Late cases (as defined here) accounted for 4.9 percent of the completed interviews.

When constructing the definition for CPS, we had access to one additional variable: number of actual and attempted personal contacts. However, the instrument only records this information for personal visits (CPS allows telephone interviews for certain months in sample). Additionally, for the counter to function properly, interviewers must access and open the case from the laptop. We have no way to know how often interviewers may visit an address, fail to make contact, but never open the laptop. Such cases will underestimate the number of contacts. Late cases for CPS were defined as those conducted during the Tuesday or Wednesday of the second week of data collection. Difficult cases were those requiring four or more personal contacts. After combining these two criteria, 4.6 percent of the May 1999 CPS interviews met the definition of a late or high contact (difficult) case.

The NCVS routinely collects information on the record of personal visits and telephone calls made for each case. However, this information is stored separately from the electronic survey data. Consequently, like the AHS-N, the NCVS definition was limited to date of interview. Late interviews were defined as those completed from the 17$^{th}$ of the month until regional office closeout (approximately the last seven days of data collection). These cases comprised 5.2 percent of the person-level interviews for the two quarters of 1999 data examined.

The NHIS used a combination of late and difficult interviews. Late interviews were those requiring additional time beyond the standard two week field period (because of refusal conversion efforts or prolonged non-contact cases). Difficult-to-contact interviews were defined based on the number of interviewer-reported personal visits required to complete the case. Cases requiring 10 or more personal visits were considered difficult to contact. Late cases combined with difficult cases

accounted for approximately 7 percent of the 1998 NHIS interviews. It should be noted that there was surprisingly little overlap between the two categories – only around 12 percent of the cases completed beyond the normal two week period also required more than 10 interviews.

**4. Limitations**

Before a discussion of findings, we caution of four serious limitations to the study data and analysis. First, the lack of consistent and detailed interview case-history information results in ambiguities surrounding the conceptual and operational definition of late/difficult interviews. For example, without knowledge of the number of contacts (attempted and made) and without interim attempt outcome codes (no one home vs. soft refusal vs. appointment made), how does one go about differentiating the truly difficult interviews from those that were merely hard to reach? Since we lack such information, our definitions are admittedly open to question. There are many reasons why late interviews and those requiring multiple contacts are not necessarily the same as reluctant or resistant interviews. For example, how far into the field period the interviewer makes the first call, the number of calls by day, the size of the household, interviewer workload and number of refusal conversions in the interviewers' assignment all contribute to how early or late a case may be completed. And, in fact, Groves and Couper's work (1998) heavily reinforces the notion that separating out different types of survey nonresponse (in particular, noncontacts and refusals) is crucial to understanding the phenomenon.

Unfortunately, having only date of interview (and in a few instances number of personal visits) precludes us from isolating prolonged noncontact groups from initial refusal groups, both of whom may end up in our definition of a late interview. Nonetheless, since these are the real conditions our surveys are currently subject to, our only alternative is to combine them under the somewhat ambiguous and undifferentiated label of "late." We concede this weakness but stop short of declaring it a fatal flaw to our study.

A second limitation stems from the high response rate all four surveys routinely achieved for the periods studied. Compared to non-government surveys, the four surveys have above average response rates (in the low 90's). Consequently, the removal of 5-7 percent of the late/difficult interviews still yields a high response rate by industry standards. Readers are cautioned that the last few interviews from a survey with a response rate in the 90's may be very different from the last few from a survey with an initial response rate of say, 60-70 percent.

Third, we explore later in the paper whether survey estimates might be different with and without late/difficult interviews included. Our method for doing so is to compare selected estimates based on all the interviews to those re-calculated without the late/difficult cases. Since the percentage of late cases is small, the overlap of cases used to produce both estimates is very large making for a high correlation between estimates. Consequently, the variances are close to zero and therefore even slight differences between estimates will test as significantly different. Our intent is not to make concrete inferences regarding estimates in the absence of late interviews but rather to take a broad view across all four surveys studies to see if any meaningful patterns of differences are detected. We also make a point of interpreting the differences from a practical or meaningful standpoint, not merely from a statistical one.

## 5. Results

Table 2 summarizes relationships between selected household demographic characteristics and the propensity of being a late/difficult interview. Findings reported for the AHS-N are the result of bivariate crosstabulations while those from the CPS, NCVS and NHIS are the result of multivariate models predicting the likelihood of being in a late/difficult interview controlling for other independent variables. For a more detailed discussion of the variable descriptions and methods, see Bates and Creighton (2000), Chiu, Riddick and Hardy (2001) and Williams (2001).

In all four surveys, it is clear that late/difficult interviews are less likely to contain persons aged 55+ (or 65+ as defined in the NHIS). This is probably related to ease of contact where interviewers have greater success finding someone at home in households containing older persons. Topic saliency of the NHIS may also explain earlier cooperation rates among seniors with older persons have more health-related incidents to report and therefore greater interest the survey.

Conversely, single person households were more likely to fall into the late/difficult category for both the AHS-N and the NHIS (the CPS and NCVS study did not include this variable). This is consistent with the survey nonresponse literature that suggests non-contact rates are much higher for single person households (Groves and Couper, 1998). This premise is also supported by additional breakouts of the late versus difficult interviews in the NHIS. Chiu et al. (2001) report that single person households comprise about 36 percent of the interviews requiring 10 or more visits (difficult interviews) compared to 28 percent of the interviews conducted after formal closeout (late interviews).

We see a positive association between renting and being a late/difficult interview in all four surveys. This may be a reflection of the types of people who reside in rented apartments and other multi-unit structures, that is, younger people or single person households. Such structures may also have a greater incidence of access impediments such as doormen or locked entry ways requiring access codes.

The relationship between race/ethnicity and the likelihood of being a late/difficult interview is less obvious than the previous characteristics in Table 2. For three of the four surveys, Hispanic origin was not found to be significant but in the AHS-N, CPS and NCVS, Blacks were disproportionately found more often in the late/difficult interview category. For the NHIS, more Hispanics and Blacks were found in the late/difficult interviews, however the effects of race and ethnicity became non-significant once other geographic and social environmental factors are controlled. We should also mention that while race was significant in the models run for the CPS and NCVS, these models were run at the person-level and did not control for variables such as household size and urbanicity. Thus, our findings are somewhat inconsistent regarding the degree to which race is associated with late/difficult interviews, but as a whole, suggest race may be a weaker predictor than others.

Previous research by Curtin, Pressor, and Singer (2000) suggests that lower socioeconomic households may disproportionately comprise initial refusals while more affluent households are disproportionately non-contacts. All four surveys contained some measure of wealth, albeit they were quite different. The AHS-N examined categories of household income as well as housing value

of owner occupied units. For both measures, affluent households were over-represented in late interviews. Likewise, the CPS and NCVS both found positive relationships between higher levels of household income and the likelihood of being of late/difficult interview. For the NHIS, the effect of income was not a significant predictor once other factors were held constant. However, it is interesting to note that in the CPS, NCVS and NHIS, the proclivity to refuse income information (or answer DK) was higher in late/difficult interviews. This correlation between item nonresponse and "lateness" could reflect some degree of reluctance or reduced cooperation in these interviews.

Three of the surveys included urbanicity in their investigation of late/difficult interviews. All three found that households located in urban areas are more likely to fall into the late/difficult category. There are many explanations for this finding including fear of crime, population density, and the types of households that populate urban areas (e.g., single person households, childless households and younger households).

Based on Table 2, can we draw any conclusions whether the late/difficult interviews resemble nonrespondents? We revisit the question posed at the beginning of the paper, that is, are there demographic similarities among the late/difficult interview households across surveys and are their characteristics different from the non-late (and presumably easier) cases? With the exception of ethnicity, we found evidence of significant differences between the characteristics of late/difficult cases and the non-late/non-difficult cases. Further, the surveys exhibited similarities in their pattern of differences.

Late/difficult interviews are less likely to contain older household members but more likely to reflect single person households, households that rent, are more affluent, have Black household members, and/or are located in an urban area. With the exception of race and tenure, these characteristics have been documented by Groves and Couper (1998) as predictors of nonresponse in household surveys. But, given the somewhat arbitrary nature of our definitions and other data limitations, we refrain from making the assumption that our late/difficult interviews are indicative of noninterviews. More likely, our late/difficult interviews represent some combination of hard-to-contact and initial refuser households that eventually do cooperate if field resources permit the extra efforts required to complete them. Therefore, we can more safely assume they represent *potential* non-respondents, but not necessarily those who never participate.

## 6. Impact on Estimates

In this section, we address the second question of interest – would survey estimates be impacted significantly if late/difficult interviews are not included? We begin illustrating the potential differences by presenting distributions of critical survey estimates by non-late/non-difficult cases versus the late/difficult cases (see Table 3). For the AHS-N we present homeownership rates and housing stress rates. Housing stress is defined as a high level of housing costs as a percent of current income. We use HUD's definition of high levels as being 31 percent or greater. For the CPS, we include a distribution of labor force participation which is used to calculate the unemployment rate. For the NCVS we examine personal and property crime victimization incidents (number of reports and type of crime). For the NHIS we examine health insurance coverage rates and limitation of physical activity.

The distributions illustrate various degrees of difference between non-late and late interviews. For example, late cases exhibit lower homeownership rates but a slightly higher housing stress rate. Differences in number of reported crimes and type of crime by type of interview are slight, but larger disparities are evident in both labor force participation and the health-related items. Late interviews reflect a higher percentage of persons currently in the labor force and fewer retirees. Late interviews also reflect a lower rate of insured persons but also a lower rate of persons with some type of health-related physical limitation. These findings follow logically from the age disparities noted previously between early and late interviews.

Our last step was to compare estimates of population characteristics from Table 3 calculated with and without the late/difficult interviews. In order to check for significant differences, late cases were converted to noninterviews and the data were re-weighted for nonresponse. Tests for significant differences between estimates with and without late cases were calculated using the standard error for late cases only since the overlapping cases from both sets add nothing to the variance. Survey design effects were taken into account when testing for differences. To reduce the possibility of Type I error and compensate for the high correlation between estimates, we used an alpha of .01 as the threshold for determining statistical significance between estimates (as opposed to the Census Bureau standard of .10).

Across all estimates, the magnitude of difference is very small between estimates with and without late/difficult cases. This is somewhat expected considering both sets contain anywhere from a 93-95 percent overlap. Thus, the practical impact of excluding these 5-7 percent of cases appears minimal. However, for some measures (notably the unemployment rate), a change in the magnitude of one-tenth of one percent can still be meaningful and impact policy-making. Additionally, the direction of differences are interesting to note. For example, housing stress rates are lower without late cases while unemployment rates are higher. Crime rates are consistently higher without late cases. And finally, physical limitation estimates of the population are slightly higher without late cases represented. Differences in non-late and late interview characteristics such as age may explain these patterns.

## 7. Practical Implications

Despite differences in operational definitions across surveys, we discovered some commonalities in the household characteristics of late/difficult interviews. These discoveries hold some practical implications when trying to keep nonresponse in check. For example, interviewers should determine attributes about the sample household as early as possible. Cases having characteristics of late/difficult interviews such as urban addresses, multi-unit structures, and efficiency apartment complexes, should be assigned early on since they may house a disproportionate number of single person households, young households, and renters. If interviewers observe proxy indicators of easy contactability and cooperation (such as neighborhoods or enclaves of older residents), they may wish to start these later in the interview period, thus allowing more time for the hard to contact.

A second practical implication is to study the correlates of late/difficult interviews and consider the bias they may have on certain survey estimates. For example, the impact of losing a disproportionate number of urban interviews will have different repercussions on a housing survey

versus a crime survey. Likewise the loss of younger people may effect an employment survey differently than a health survey. Knowing what kinds of households are likely to be late/difficult cases also helps us understand how to design better nonresponse adjustments in the absence of these cases. They also suggest how an extra 5 percent nonresponse may effect the degrees of bias for different age and race subpopulation estimates.

Finally, a critical recommendation became obvious as a result of preparing this paper -- agencies should begin to routinely collect and retain case history information of interviewer contact attempts. This should include number of contacts and attempted contacts, interim outcomes that differentiate non-contacts from initial refusers, and final outcome codes that differentiate the same. Admittedly, this information has been difficult to capture under the DOS-based CAPI environment currently used at the Census Bureau. However, as Census migrates its CAPI surveys to the Windows-based Blaise software, the task will hopefully become easier. Such information will provide a wealth of data currently unavailable and allow us to disentangle the truly difficult interviews from the harder to contact. If this can be achieved, we suggest repeating the analysis presented here to more accurately assess the impact that growing nonresponse rates may have on survey quality, nonresponse bias, and official survey estimates.

**References**

Bates, N. and Creighton, K. (2000). "The Last Five Percent: What Can We Learn From Difficult/Late Interviews?" *Proceedings of the Section on Government Statistics and Section on Social Statistics*, American Statistical Association, pp. 120-125, Alexandria: VA.

Chiu, P., Riddick, H. and Hardy, A. (2001). "A Comparison of Characteristics between Late/Difficult and Non-Late/Difficult Interviews in the National Health Interview Survey." A paper to be presented at the American Statistical Association Joint Statistical Meetings, August 5-9, Atlanta, GA.

Curtin, R., Presser, S. and Singer, E. (2000). The Effects of Response Rate Changes on the Index of Consumer Sentiment. Public Opinion Quarterly, Vol 64, 4, pp. 413-428.

Groves, R. and Couper, M. (1998). Nonresponse in Household Interview Surveys. New York: Wiley.

Keeter, S., Miller, C., Kohut, A., Groves, R. and Presser, S. (2000). Consequences of Reducing Nonresponse in a National Telephone Survey, *Public Opinion Quarterly*, Vol. 64, 2, pp. 125-148.

Williams, B. (2001). "Comparison of Early and Late Interviews from the 1999 American Housing Survey," U.S. Census Bureau internal memorandum.

**Table 1. Definition of Difficult/Late Cases**

| Survey | Number of Contacts Cutoff | Date Cutoff | % "Late/Difficult" Cases |
|---|---|---|---|
| AHS -N | not available | Nov. 4th (last 2 weeks of data collection) | 4.9% |
| CPS | 4+ personal visits | Tues.& Wed. of 2nd week | 4.6% |
| NCVS | not available | Approx. 17th of the month (last week of data collection) | 5.2% |
| NHIS | 10+ contacts | 16+ days after assignment | 7.0% |

**Table 2. Demographic/Household Characteristics and How They Relate to Late/Difficult Interviews**

| | Survey | | | |
|---|---|---|---|---|
| Characteristic: | AHS-N[a] | CPS[b] | NCVS[b] | NHIS[b] |
| Older person(s) in household | — | — | — | — |
| 1-person household | ++ | | | ++ |
| Renters | ++ | ++ | ++ | ++ |
| Hispanic member(s) | ++ | n/s | n/s | n/s |
| Black member(s) | ++ | ++ | ++ | n/s |
| High Income/Housing Value | ++ | ++ | ++ | n/s |
| In central cities | ++ | + + | | ++ |

[a] associations based on bivariate analysis. [b] associations based on multivariate analysis.
— = significant, negative relationship, ++ = significant, positive relationship, n/s = non-significant relationship

**Table 3. Distribution of Selected Survey Items by Non-late/Non-difficult and Late/Difficult Interviews**

| | **AHS-N** | | **CPS** | | **NCVS** | | **NHIS** | | p-value |
|---|---|---|---|---|---|---|---|---|---|
| Survey Measure | Not late | Late | Not late | Late | Not late | Late | Not late | Late | |
| **Homeownership**: | | | | | | | | | |
| Owner | 67% | 58% | | | | | | | |
| Renter | 31% | 40% | | | | | | | |
| No cash rent | 2% | 1% | | | | | | | <.001 |
| **High Housing Stress**: | | | | | | | | | |
| Yes | 27% | 30% | | | | | | | |
| No | 73% | 70% | | | | | | | <.01 |

| (Table 3 con't.) | | | Not late | Late | Not late | Late | Not late | Late | p-value |
|---|---|---|---|---|---|---|---|---|---|
| **Labor Force Status:** | | | | | | | | | |
| Employed -at work | | | 62% | 69% | | | | | |
| Employed -absent | | | 2% | 2% | | | | | |
| Unemployed | | | 3% | 3% | | | | | |
| Retired | | | 16% | 8% | | | | | |
| Disabled | | | 4% | 3% | | | | | |
| Not working -other | | | 13% | 15% | | | | | <.001 |
| **# Crime Incidents:** | | | | | | | | | |
| None | | | | | 89% | 89% | | | |
| One | | | | | 9% | 9% | | | |
| Two | | | | | 1% | 2% | | | |
| Three or more | | | | | 1% | 0% | | | >.01 |
| **Type of Crime:** | | | | | | | | | |
| Personal | | | | | 24% | 23% | | | |
| Property | | | | | 76% | 77% | | | >.01 |
| **Health Insurance:** | | | | | | | | | |
| Yes | | | | | | | 85% | 80% | |
| No | | | | | | | 14% | 17% | |
| DK/Refused | | | | | | | 1% | 3% | <.001 |
| **Activity limitation**: | | | | | | | | | |
| Yes | | | | | | | 13% | 8% | |
| No | | | | | | | 87% | 92% | <.001 |

**Table 4. Estimates with and without late/difficult interviews**

| | Estimates for Total Population | |
|---|---|---|
| | All Cases | Without Late Cases |
| Homeownership Rate | 66.92% | 66.90% |
| Housing Stress Rate | 27.25% | 27.10%* |
| Unemployment Rate | 4.05% | 4.10%* |
| Person Crime Rate[a] | 36.8 | 37.2 |
| Property Crime Rate | 209.1 | 212.0** |
| Violent Crime Rate | 35.8 | 36.2 |
| Health Insurance Rate | 84.33% | 84.44% |
| Activity Limitation Rate | 12.65% | 12.80%** |

[a] Crime rates reflect victimization rates per 1,000 people

Difference significant a the *.01, **.001 level.