

Current and Future Uses of Address Lists for the National Health Interview Survey Sample

Chris Moriarity

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

Abstract

The National Health Interview Survey (NHIS) is a continuous survey sponsored by the National Center for Health Statistics (NCHS) that has collected health data using personal interviews since 1957. The Census Bureau has been the NHIS field collection agent since the inception of the NHIS. Beginning with the 1985 NHIS, Census Bureau personnel have created address lists via canvassing sample areas (field listing), and these lists were then used to select sample addresses for inclusion in the survey.

The current NHIS address list creation procedure is expensive, and it will become even more expensive within a few years as other federal agencies currently sharing field canvass costs with NCHS move to using the Census Bureau's Master Address File (MAF) as their primary sample address source. The MAF is not an option for the NHIS, because the Census Bureau cannot release MAF addresses to NCHS. NCHS moved to field canvassing in 1985 so the NHIS sample addresses could be released to NCHS for followback surveys, record linkage to administrative data, etc.

NHIS began using an address list in the 2013 NHIS as part of a sample augmentation in selected states. NHIS also will transition to using one or more address lists as the primary sample address source for the next NHIS sample design period beginning in 2016. I discuss our planning processes for current and future uses of address lists for the NHIS sample.

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. NCHS anticipates obtaining 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 persons at a sampled address are included in 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.

The NHIS sample consists of clusters of living quarters (addresses) chosen within a first-stage sample of U.S. counties. The NHIS sample design is complex, with multiple stages of sampling. The first stage of sampling is selection of primary sampling units (PSU), which are one or more contiguous U.S. counties. The PSUs are assigned to sampling strata, and then one or two PSUs are selected from each sampling stratum. Once the sample PSUs are selected for a given sample design period, they delineate the geographic areas where the NHIS sample addresses will be selected for the entire sample design period. There are additional stages of sampling within each PSU, yielding a sample of addresses after the final stage of sampling. This sampling method is used to control the costs related to personal visit interviewing. The cost of conducting personal visit interviews in a simple random sample of U.S. living quarters would be prohibitive, due to the amount of travel that would be required.

The current sample design of the NHIS is based on Census 2000 information. The current sample design is very similar to the previous sample design, which was in effect from 1995 to 2005 and was based on 1990 Census

information. The NHIS has undergone changes in the sample design at intervals of approximately ten-year duration, using information from the previous decennial census.

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

NCHS started receiving supplemental funding in 2010 to expand the NHIS sample in selected states, beginning with the 2011 NHIS. The current NHIS sample design period (2006–2015) has a state-level stratification, but only a limited number of states normally have adequate sample to support reliable estimation. The 2011 and 2012 NHIS samples were augmented by expanding the number of sample addresses in existing PSUs. The two initial augmentation sources were addresses eliminated from the NHIS sample in previous years due to budget shortfalls, and addresses assigned for interview beyond 2015 ("reserve sample"). Beginning in 2012, addresses initially subsampled out during a within-PSU stage of sampling also were used as an augmentation source. By the end of 2012, both of the two initial augmentation sources had been used up.

For the 2013 NHIS, we began augmenting the NHIS using sample addresses in new sample PSUs, in addition to continuing the augmentation using addresses in existing PSUs that had been subsampled out originally that was begun in 2012. Some of the sample addresses in the new sample PSUs came from a commercial address list. Additionally for the 2013 NHIS, we began to augment the NHIS sample using sample addresses from a commercial address list in selected areas where the entire area already was included in the NHIS sample.

The next sample design of the NHIS will be implemented at the beginning of 2016. A major change is that NHIS will transition to using one or more commercial address lists as the primary sample address source for the next NHIS sample at that time.

I describe the current and planned uses of commercial address lists in the NHIS below.

2013 NHIS sample augmentation using a commercial address list

The sample augmentation that began with the 2011 NHIS was in 32 states and the District of Columbia. No sample augmentation occurred in the 18 most populous states because their typical annual sample sizes were considered to be adequate for precise state-level estimates. As previously mentioned, the 2011 and 2012 NHIS samples were augmented by expanding the number of sample addresses in existing PSUs. The two initial augmentation sources were exhausted by the end of 2012.

For the 2013 NHIS, we wanted to consider the additional option of augmenting the NHIS using sample addresses in new sample PSUs. We first needed to determine if such an option was feasible. The research we conducted is summarized in Moriarity and Parsons (2013). Once we determined that we could expand the NHIS into new sample PSUs, we needed to determine how to obtain sample addresses in the new PSUs.

One possibility for obtaining sample addresses in the new PSUs was field listing, the same method used to obtain most of the current NHIS sample addresses. A barrier to implementing this possibility is that the Census Bureau had converted many of their survey support systems from Census 2000 geography to 2010 Census geography. The current NHIS sample design is based on Census 2000. It was not clear that we could develop sampling methods in the new PSUs that were consistent with the current NHIS sampling methods while coping with the 2010 Census-based survey support system.

The other possibility for obtaining sample addresses in the new PSUs was using addresses from a commercial address list that included 2010 Census geography. This possibility allowed us to develop consistent within-PSU sampling procedures in the new PSUs. However, we lacked experience with using this type of resource. Thus, we did not have advance knowledge of the quality of the geocoding nor the accuracy of the list (i.e., high coverage of extant residential addresses, low incidence of invalid addresses and duplicates).

The Census Bureau acquired a commercial address list for the areas included in the new sample PSUs and the selected areas already in the NHIS sample that were to receive augmentation from this initiative. Census Bureau staff then selected a sample of approximately one thousand 2010 Census blocks. They conducted an evaluation where they compared the address count from the commercial address list at the block level to 2010 Census figures at that same level. NCHS and the Census Bureau then agreed on a protocol for selecting a ~50% subsample of the initial sample of ~1000 blocks to be sent to field listing, where most of the subsample was to come from blocks that had address counts that varied considerably from 2010 Census counts. The field listings were used to select the address sample in the sample blocks where field listing occurred; the commercial address list was used to select the address sample in the remaining sample blocks.

We plan to conduct an evaluation that compares the commercial address list to the field listing in the subsample when resources are available. The evaluation will start to help us assess the quality of the commercial address list in the areas where we used it for the 2013 NHIS sample augmentation. We also plan to compare the survey administration results for areas where we used addresses from the commercial address list to the survey administration results for other areas. For example, we plan to determine whether there are important differences in the percentage of addresses that are out of scope in the two types of areas.

2016 NHIS sample design: new sample address source

The NHIS sample designs for 1985-1994, 1995-2005, and 2006-2015 have obtained sample addresses primarily through field listing. Following PSU sampling and within-PSU sampling to select relatively compact sample areas, Census Bureau personnel have created address lists via canvassing these areas. The lists have been used to select NHIS sample addresses.

NCHS moved to field listing in 1985 so the Census Bureau could share NHIS sample address information with NCHS. Prior to 1985, NHIS sample addresses came primarily from decennial census lists, which are confidential and cannot be shared. NCHS has used the sample address information for followback surveys and record linkage to administrative records. Beginning in 1996, the Medical Expenditure Panel Survey, sponsored by the Agency for Healthcare Research and Quality (AHRQ), has been based on a reuse of a portion of the previous year's NHIS sample address records, under an interagency agreement between NCHS and AHRQ.

Field listing requires significant resources, so it is costly. The cost has been shared with other federal agencies who use the Census Bureau as their field collection agent for demographic surveys, e.g., the Current Population Survey, which is sponsored by the Bureau of Labor Statistics. In the past the other demographic surveys conducted by the Census Bureau have required some level of field listing, primarily in rural areas and areas experiencing rapid expansion or contraction, relative to the previous decennial census.

The other federal agencies who use the Census Bureau as their field collection agent for demographic surveys have agreed to greatly reduce field listing for their post-2010 Census sample designs, relying mostly on the Census Bureau's Master Address File (MAF) as their sample address source. The MAF addresses are confidential; they cannot be shared. NCHS has decided to continue with a sample address creation method for the 2016 NHIS sample design whereby the Census Bureau can continue to share sample addresses with NCHS. Thus, relying mostly on the MAF is not an option for the NHIS. To avoid incurring much higher field listing costs starting with the 2016 NHIS sample design, NCHS must use a new address source. NCHS decided to use one or more commercial address lists as the primary sample address source for the 2016 NHIS sample design.

The decision to use a new sample address source is a major change for the 2016 NHIS sample design. A number of issues need to be considered, such as:

1. Accuracy of the source.
2. Geocoding quality.
3. Coverage of persons who are eligible for inclusion in the NHIS, such as college students residing full-time in college dormitories while attending school.

4. How to update the source over the sample design period.

5. How to select annual samples that do not overlap with addresses previously selected during the sample design period.

Commercial address lists are used primarily for contact by mail. The list addresses are a mix of house number/street name, rural route/highway contract route, post office box, etc. Addresses that are not house number/street name are not suitable for a personal visit interview survey such as the NHIS. Thus, some level of field listing will still be required in the 2016 NHIS sample design.

We are aware that situations exist that require special procedures, such as a large apartment building with a single mail delivery point. We assume that we will have to do field listing at least part of the time when a building of this type is included in the NHIS sample.

Research is underway to prepare for the use of one or more commercial address lists for the 2016 NHIS sample design. The evaluations we are planning for the 2013 augmentation will provide some useful information, but many questions will remain. The 2013 augmentation was done in selected areas, while the 2016 NHIS sample design will be a sample of the entire U.S. The 2013 augmentation implementation did not have to develop methodology for issues such as updates and selection of multiple nonoverlapping annual samples. The 2016 NHIS sample design will have to develop solutions for both of these issues.

Summary

The NHIS has just begun using commercial address lists. We have started to acquire some experience with working with this type of resource. Looking to the future, we face many questions about the use of commercial address lists that still require answers.

Using one or more commercial address lists for the 2016 NHIS sample design address source is a major change. We have limited time to develop methodologies to use the new sample address source while simultaneously assessing the source's quality. We know that some field listing will be required, but we are not in a position yet to assess the required level. Cost constraints likely will limit the amount of field listing we do primarily for quality assessment purposes.

Part of our strategy is not to acquire just a sample of addresses, but to acquire the entire address list in designated areas. The initial cost of acquiring the entire list is higher than purchasing just a sample. Having the entire list gives us more flexibility to implement and alter our sampling methods as needed, and we are able to do more comparisons to external benchmarks as part of our assessment of the list's quality.

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 MG, Poe GS. The National Health Interview Survey design, 1973-1984, and procedures, 1975-83. *Vital Health Stat* 1(18). 1985.

Massey JT, Moore TF, Parsons VL, Tadros W. Design and estimation for the National Health Interview Survey, 1985-94. *Vital Health Stat* 2(110). 1989.

Moriarity C, Parsons VL. Expanding the Number of Primary Sampling Units for the National Health Interview Survey. 2013 Proceedings of the Joint Statistical Meetings, 1911-1916.

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.