

## **Overview of the 2016 National Health Interview Survey Sample Redesign**

### **Introduction**

The National Health Interview Survey (NHIS) is the nation's primary source of general health information for the resident civilian noninstitutionalized population. NHIS is sponsored by the Centers for Disease Control and Prevention's National Center for Health Statistics (NCHS). In accordance with NCHS specifications, the U.S. Census Bureau, under a contractual agreement, participates in the planning and collection of data for NHIS. NHIS has continuously collected data since July 1957.

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 current NHIS sample design has been in place since 2006. The next NHIS sample design will be implemented at the beginning of 2016.

### **Continuing Features of the 2016 NHIS Sample Design**

The NHIS has been a personal interview survey since its inception, and will continue to be so. The sample design 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 the highest priority for the NHIS sample design. To meet secondary geographical estimation objectives, while still maximizing precision for national-level estimates, the NHIS core sample (35,000 households in a typical year) has been allocated proportional to population size within each State. A state sample allocation will continue to be a feature of the NHIS sample design.

### **New Features of the 2016 NHIS Sample Design**

Sample Frame Change: Since 1985 the NHIS sample designs have used a field listing to develop the sample frame of addresses. For previous designs 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.

Flexibility: A main goal for the 2016 NHIS sample design is to have flexibility to increase or decrease annual sample sizes, and/or to have flexibility to shift annual sample sizes on a state-by-state basis. The core NHIS sample will consist of two parts: a stable part (~70% of the total) that 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, in accordance with decisions made by the NCHS Director at least one year in advance. Reserve state samples for increasing state-level precision will also be

available if additional funding is provided.

College Housing Frame: To ensure an adequate sample from college students living full-time in residence halls, a separate sample frame has been created to sample this population in the primary sampling units (PSUs) included in the NHIS. This is necessary because college students living full-time in residence halls are the largest subgroup of persons who live in non-institutional group quarters, yet recent research conducted by the Census Bureau suggests that commercially-available address lists may not provide an acceptable level of coverage of this population.

## **Chronology of the 2016 NHIS Sample Redesign**

Full-scale planning for the 2016 NHIS sample design began in 2011. A high-level schedule and task list was created in 2012. Three workgroups were created in 2012 to begin implementing the tasks. One workgroup concluded its work in 2014; the other two workgroups will be active through 2015 and after the sample redesign is implemented on January 1, 2016.

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 were appointed to be co-managers of the sample redesign project. Subsequently the Sample Redesign Committee membership has expanded to six members, including the Census Bureau's Survey Director for the National Health Interview Survey. The committee meets at least 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 high-level 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. Three Census Bureau/NCHS workgroups were created in late 2012 to begin implementing the tasks in the redesign schedule.

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 was not guaranteed by the end of 2012 via an interagency agreement between the Census Bureau and the USPS.

An interagency agreement between the Census Bureau and the USPS did not occur by the end of 2012, so NCHS decided at the beginning of 2013 to use a commercially-available list as the NHIS sample address source. Following the competitive process, a contract was signed 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".

One Census Bureau/NCHS redesign workgroup ("Workgroup 4.2") conducted research on the

MSG File. The 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 would be created via field listing.

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

The Census Bureau was able to conduct a comparison of the MSG File with the Census Bureau's Master Address File (MAF). This comparison provided information about the coverage of the MSG File, treating the MAF as the benchmark. The coverage assessment informed the redesign workgroup's decision in 2014 of where to use the MSG File as the NHIS sample frame and to identify areas that might require listing.

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 and independent cities in Maryland, Nevada, and Virginia are treated as county equivalents.

NCHS, in consultation with the Census Bureau, selected a core set of PSUs (which will always be in sample) to enable the production of quality national estimates and annual state-level estimates in a limited number of states, and a much larger set of reserve PSUs to be easily available for sample expansions to improve the quality of national estimates and enable state estimates on an annual basis in as many states as funding will allow.

The initial selection of sample PSUs was completed in 2013, with adjustments between the core and reserve PSU samples taking place in late 2014 to ensure that state-level estimates in all states would be possible over a three-year period using only the core PSUs. In more populous states, where numerous PSUs are already in the core sample, sample expansions would take place primarily by increasing sample in core PSUs, with reserve PSUs available if needed. This approach is designed to permit sample expansion to occur primarily in less populous states, consistent with the sample expansions that have occurred in 2011-2015.

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 information reseller 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 (IPEDS), has been created to sample college students living full-time in residence halls in NHIS PSUs. Although this project has been reviewed and approved under separate cover (OMB# 0920-0222), for purposes of completeness a brief overview is provided below.

IPEDS is a free publicly-available source of college housing information that collects information annually from every college, university, and technical and vocational institution that participates in federal student financial aid programs (Knapp et al, 2012). In order for colleges/universities to participate in the Free Application for Federal Student Aid (FAFSA) program, they must provide the information collected by IPEDS, making this a reliable source of information. IPEDS data contains college name, location, telephone number, and number of students housed on campus – all information vital to building the College Housing Frame. Because IPEDS is conducted annually with results released publicly the following year, it provides accurate and timely information for constructing the College Housing Frame.

Colleges offering student housing in sample PSUs are identified from the IPEDS website and information about each college is collected, including contact information for the school and the total possible number of students able to live in on-campus housing provided by the school. The total on-campus population for each school is converted into measures, with each measure representing two housing unit equivalents. Measures within a PSU are grouped together and sampled. Each of the measures sampled within each PSU is associated with a college (multiple measures could reside in the same college); this is how, ultimately, the colleges that fall into the sample are identified. Within each state, the sampling rate is adjusted to the college population in the PSUs in sample as if the college dorm frame in the state was a cluster of non-self-representing (NSR) PSUs, and the restriction of the college dorm frame to the PSUs in sample was the NSR PSU chosen from that cluster.

As the IPEDS website does not provide specific college dormitory information but rather the total housing the school offers, all colleges selected during Within-PSU sampling are then called by the Census Bureau's National Processing Center (NPC) to collect information about the dormitories located on campus. This information will include the name, number of units, address information, and contact information for each dormitory. In addition, the telephone lister asks whether the dormitory contains a register or not (an up-to-date list of all rooms/beds in the dormitory); this is used to determine the final units in the next stage of sampling.

It is possible that a college dormitory could be selected for sample in both the commercial address frame and the college housing frame. To ensure that units only receive one chance of selection, a sample unit in the commercial address frame that is found to be a college dormitory

is assigned a Type C non-interview because that unit would have been given a chance of selection in the College Housing Frame.