

MONITORING SCREENING PROCEDURES FOR THE HEALTH INTERVIEW SURVEY

Christine Milliken, Bureau of the Census
Washington, DC 20233

KEY WORDS: Evaluation, Cost Effectiveness, Efficiency

Introduction

Beginning in January 1995, The Census Bureau implemented a new screening procedure for The National Health Interview Survey (NHIS). This new screening procedure must be monitored closely during the early stages of implementation to determine its effectiveness in achieving the desired sample in addition to its impact on survey interviewing costs. This paper will outline the extensive screening evaluation plan developed by the Census Bureau to monitor the screening activities of Field Representatives nationwide. The screening results are based on production data received from the twelve Census Regional Offices (ROs).

Background

The National Center for Health Statistics (NCHS) sponsors the NHIS which collects health information on the population of the United States. The sample size is approximately 45,000 housing units interviewed annually. Data is obtained about health, doctor visits, hospitalization, and other selected information for household members.

The 1990 sample design of the NHIS is unique because it combines screening of sample units with oversampling of minorities in higher minority strata. The ten year sample was selected by taking two or three clusters per sample designation depending on the minority density of the combined block that the sample has been selected in. (WG-9, 1991)

During the initial NHIS old construction sampling, each hit was assigned to an NHIS oversampling stratum code. The stratum code definitions can be found in Table 1. The stratum code determines the number of measures per sample within a hit and the screening sampling rate. The screening sampling rate enables us to designate cases to be screened ("S" cases) or interviewed regardless of the Black/Hispanic status of the household ("I" cases). (WG-9, 1991)

Specifying a case to be screened means that if there is not a Black or Hispanic person living within the household, the unit will be screened out of sample. This screening of sample units can be accomplished through the use of neighbors. Currently, our field procedures require FRs to visit "S"

households twice on two different days. If no one is home, the FR may go to a "neighbor" to determine the minority status of the sample unit. If the neighbor doesn't know the sample unit's race and ethnicity, or if the neighbor says the sample unit's occupants are Black or Hispanic, the FR must continue her attempts to contact the sample unit. If the neighbor says no Blacks or Hispanics occupy the sample unit, the FR must visit another neighbor for confirmation. If two neighbors agree, then the FR may screen out the sample unit.

Not all sample units are designated to be screened. A proportion of units will be assigned to be interviewed with certainty ("I" cases). All new construction sample units are designated as "I" because we do not screen in the permit frame.

Combined with oversampling, screening allows us to include Blacks and Hispanics in sample at a higher rate than nonminority households. The final result is that we will have more interviews obtained from Black/Hispanic households. Thus, the statistics that NCHS publishes relating to health issues of Blacks and Hispanics will be more reliable.

Since field screening is a new process, an evaluation is necessary. Prior testing of screening has been fairly limited in its scope, and it's important to monitor the data received from the regions to see if our assumptions on screening are true. Based on our findings, the screening populations could theoretically be expanded in the future to include other minority groups such as American Indian, Eskimo, or Asian/Pacific Islander.

Evaluating neighbor screening is also important to determine its implications on costs. If survey interviewing costs are increasing significantly because of the use of neighbors to screen sample units, we will not want to continue this procedure. We would have no way of knowing exactly what FRs are doing without this evaluation.

Evaluating screening will also show us where our assumptions about the existence of specific population groups are no longer valid, or where they were incorrect. For example, if we have oversampled too much in areas where there are no longer any Blacks or Hispanics, we will find that a large proportion of our sample units are screening out. The FRs who are working these areas will have smaller than normal assignment sizes. Contrarily, we might find that a

larger proportion of units than we expected are screening into sample thus creating an increased FR workload. If an FR is traveling to get to her assignment area, this increase in workload could force the FR to travel additional days consequently increasing the cost of this assignment. If this is happening frequently, it would directly impact survey interviewing costs.

This evaluation could have a direct impact on the procedures affecting field representatives. Likewise, it could also impact the screening sampling done by the ROs by allowing them to designate more (or fewer) households to be screened. We expect that changes made as a result of this evaluation will enable us to obtain the desired goal of more reliable data for the Black and Hispanic populations without a drastic increase in survey interviewing costs.

Our requirements for the neighbor screening and sample household screening field procedures were that they:

- be easy to implement
- be generic for all parts of the country
- discourage Field Representatives from falsifying neighbor information in order to screen out households that they did not want to interview (dangerous areas, etc.)
- save costs
- not allow for bias in the survey by screening out sample households (using neighbors) that should have been interviewed

Neighbor screening is highly controversial, and potentially dangerous to the integrity of the sample. FRs are visiting non-sample households, and asking about the race and ethnicity of their neighbors. Since the NHIS could be interviewing in this same area for the next ten years, relationships that the FR builds with neighbors can affect a potential interview a few years later.

By the summer of 1992, some preliminary plans had been developed for how we would conduct neighbor screening. At that time, however, it was undecided on whether neighbor screening would even be feasible for such a large scale survey. Our objectives for the hothouse tests were to test different sets of neighbor screening procedures and neighbor screening questionnaires before the larger screening pretest scheduled for 1993.

We conducted six hothouse tests in three different regions. Afterwards, we held a debriefing in which the Field Representatives shared their experiences in the field. Based on this hothouse test, we were able to draw a few conclusions. First, neighbor screening was feasible. FRs liked going to neighbors when the sample household occupants were not available, and

neighbors were able to report race/ethnicity fairly accurately. Second, neighbor screening worked best when the rules on what defines a "neighbor" were somewhat flexible. For example, FRs preferred an approach where they could find a suitable neighbor in thirty minutes - they did not work well when a structured approach for neighbor contact was given to them. Third, and perhaps most important was the fact that our hypothesis about different neighborhoods "acting" differently with the same approach was correct. Neighborhoods vary greatly in their characteristics and these differences influence how well neighbors know each other and report on race/ethnicity status. (Hayes, 1992)

Because neighbor screening was supported by our data from the hothouses, we developed a larger scale screening pretest during the summer of 1993. The objectives of the pretest differed slightly from the hothouse tests because we intended to include testing of sample household screening. This was in addition to testing of neighbor responses for accuracy on race/ethnicity status.

The sample size for the pretest consisted of 300 "target" households plus two neighbors for each target. The total of target households and neighbors were 900 units. Overall, we learned from the pretest that neighbor screening is feasible. The FRs who participated in the test liked the option of going to neighbors, and thought it would save time and money. FRs also had suggestions that we have decided to implement such as including screened out houses in reinterview (these cases have become part of our screening QC), and requiring FRs to record the names and addresses of neighbors to deter falsification. (Waite, 1993)

We learned during the pretest that the error rate of using only one neighbor was higher than the error rate when using two neighbors. We found that most neighbors could recognize Black occupants of a target household. The majority of the problem cases were when a neighbor identified the occupants of a target household as being non Black/Hispanic when the household occupants were Hispanic. Based on this, it was decided that during the survey implementation, FRs would be required to visit two neighbors to confirm a household is non Black/Hispanic before screening a unit out of sample.

When neighbors incorrectly screen out a Black/Hispanic sample household, this creates a bias in the sample design. There are four circumstances which must occur to create this situation. First, the sample unit would have to be designated as "S". Second, the FR would have to fail to make contact with the sample unit in the first two attempts (in the

1980 design, FRs made contact with approximately 55% of the sample in the first two attempts). Third, the sample unit would have to be minority (approximately 30% of the sample). Fourth, two neighbors would have to incorrectly identify the sample unit as nonminority. These circumstances together result in less than .2% of minority households being incorrectly screened out. Because this number was so small, and alternative plans too overwhelming on FR burden, we accepted that this bias would occur. (Moore, 1994)

Because we found during the pretest that the error rate increases when only one neighbor is used, we did not test this during survey implementation. This could risk misclassification of the sample unit by neighbors resulting in screening out a minority interview.

Screening Evaluation

We obtained information from the following sources to monitor and evaluate screening:

- Data files from across the country on a weekly basis through a modified System to Automate Regions (STAR) system.
- A small test of eight Field Representatives from each region to compare the costs of a completed interview from 1994 to 1995.
- Field Representative and RO supervisory feedback on how sample household screening and neighbor screening is working.
- A Quality Control (QC) program set up to examine a 5% sample of households screened out of sample ("S" cases only).
- A falsification check to compare the number of cases we expect an FR to screen out with the number of cases the FR is *actually* screening out.

All twelve ROs had a system in place during the 1980 design called the System to Automate Regions (also known as STAR). This was a case management program used to print mailing labels for advance letters, organize Field Representative assignments, etc. Modifications were made to this program to collect information on each case assigned for interview. This is the data we used to evaluate screening. The data was keyed by the ROs and sent to headquarters on a weekly basis.

Over time, we do not expect the percent of Blacks/Hispanics to vary between the cases designated as "S," and the cases designated as "I" within an oversampling stratum. For those strata that contained sufficient data, our results are supporting our expectations because there is no evidence that the percent of Blacks/Hispanics is different between the two groups. (See Table 1)

We were able to compare, based on the data we have collected, the post interview percentages of minority populations against what we expected them to be. Our expectations were based on 1990 census data. Where we have received enough data, the results of our study shows the number of minorities are consistent with what we originally expected. Within each oversampling stratum, we might suspect problems with the screening process if the proportion of minorities were significantly higher for the "I" cases than the "S" cases.

Using Table 1, we were also able to see if we are retaining the correct proportion of non-minorities for our sample. We can do this by looking at the number of cases that are assigned to be interviewed regardless of the minority status of their occupants ("I" cases).

Looking at Stratum #01, we noticed that approximately 78.0% of the cases were being retained. We were expecting this to be approximately 70.0%. Upon investigation of the nonminority retention rates, we discovered an error in the assignment of the screening sampling rates. These are the rates that are used in the ROs to assign cases to be interviewed or screened. (Waite, 1995)

Based on data received from our screening observer evaluation worksheet and FR screening questionnaire, there is a noticeable lack of neighbors to conduct neighbor screening with. In many cases, within a particular neighborhood, people work at the same times during the day. If the occupants of the sample household aren't home, chances are good that neighbors won't be home either. Also, in some segments, every house is in sample. Because FRs can not use current sample units for neighbor screening, they have no neighbors left to visit. Another situation is when the segment is located in a rural area. Because households are so far apart we assume that visiting neighbors is not practical although we have no data to support this.

Of the 4192 screening cases, FRs visited neighbors only 668 times or 15.9% of the time. This figure is instrumental in determining the cost effectiveness of neighbor screening. Obviously, if FRs aren't even going to neighbors, the cost savings achieved through this is minimal. It might be more efficient for FRs to screen only at sample units. This would eliminate the need for additional neighbor screening procedures and the extra neighbor screening questionnaire.

There were 43 times that FRs visited a sample unit based on neighbor response (i.e., a neighbor reported that there was a Black/Hispanic occupant at the sample unit). Of these 43 times, the FR eventually contacted the sample unit 26 times. Neighbors were correct in predicting that these

households were Black/Hispanic 24 times (92.3%). The breakdown of these 24 are the following:

Black	5
Hispanic	19

[note that the remaining 19 were various noninterviews]

It should be noted that the accuracy of neighbor responses is quite high (92.3%). This means that if an FR does visit neighbors, the neighbors are able to report on the race and ethnicity of sample households. So, even though FRs don't go to neighbors too often, neighbors respond accurately when they are used.

This accuracy of neighbor response is not as significant as measuring how often neighbors incorrectly report that there are no Black/Hispanic occupants at the sample unit. However, we did not collect any data that would enable us to evaluate these errors.

We looked at the breakdown of the number of visits to Sample Households for the "S" and "I" cases for interviews and noninterviews. We expected the number of visits for the "S" cases to be small because of neighbor screening. Our data showed that FRs are making three or more visits to the sample household 46.8% of the time for "I" cases, and 37.3% of the time for the "S" cases.

The percentage for the "S" cases appears high, but it might be indicative of the fact that FRs are continuing attempts at sample units because a neighbor has screened it in. The percentage of the number of visits to one or more neighbors probably doesn't accurately display this, because if the first neighbors confirms that the sample household is Black or Hispanic, the FR might just continue attempts to contact the unit without filling out a neighbor questionnaire. We wouldn't have any documentation of that neighbor visit.

We expected our screening evaluation results to show the number of calls to "S" cases to be somewhat less than the number of calls to "I" cases. This is because, hopefully, the FR will be visiting neighbors, and in some cases, screening those "S" cases out of sample. This doesn't apply to the "I" cases because the FR must continue her efforts to contact them. Our data showed that was an average of 2.50 calls to "S" households, while the "I" households required an average of 3.00.

Because of the cost concerns of screening, we wanted a quick study to show us if costs were rising. Therefore, we had each RO choose eight field representatives for participation in this Benchmark Cost analysis.

The preliminary results of this study were not surprising. We found that the total number of hours

per case (for both completed interviews and noninterviews) decreased because of the number of cases being screened-out. The full NHIS interview takes, on average, two to three hours including the supplements while a screened-out case could take an average five to ten minutes. If we notice that the total number of hours per case is increasing, it may be because FRs are spending too much time trying to find suitable neighbors for neighbor screening.

The other variable we were interested in studying was the number of miles per case (once again, "case" refers to both completed interviews in addition to noninterviews). This showed a decrease potentially because of neighbor screening. If an FR is able to screen a unit out with neighbors, she will not have to continue traveling back to that sample household to find someone at home. Therefore, the number of miles traveled for that case would be lower than for those cases that require numerous call attempts. This decrease is also attributed to an increased clustering in sample units for 1995.

The hours per completed interview and miles per completed interview have increased slightly. Because the survey sample was designed to have the same number of completed interviews after screening that we had in 1994, this increase in hours and miles show an increase in cost for 1995.

There are two aspects to the QC plan that we used to monitor screening. The first aspect is the Quality Control (QC) program developed by Census which chooses 5% of the cases selected to be screened, that were screened out (either by the sample household, or by neighbors). This can help detect FRs who are screening out units occupied by Blacks or Hispanics in error. The second aspect is the evaluation of screening rates by FR to determine FR falsification. Using the census minority composition for each strata, we compute expected screen out rates for each FR. Then, using a regression model, we determine what the screen out rate should be. Using a 95% screen-out tolerance limit, we identify those FRs who appear to be screening "out of control." These FRs are placed in a supplemental QC where their next assignment will be monitored. After quarter one of 1995, three FRs were placed in the supplemental QC program.

Summary

As a result of the screening evaluation, we determined that we made an error in assigning the screening sampling rates. Fixing the problem has decreased the number of cases the ROs designate as "I." Also, we have made modifications to the screening exit statement which will aid less

experienced FRs in ending an interview at an "S" household with no Blacks or Hispanics. In addition, three FRs have been placed in supplemental QC where they will be monitored for potential falsification.

The screening evaluation will be an ongoing process. As we receive more data, we will be better equipped to draw more significant conclusions about screening cost effectiveness and implications. Our Benchmark Cost analysis showed, however, that the hours and miles per completed interview have increased slightly for 1995. This will result in increased cost for the survey. Our preliminary results on neighbor screening showed that neighbors appear to be screening sample households effectively. The use of neighbors is limited, however, simply because eligible neighbors are difficult to find.

It is important to realize the limitations of this evaluation plan. The main analysis of results is based on data keyed by the ROs for each case ("case" refers to both interviews and noninterviews). Therefore, the data may not be entirely accurate because of keying errors. Because of these keying errors, the totals reflected in the tables may not always add up. Also, because the neighbor questionnaires were not keyed, we don't have reliable information on neighbors. And, because FRs do not always report a neighbor visit we have no way of knowing the entire number of times a neighbor is visited.

Overall, the conclusion is that further data is needed for a more significant analysis. We plan to continue in our efforts to collect data to more thoroughly investigate the impacts of neighbor screening on costs. During this continued evaluation, it is likely that more issues will arise which will cause us to refine the screening procedures thus helping to achieve the goal of better statistics for minorities.

References

(1991) U.S. Department of Commerce, Bureau of the Census. "Overview of the 1990 Redesign Health Interview Survey Screening Operations." Memorandum for Documentation from HIS Screening Area of Concentration Output Systems Workgroup (WG-9), October 3, 1991.

(1992) U.S. Department of Commerce, Bureau of the Census. "National Health Interview Survey (NHIS) - Summary of Screening Hothouse Tests Results; Proposal for Pretest." Memorandum for James Massy and Preston Jay Waite from Jeffrey Hayes, October 30, 1992.

(1993) U.S. Department of Commerce, Bureau of the Census. "National Health Interview Survey (NHIS) - Summary of Screening Pretest Results."

Memorandum for Owen T. Thornberry and James Massey from Preston Jay Waite, October 26, 1992.

(1994) U.S. Department of Commerce, Bureau of the Census. "NHIS Screening: Bias in the Screening Procedure." Memorandum for Larry Cahoon from Thomas Moore, February 23, 1994.

(1995) U.S. Department of Commerce, Bureau of the Census. "Error in the NHIS Take-Every's." Memorandum for Sherry L. Courtland from Preston Jay Waite, May 9, 1995.

Acknowledgements

Special thanks to Fred Meier for data processing support. Also, thanks to John Bushery and Debbie Fenstermaker for their helpful comments and suggestions.

For further information (or to obtain a copy of the complete paper), please contact the author. This paper reports the general results of research undertaken by the Census Bureau staff. The views expressed are attributable to the author and do not necessarily reflect those of the Census Bureau.

TABLE 1: % Black/Hispanic in each Oversampling Stratum by Screening Status Code (through 5/22/95)

Stratum	Definitions		Screening Status Code 'S'				Screening Status Code 'I'			
	% Black	% Hispanic	# of Cases	Black	Hispanic	Black or Hispanic	# of Cases	Black	Hispanic	Black or Hispanic
01	<10%	<5%	1246	22	25	46	4809	148	92	222
				1.8%	2.0%	3.7%		3.1%	1.9%	4.6%
02	<10%	5-9%	553	17	43	55	528	21	36	54
				3.1%	7.8%	9.9%		4.0%	6.8%	10.2%
03	<10%	10-29%	396	23	53	74	397	17	30	46
				5.8%	13.4%	18.7%		4.3%	7.6%	11.6%
04	<10%	30-59%	364	25	111	127	341	34	101	128
				6.9%	30.5%	34.9%		10.0%	29.6%	37.5%
05	<10%	60+%	330	16	269	271	136	10	106	111
				4.8%	81.5%	82.1%		7.4%	77.9%	81.6%
06	10-29%	<5%	41	13	1	13	409	78	11	87
				31.7%	2.4%	31.7%		19.1%	2.7%	21.3%
07	10-29%	5-9%	72	17	9	25	80	15	5	19
				23.6%	12.5%	34.7%		18.8%	6.3%	23.8%
08	10-29%	10-29%	71	16	8	22	64	8	6	14
				22.5%	11.3%	31.0%		12.5%	9.4%	21.9%
09	10-29%	30-59%	111	28	41	63	100	23	33	54
				25.2%	36.9%	56.8%		23.0%	33.0%	54.0%
10	10-29%	60+%	54	11	41	51	29	4	25	29
				20.4%	75.9%	94.4%		13.8%	86.2%	100.0%
11	30-59%	<5%	19	11	0	11	120	54	0	54
				57.9%	0.0%	57.9%		45.0%	0.0%	45.0%
12	30-59%	5-9%	0	0	0	0	16	3	2	4
				0.0%	0.0%	0.0%		18.8%	12.5%	25.0%
13	30-59%	10-29%	0	0	0	0	16	15	0	15
				0.0%	0.0%	0.0%		93.8%	0.0%	93.8%
14	30-59%	30-59%	62	27	19	44	61	30	17	44
				43.5%	30.6%	71.0%		49.2%	27.9%	72.1%
15	30-59%	60+%	9	8	1	9	8	5	0	5
				88.9%	11.1%	100.0%		62.5%	0.0%	62.5%
16	60+%	<5%	89	78	2	80	364	309	1	310
				87.6%	2.2%	89.9%		84.9%	0.3%	85.2%
17	60+%	5-9%	3	3	0	3	12	12	0	12
				100.0%	0.0%	100.0%		100.0%	0.0%	100.0%
18	60+%	10-29%	7	6	0	6	26	26	1	26
				85.7%	0.0%	85.7%		100.0%	3.8%	100.0%
19	60+%	30-59%	16	13	2	15	16	11	7	16
				81.3%	12.5%	93.8%		68.8%	43.8%	100.0%
20	60+%	60+%	0	0	0	0	0	0	0	0
				0.0%	0.0%	0.0%		0.0%	0.0%	0.0%
21	New Construction		0	0	0	0	349	29	27	54
				0.0%	0.0%	0.0%		8.3%	7.7%	15.5%
Blank			287	73	79	140	405	83	61	134
				25.4%	27.5%	48.8%		20.5%	15.1%	33.1%
Total			3730	407	704	1055	8286	935	561	1438
				10.9%	18.9%	28.3%		11.3%	6.8%	17.4%
Weighted Total				11.1%	11.1%	21.2%		11.1%	9.8%	20.1%