

REASSESSING CASE DISPOSITION STATUS IN A LARGE NATIONAL SURVEY WHEN RESPONDENTS REFUSE TO REPORT THEIR NAMES

Steven L. Botman, John Horm, NCHS

Steven L. Botman, NCHS, 6525 Belcrest Rd., Hyattsville, MD 20782

KEY WORDS: Survey design, response rates, data quality, data collection

INTRODUCTION AND SUMMARY

The National Health Interview Survey (NHIS), a continuous annual survey of 45,000 households conducted by the National Center for Health Statistics (NCHS) a component of the Centers for Disease Control, usually obtains a household response rate of about 95 percent; however, about 2 percent of the NHIS respondents refuse to provide their name. The NHIS attempts to collect the names of all persons residing at the NHIS sample address (1) in household rostering and (2) in the collection of the information to permit the matching of survey data to other records in the Department of Health and Human Services. In a sense when faced with a reluctant respondent the NHIS, thus, allows anonymous reporting. We do not know why the respondents do not provide names. We do not think its confidentiality related, because the NHIS provides respondents with a strong assurance of confidentiality. For whatever reason some individuals may not positively respond to requests for information in a Federal survey--perhaps they do not see such activities as a useful civic obligation. Currently the NCHS classifies the sample disposition of cases regardless of whether the NHIS household respondent provides the name of each person.

Using the 1990 NHIS, this paper explores some effects of this decision to permit anonymous reporting, in part based on the quality of NHIS response. Also examined are the correlates of and the clustering of NHIS sample by respondent's provision of the names.

In addition, the paper examines the sample disposition status not for retention in the NHIS but for retention in the NHIS sample used as a sampling frame for other surveys. The NHIS respondent sample has been re-used for surveys with smaller samples such as the National Survey of Family Growth and the Medical Expenditure Panel Survey. That is, the NHIS sample disposition code could be tailored based on the intended use.

In compiling sample frames for other surveys, the decision on NHIS sample disposition has survey quality as well as survey cost implications. NHIS cases without a reported name when selected for sample re-use often have both lower subsequent response rates and higher subsequent costs per case. In such NHIS sample re-use, for relatively and absolutely few cases we may not have enough information for location and identification--we, for example, may not have exact age. These cases higher cost is due in part to the additional expense in

tracing and locating movers between survey interviews. The adverse effect on data quality reflects both potential increases in nonresponse rates if such cases are retained or decreases in survey coverage rates if such cases are dropped from the frame in sample re-use.

This problem in case identification without a name is aggravated with a long lag between the two survey interviews; with a short lag, case identification and location become much less troublesome. A shorter lag between interviews, however, does not help in obtaining a response for another survey if the lack of a name at the initial interview was due to respondent hostility.

In section 1 we discuss some tabulations showing how such cases are clustered in the sample. In section 2 we look how some variables in the NHIS data set correlated with a respondent not providing a name. In section 3 we look at CH-AID analysis for refusal to provide name. In section 4 we summarize the paper and presents some tentative recommendations.

It should be noted that we observed variability in how interviewers handled the refusal of name, despite interviewer instruction on how this information was to be entered. Some respondents said they refused to provide such information and the NHIS interviewers recorded this information. Other respondents provided names such as John Doe and the interviewers recorded this information. In this report we assume that refusals to the request for names were reported with last name as either refused or Doe. There could have been other variations in names that we did not identify as refusals, such as John Smith. The NHIS conversion to computer assisted data collection in 1997 should reduce the frequency of such problems.

Section 1. Clustering of cases without a name

Table 1 shows the distribution of these cases by age, sex, and race. There's only a slight variation. Although not shown in this table, about 2 percent of Hispanic persons did not report name.

This lack of variation by basic demographic variables is not surprising. We expect to find systematic nonresponse for the request for respondent name by household not person characteristics.

To confirm this, we then looked at the distribution of these cases by household. Table 2 shows the results the distribution of the 46,443 households in the 1990 NHIS sample by the household total and by the number of persons not providing names. It is seen that in most cases NCHS obtains the names of all persons in the household or none.

In NHIS data collection the interviewers separately enumerate individuals or separate families that reside in the same household. For the purpose of the NHIS, an individual is someone not related to other household residents. We have not checked whether the provision of some but not all names in sample household is related to whether more than one family resides in the household.

Table 1. Percent of the 1990 NHIS sample aged 18 years and over without a reported name according to age, sex, and race.

Race	Sex	Age	%
.	.	.	2
.	Male	.	2
.	Female	.	2
Black	.	.	2
Nonblack	.	.	2
Black	Male	.	2
Black	Female	.	2
Nonblack	Male	.	2
Nonblack	Female	.	2
Black	Male	18<=age<=19	2
Black	Male	20<=age<=24	0
Black	Male	25<=age<=29	2
Black	Male	30<=age<=34	2
Black	Male	35<=age<=44	3
Black	Male	45<=age<=49	1
Black	Male	50<=age<=54	2
Black	Male	55<=age<=64	2
Black	Male	65<=age<=74	1
Black	Male	75<=age	1
Black	Female	18<=age<=19	2
Black	Female	20<=age<=24	2
Black	Female	25<=age<=29	2
Black	Female	30<=age<=34	3
Black	Female	35<=age<=44	3
Black	Female	45<=age<=49	1
Black	Female	50<=age<=54	1
Black	Female	55<=age<=64	2
Black	Female	65<=age<=74	1
Black	Female	75<=age	1
Nonblack	Male	18<=age<=19	3
Nonblack	Male	20<=age<=24	2
Nonblack	Male	25<=age<=29	2
Nonblack	Male	30<=age<=34	2
Nonblack	Male	35<=age<=44	3
Nonblack	Male	45<=age<=49	2
Nonblack	Male	50<=age<=54	3
Nonblack	Male	55<=age<=64	2
Nonblack	Male	65<=age<=74	2
Nonblack	Male	75<=age	1
Nonblack	Female	18<=age<=19	2
Nonblack	Female	20<=age<=24	2
Nonblack	Female	25<=age<=29	2
Nonblack	Female	30<=age<=34	2
Nonblack	Female	35<=age<=44	2
Nonblack	Female	45<=age<=49	3
Nonblack	Female	50<=age<=54	3
Nonblack	Female	55<=age<=64	2
Nonblack	Female	65<=age<=74	2
Nonblack	Female	75<=age	2

In some of the cases where we have partial household response to the request for names, we may be enumerating separate families within household units. We could but have not checked whether this is the situation.

Table 2. Number of households in the 1990 NHIS sample by the number of persons (aged 18 years +) in the household and the number (aged 18 years +) providing names.

Persons in Household	Persons In Household Not Reporting Name	Households
Total	-	46,443
1	0	14,839
1	1	374
2	0	24,223
2	1	56
2	2	526
3	0	4,590
3	1	25
3	2	8
3	3	67
4	0	1,267
4	1	10
4	2	3
4	3	4
4	4	24
5	0	308
5	1	12
5	2	1
5	4	2
5	5	5
6	0	73
6	1	2
6	3	1
6	6	1
7	0	16
8	0	5
9	0	1

Table 3 displays the number of second-stage sampling units (SSSUs) in the NHIS sample based on the number of households with one or more persons without a reported name. Members of a SSSU are often similar in socio-economic characteristics. On the average an NHIS SSSU contains about 6 households with completed interviews.

Table 3 indicates only slight clustering of this characteristic with the design. Sixteen SSSUs out of 8,112 SSSUs in the design had 3 or more households without reported names.

Table 3. Number of second-stage sampling units (SSSUs) in the NHIS according to the number of households (HHs) with a person not reporting name

Households not Reporting a Name per SSSU	SSSUs
.	8,112
0	7,112
1	895
2	89
3	16

Section 2. Variables related to refusal of respondent name

The refusal to provide a name, however, is disproportionately related to the particular response to several other variables in the NHIS. The variables with such predictive ability fall into three general classes. First, variables related to NCHS's ability to recontact the household. Second, variables generally considered to be non-threatening and used as basic analytic covariates. Third contextual variables related to the interview itself, such as the respondent's proxy status. If proxy status is unknown, we have a potential measure of the interviewer's stress during the interview.

An example of the first kind of variable is the household respondent's refusal to provide an available telephone number. In this case the household respondent may consider such a request to invasive, perhaps the number itself may be unlisted. In fact the respondent may not want to be recontacted for any additional information. Even if in sample re-use we have no difficulty in identifying and locating the individual identified in the NHIS, contacting that person is more difficult without a telephone.

An example of the second kind, is educational attainment, health status, annual family income above or below \$20,000, height, and kind of industry that the persons is employed in. These are basic survey covariates, although not apparently threatening in detail or sensitivity. They may reflect latent hostility in the interview. The provision of an unknown response to these question, however, is clearly related to the likelihood for an unknown response to the request for respondent name.

Table 4 shows that an unknown to such variables is reveals a strong correlation to also having an unknown response for name.

Table 5 looks at this same matter more directly. If name is unknown, the percent of unknowns on several basic variables was quite high. Education, income and health status are basic variables in the NHIS. The NHIS has typically not imputed for missing data items. If

imputation for missing data items was done, such variables would likely have to be done first.

Table 4. Percentage of persons with an unknown response for selected variables that did not provide their name

Data item	Percent not providing name
Telephone number	25
Education	41
Income*	29
Industry	26
Health Status	20
Proxy status	37

* Above or below \$20,000 annually

Table 5. Percentage of persons with unknown response for selected variables by provision of name status

Data item	Name known	Name unknown
<u>Percent with unknown response category</u>		
Telephone number	0.4	4.6
Education	0.7	20.7
Income	2.3	41.3
Industry	0.0	0.1
Health status	0.3	3.4
Proxy status	0.7	18.0

However, while such variables can be used to identify cases disproportionately not reporting a name, Table 5 shows the distribution of such cases. Unknown responses for such variables have high predictive ability for a refusal for name; however, for cases with unknown names, except for income, respondents are not consistently reporting unknowns for other variables. The NHIS interview is thus of lower quality for these cases without a respondent name but still includes a definite response for such key survey variables.

Section 3. Variables predicting refusal of name

To begin identifying variables and interrelationships among variables predicting no name in the NHIS. We used a CH-AID procedure. We wanted to see whether the variables so identified would be those affecting data quality on the typical NHIS data items, affecting variables used for case contact, or measuring respondent hostility.

Figure 1 provides a CH-AID analysis. The first variable is provision of annual family income above or below \$20,000. This is useful as an analytic covariate and one that indicates respondent hostility. This data item hardly seems detailed enough to be intrusive, for many respondents this information could be inferred by the NHIS interviewer.

Among those providing a definitive response to annual family income above or below \$20,000, 1 percent refused to provide respondent name. Of the 2,766 respondent whose response was unknown to this item, 28 percent refused to provide their names. These account for almost 1/3 of all cases without a respondent name and likely were difficult interviews to conduct. In the NHIS we have some administrative information on the interview including proxy status. Among the 2,766 respondents, 343 had proxy status, an interviewer completed item, unknown. This likely measures interviewer stress during the interview.

For persons who did report annual family income above or below \$20,000, disproportionate refusal to provide respondent name was found among those persons who had a telephone but did not provide the number. One scenario for these cases is that they consented to the NHIS interview but not any re-use or re-contact. The troublesome part of this CH-AID analysis is that some variables related to sample re-use and data quality are showing up high in the tree.

Section 4. Summary and Recommendations

In the paper, the lack of reporting for name on the NHIS interviewer has been associated with a higher item unknown rate for several variables. However, the lack of name has not been shown consistently related to item unknowns concurrently for other variables. The NHIS cases without a reported name are thus providing information for the NHIS but we have no direct way to measure the accuracy of item responses for these cases

other than item nonresponse, but we speculate that they may be of lower quality. However, it may make sense in the processing of the entire NHIS for NCHS to carefully look at case disposition status for respondents with unknown simultaneously as a response for several key variables.

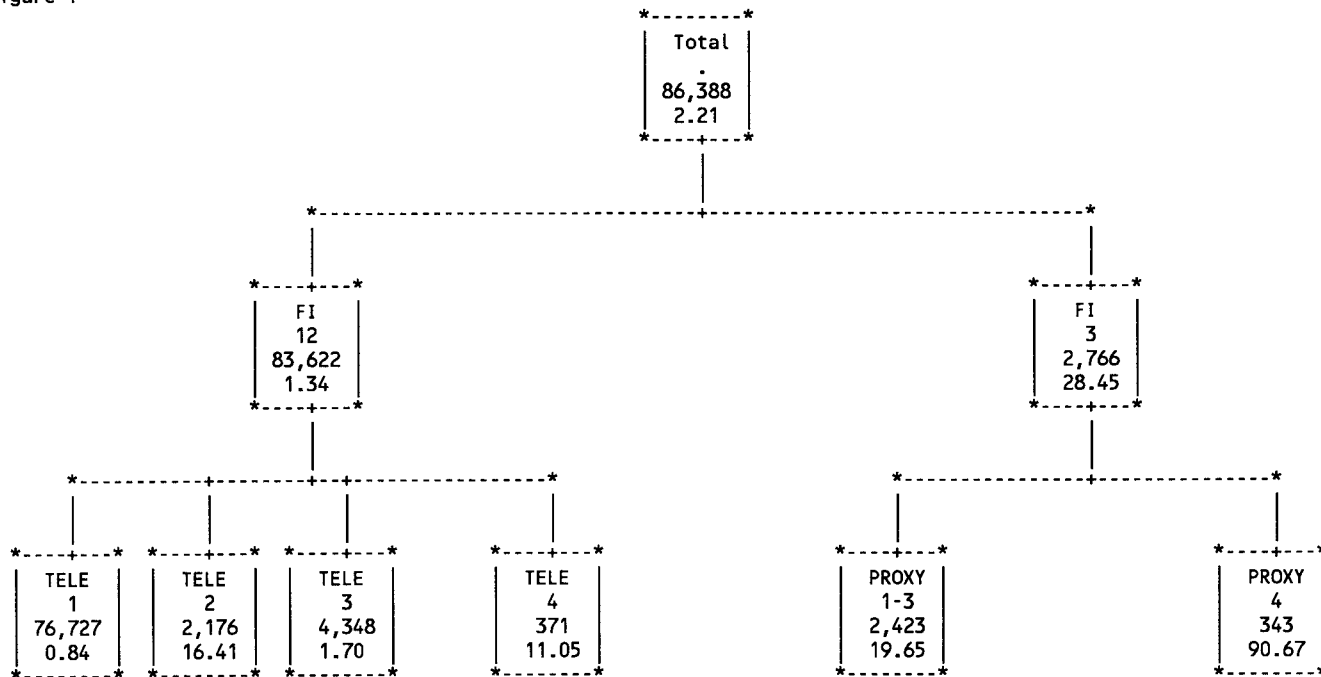
For the NHIS sample re-use the lack of name, however, causes a number of survey problems in sample identification and location, especially with long lags between the two interviews, or in potential respondent hostility, or both. These variables are occurring high in the CH-AID analysis. With many of these cases being quite expensive to secure complete interviews, design of such linked designs should consider classifying these cases as nonrespondents in the NHIS frame.

Table 6. Number of unknown response to selected socio demographic characteristics according to NHIS respondent's provision of name

Number of variables with Unknown as Response*	With Out Name (%)	With Name (%)
.	100	100
0	50	94
1	23	4
2	7	0
3	7	0
4	9	0
5	2	0
6	1	0

* Family income above or below \$20,000, health status, educational attainment, class of worker under employment status, Veteran's status, and height

Figure 1



715

FI = 1 (Annual family income GE \$20,000); 2 (LT \$20,000); 3 (Unknown). Tele = 1 (household had telephone and number reported); =2 ((household had telephone and number not reported); 3 (no telephone number); 4 (unknown). Proxy = 1 (person present during interview); 2 (person present part of interview); 3 (proxy); and 4 (unknown)