

2016 Sample Redesign of the National Health Interview Survey

Chris Moriarity, Van Parsons

National Center for Health Statistics, 3311 Toledo Road, Hyattsville, MD 20782 USA

Abstract

The National Health Interview Survey (NHIS) has undergone a sample redesign at ~10 year intervals after each decennial census. The current NHIS sample design began in January 2006 and will run through the end of 2015. Planning for the 2016 sample redesign began in 2012 when an interagency group (Census Bureau, National Center for Health Statistics (NCHS)) developed a redesign milestone schedule. A major change, relative to the last three NHIS redesigns, is the sample address source. The current method of field listing all sample areas is no longer affordable for the next NHIS sample design, so NCHS will be using addresses purchased from one or more commercial vendors (supplemented with a limited amount of field listing). We describe the planning activities for the 2016 sample redesign of the NHIS.

Key Words: Sample Survey

1. Introduction

The National Health Interview Survey (NHIS) is the principal source of information on the health of the civilian noninstitutionalized population of the U.S. It is a continuous survey that has been in operation since 1957. The current NHIS sample design was implemented in 2006 and will be in place through the end of 2015. Within the 2006 sample design, NCHS obtains completed interviews at approximately 35,000 living quarters (households and noninstitutional group quarters such as college dormitories) each calendar year if there are no sample reductions or augmentations. All eligible persons residing at a sampled address are covered by the NHIS interview, yielding a sample of approximately 87,500 persons each year if there are no sample reductions or augmentations. Sample sizes can increase or decrease appreciably, according to the availability of funding. Each interview is conducted via a personal visit to the living quarters by an employee of the U.S. Census Bureau, which is the data collection agent for the NHIS.

Additional information about the NHIS is available online at the NHIS home page, <http://www.cdc.gov/nchs/nhis.htm>. The reference section of this paper includes publications that describe NHIS sample designs all the way back to when the survey began in 1957. All NCHS publications that describe the historic NHIS sample designs are available online at:

<http://www.cdc.gov/nchs/nhis/methods.htm>

Since its inception, the NHIS sample has been redesigned following each decennial census of the population to accommodate changes in survey requirements and to take into account changes in the population and its distribution. The next NHIS sample design will be implemented at the beginning of 2016. This paper describes the planning activities for the 2016 sample redesign of the NHIS.

2. Continuing Features of the 2016 NHIS Sample Design

The NHIS has always been a personal interview survey, and will continue to be so. The sample distribution for a personal interview survey usually has some level of geographic clustering, to reduce interviewer travel. This has been, and will continue to be, a feature of the NHIS sample design.

The precision of national-level annual estimates has been and remains a high priority for the NHIS sample design. To meet other geographical estimation objectives, while still maximizing precision for national-level estimates, the NHIS "base" sample (i.e., the sample with no reduction/augmentation) usually has been allocated proportional to population size within each state. A state sample allocation to each state will continue to be a feature of the NHIS sample design. This allocation will support annual state-level estimates with sufficient levels of precision for the most populous states.

3. New Features of the 2016 NHIS Sample Design

There are two major new features of the 2016 NHIS sample design: a change in the sample address source, and increased flexibility. Each is briefly described in turn, and more details appear later.

Sample Frame Change: The NHIS sample that is selected for interviewing is a sample of addresses. The NHIS sample frame of addresses is a list of addresses within the geographic areas selected into sample. NHIS sample designs since 1985 have used field listing to develop most of the sample frame of addresses. This was feasible because NCHS shared the cost of field listing with other federal agencies that sponsor demographic surveys (e.g., The Current Population Survey) conducted by the Census Bureau. This will not be feasible for the 2016 NHIS sample design because the other federal agencies that sponsor demographic surveys conducted by the Census Bureau will no longer use field listing nationwide. In 2016 and beyond, field listing will be only a small part of the sample frame development process. Instead, in most geographic areas, NCHS will be using addresses purchased from one or more commercial vendors.

Flexibility: A main goal for the 2016 NHIS sample design is to have flexibility to increase or decrease annual sample sizes, and/or flexibility to shift annual sample

sizes on a state-by-state level. In order to maintain year-to-year stability, the "base" NHIS sample, which as mentioned above is the anticipated annual sample size if there are no sample reductions or augmentations, will consist of two parts. One part (~70% of the total) will remain the same from year to year, with state-level sample allocation proportional to state population size. The other part (~30% of the total) will be allowed to change from year to year during the sample design period. The 70/30 split provides a large degree of stability, a feature of all previous NHIS sample designs, while allowing for the possibility of change.

4. Chronology of the 2016 NHIS Sample Redesign

The timeline for implementing the 2016 NHIS sample design is much shorter than the timeline was for implementing the current NHIS sample design. Planning for the current NHIS sample design, initially to be implemented in 2005, began in 1998, a span of more than six years. The current NHIS sample design implementation was delayed by a year to 2006, adding another year to implementation time. In contrast, implementation of the 2016 NHIS sample design is occurring over a span of a little over four years.

The NHIS Sample Redesign Committee was created in August 2011. Initially the committee consisted of four NCHS employees. Two of the four original committee members (Moriarity, Parsons) were appointed to be co-managers of the sample redesign project. Subsequently the Sample Redesign Committee membership has expanded to seven members, including the Census Bureau's Survey Director for the National Health Interview Survey. The committee meets once a month to monitor the sample redesign project's progress.

A Census Bureau/NCHS NHIS redesign team was created at the beginning of 2012. The team met frequently between January and September 2012 to develop a sample redesign schedule that includes major tasks and the amount of time needed to accomplish them, with a sample redesign implementation deadline of January 1, 2016. Once the schedule was completed, three Census Bureau/NCHS workgroups were created in the fall of 2012 to begin implementing the tasks in the redesign schedule. One workgroup concluded its work in 2014; the other two workgroups will continue to be active after the sample redesign is implemented on January 1, 2016.

The initial schedule assumed that the United States Postal Service (USPS) Delivery Sequence File (DSF) would be the NHIS sample address source. This would be within the current scope of Section 412 of Title 39 USC, which states in part: "The Postal Service shall provide . . . address information . . . as may be determined ... to be appropriate for any census or survey being conducted by the Bureau of the Census". As a contingency, the redesign team prepared a modified schedule in the fall of 2012 with task changes/new tasks that are needed if access to the DSF for use as the NHIS sample address source was not granted by the end of 2012.

An interagency agreement between the Census Bureau and the USPS that provided access to the DSF for use as the NHIS sample address source did not occur by the end of 2012, so NCHS decided at the beginning of 2013 to use a commercially available address list as the main NHIS sample address source. Following the competitive process, a contract was signed by the Census Bureau in 2013 with Marketing Systems Group (MSG) for the use of MSG's Address Based Sampling database. Henceforth this resource is referred to as "the MSG File". The contract terms included MSG providing the Census Bureau with periodic updates of the file.

One Census Bureau/NCHS redesign workgroup ("Workgroup 4.2") conducted research on the MSG File. The most important goal of this research was to determine the NHIS sample areas where the MSG File was suitable for use as the NHIS sample address frame; the NHIS sample address frame for the remaining NHIS sample areas will be created via field listing.

As is the case with the USPS DSF and with other commercially available address lists, the primary use of the MSG File is for mailing. In some rural areas of the U.S., addresses are not in house number/street name format, e.g., 123 Main St.; instead, they can be in rural route number/box number format, e.g., Rural Route 2, Box 127. The USPS knows how to deliver mail to Rural Route 2, Box 127, but it not feasible to send an NHIS interviewer to this location. NHIS sample areas without house number/street name addresses are designated for field listing.

Census Bureau staff in Workgroup 4.2 were able to conduct a comparison of the MSG File with the Census Bureau's Master Address File (MAF) and provide summary-level information about the comparison results to the full workgroup. The results of this comparison provided information about the coverage of the MSG File, treating the MAF as the benchmark. The coverage information informed the redesign workgroup's decision in 2014 of where to use the MSG File as the NHIS sample frame, which in turn determined where field listing would be used to create the remainder of the NHIS sample frame. The decision was a combination of achieving high anticipated coverage rates (85% or greater) while keeping the field listing workload within budget.

Workgroup 4.2 researched various "filtering" options to select a high proportion of MSG addresses considered to be valid residential addresses by the MAF, while simultaneously minimizing the number of MSG addresses considered to be invalid by the MAF, using the information available in the MSG File. The workgroup identified a filtering algorithm that achieved a good combination of the two criteria. This is the filtering algorithm that will be used.

Additional research by Census Bureau staff in Workgroup 4.2 determined that it was worthwhile for the Census Bureau to perform an independent assignment of geographic codes to the addresses in the MSG File. The research findings

indicated a high level of agreement between MAF geographic codes and MSG File geographic codes, except at the census block level. Accurate geographic codes down to the census block level help to guide interviewers to the correct area to locate a given sample address.

Another Census Bureau/NCHS redesign workgroup ("Workgroup 3.10") focused on selection of NHIS primary sample areas and developing the procedures for the stages of subsampling within the primary areas.

The first step was to select a sample of NHIS primary sample areas, referred to henceforth as primary sampling units (PSU). Within each state, a set of PSUs was created. Each PSU consisted of one or more adjacent counties. Counties with small populations were combined with adjacent counties. Counties in the same Metropolitan Statistical Area (MSA), using February 2013 MSA definitions, were grouped together into a PSU. The District of Columbia is a state equivalent, consisting of a single PSU. Entities such as parishes in Louisiana, independent cities in Maryland, Missouri, Nevada, and Virginia, and boroughs and census areas in Alaska are treated as county equivalents.

NCHS members of Workgroup 3.10, in consultation with the Census Bureau, completed the sample selection of PSUs in 2013. Both a core and reserve set of sample PSUs have been chosen in a way to provide flexibility for sample expansion if funding is provided. In less populous states, a sample of reserve PSUs has been identified for use in the event that sample expansion were to occur in these states. If sample expansion does not occur, the core sample of PSUs is identified for use. In more populous states, if sample expansion were to occur, additional sample areas will be used within existing PSUs with reserve PSUs available if needed. With this approach, sample expansion can occur primarily in less populous states if desired. This is consistent with the sample expansions that occurred in 2011-2015, which yielded a larger number of state-level estimates.

In the absence of funding for sample expansion, the number of PSUs in less populous states is limited by the state-level sample allocation and the PSU-level minimum interviewer workload requirement of ~100 addresses annually. More populous states have sufficient core sample PSUs for stable state-level estimates. Thus, modest sample expansions in more populous states can be implemented by sample expansion in existing PSUs, while it is advantageous in less populous states for sample expansions to be implemented using new PSUs if the expansion amounts are sufficiently large.

A third redesign workgroup ("Workgroup 5.4"), consisting only of Census Bureau employees, is responsible for monitoring the progress of the redesign work, and assessing if the redesign project is on schedule. This workgroup also is responsible for developing the necessary procedures to process the MSG File to create the set of sample addresses in the sample areas where the MSG File is being used as the sample address source.

The combination of the MSG File and a limited amount of field listing will provide most of the NHIS sample. One additional resource has been created to assure an adequate sample from one portion of the U.S. eligible population: college students living full-time in residence halls (also known as "dormitories"). College students living full-time in residence halls are the largest subgroup of persons who live in non-institutional group quarters. Recent research conducted by the Census Bureau suggested that commercially available address lists may not provide an acceptable level of coverage of college students living full-time in residence halls. Thus, a separate sample frame, based on the Integrated Post-Secondary Education Data System, is being created to sample college students living full-time in residence halls in NHIS PSUs.

5. Allocation of the Year 2016 NHIS Base Sample

Given that the year 2016 is the first year of the 2016 sample design, NCHS started the planning process for the allocation of the flexible part of the year 2016 NHIS base sample more than a year in advance, in the fall of 2014.

NCHS staff conducted extensive research on various sample allocation scenarios, and identified 4 allocations for illustrative purposes. Allocation 1 is best for the precision of national-level estimates. Allocation 2 preserves most of the precision of national-level estimates while allowing for three-year estimates for all states and the District of Columbia (DC). Allocation 3 allows two-year estimates for all states and DC; the price is a considerable loss in national-level precision and minority sample size. Allocation 4 attempts to maximize the number of one-year state estimates; again, the price is a considerable loss in national-level precision and minority sample size. All four allocations incorporate the stable annual core of ~25,000 household interviews (~70% of the NHIS base sample), allocated proportional to state population size. Details of the four allocations follow.

Allocation 1: Allocate the entire flexible part of the base sample proportional to state population size. This is similar to historic NHIS allocations, including the current one. This allocation yields the most precise national-level estimates. Nominal annual sample sizes range from ~3600 completed household interviews in California to ~75 in Wyoming. If implemented, this allocation would be expected to yield annual state-level estimates with suitable precision for ~21 states; two-year estimates for ~34 states; three-year estimates for ~40 states.

Allocation 2: Relative to allocation 1, this allocation of the flexible part of the base sample tends to slightly reduce allocations in some of the more populous states, and increase allocations appreciably in the 10 least populous states and DC. This allocation leads to relatively small loss in the precision of national-level estimates (~1.3%) and minority group sample sizes (~1%), relative to allocation 1. Nominal annual sample sizes range from ~3450 completed household interviews in California to ~250 for the 10 least populous states and DC. If

implemented, this allocation would be expected to yield annual state-level estimates with suitable precision for ~21 states; two-year estimates for ~34 states; three-year estimates for all states and DC.

Allocation 3: Relative to allocation 1, this allocation of the flexible part of the base sample reduced allocations in the 25 most populous states and increased allocations in the 25 less populous states and DC. Relative to allocation 1, this allocation results in noticeable loss in the precision of national-level estimates (~13%) and minority group sample sizes (~11%). Nominal annual sample sizes range from ~2600 completed household interviews in California to ~525 for the 25 least populous states and DC. If implemented, this allocation would be expected to yield annual state-level estimates with suitable precision for ~14 states; two-year estimates for all states and DC.

Allocation 4: Relative to allocation 1, this allocation of the flexible part of the base sample reduced allocations in the 11 most populous states and increased allocations in the next 30 states (ranked by population size) to obtain ~650 completed household interviews in those 30 states. Relative to allocation 1, this allocation results in sizable loss in the precision of national-level estimates (~7.6%) and minority group sample sizes (~12%). Nominal annual sample sizes range from ~2600 completed household interviews in California to ~100 for the nine least populous states and DC. If implemented, this allocation would be expected to yield annual state-level estimates with suitable precision for ~41 states; to obtain state-level estimates for the nine least populous states and DC, at least five years of data would have to be aggregated.

These four options were presented to the NCHS Director in October 2014. The Director chose Allocation 2. Planning for implementing the year 2016 sample allocation then proceeded with the decision milestone reached.

6. Future Steps

Several important tasks remain to be done in late 2015 or early 2016, such as defining the panel structure that partitions the base NHIS sample into four essentially identical subsamples, and defining variance estimation structures for 2016 and beyond.

The sample design project does not end on January 1, 2016. Monitoring and evaluation will be ongoing as the sample design is implemented in 2016. Changes can be expected, based on what is learned from the monitoring/evaluation. For example, there likely will be areas where the sample address source is changed from the MSG File to field listing, and vice versa. Preliminary planning has begun for several monitoring/evaluation projects. For example, Census Bureau staff will continue several aspects of the earlier work by Workgroup 4.2.

7. Conclusion

Compared to the previous sample redesign process, the 2016 NHIS sample redesign process has moved at a faster pace due to the amount of change being implemented and a smaller amount of time available for implementation. Implementation will occur on schedule in January 2016.

The NHIS sample address source change for 2016 and beyond is a major change that will require ongoing monitoring and evaluation.

More details about the 2016 NHIS sample redesign will appear in a future NCHS publication, similar to the recent entries in the reference list below.

References

- Botman S, Moore T, Moriarity C, and Parsons V. Design and Estimation for the National Health Interview Survey, 1995-2004. *Vital Health Stat* 2(130). 2000.
- Kovar M, Poe G. The National Health Interview Survey design, 1973-1984, and procedures, 1975-83. *Vital Health Stat* 1(18). 1985.
- Massey J, Moore T, Parsons V, Tadros W. Design and estimation for the National Health Interview Survey, 1985-94. *Vital Health Stat* 2(110). 1989.
- National Center for Health Statistics. The statistical design of the Health Household-Interview Survey. Health Statistics. PHS Pub. No. 584-A2. Public Health Service. Washington: U.S. Government Printing Office. 1958.
- National Center for Health Statistics. Health Interview Survey procedure, 1957-1974. National Center for Health Statistics. *Vital Health Stat* 1(11). 1975.
- Parsons V, Moriarity C, Jonas K, Moore T, Davis K, and Tompkins L. Design and Estimation for the National Health Interview Survey, 2006-2015. *Vital Health Stat* 2(165). 2014.