The National Health Interview Survey: Sample Design, Sample Augmentation, Sample Redesign Plans

Chris Moriarity, NCHS

Symposium of the U.S. Statistical Agencies, November 13, 2013

cdm7@cdc.gov
Presentation outline

National Health Interview Survey (NHIS) sample design features

NHIS sample augmentation, beginning in 2011

A big change for the next NHIS sample design cycle, to be implemented in January 2016
The purpose of NHIS

To collect data on the health status and health services utilization of the U.S. civilian non-institutionalized population

To collect data on specific topics of current public health concern

To provide official national health data to the public
NHIS target population

Persons living in the 50 States and D.C. (not Puerto Rico, etc.)

Civilians (not active military)

Persons who live in households or non-institutional group quarters such as college dormitories (not persons living in institutional group quarters such as prisons)
NHIS interview protocol

Personal visit interviewing at sample addresses, using electronic questionnaire

Information is collected about all eligible persons of all ages at the sampled address

One adult and one child (if children present) randomly selected per family for more extensive questions
Analysis of NHIS data

Account for NHIS sample design - weights, variance estimation

1995-2005 NHIS sample design - Series 2 report #130 - available online at www.cdc.gov/nchs/nhis/methods.htm

2006-15 NHIS design similar but not identical to 1995-2005 design (Series 2 report forthcoming)
Sample design features

A "complex" design - a random sample, but not a simple random sample

Multiple stages of sampling, beginning with geographic areas (counties or grouped counties)

Sample universe partitioned into sampling strata and substrata
Sampling steps

1. Define Primary Sampling Units (PSU): counties, or groups of adjacent counties

2. Group PSUs into sampling strata within each State

3. Within each sampling stratum, select a sample of PSUs to represent all PSUs in the stratum
4. Substratify census blocks within PSUs by concentrations of several race/ethnicity groups

5. Within the PSU substrata, partition the census blocks into chunks - each is to provide a decade of annual samples - then select a sample of chunks
NHIS sampling frame

Two parts:

"Old construction" - Based on information from Decennial Census prior to beginning of sample design period (e.g., 2000 Census for current 2006-15 NHIS sample design)

"New construction" - Sample of building permits in the PSUs already selected for the "old construction" sample
NHIS - not a simple random sample

The 87,500 persons and 35,000 households are geographically concentrated in sample counties.

The proportions of black, Hispanic, and Asian persons in the NHIS are higher than in the general population (oversamples) - Asian oversample a new design feature.
NHIS oversampling methods

Different sampling rates in the PSU substrata

Screening: NHIS sample split into two parts prior to interviewing - in one part, household is "screened out" if it does not contain any eligible black, Hispanic, or Asian persons
Software for NHIS variance estimation

Reference: excellent Web page maintained by Alan Zaslavsky

http://www.hcp.med.harvard.edu/statistics/survey-soft/

Software list, comparative summaries, review articles
<table>
<thead>
<tr>
<th>Software Package</th>
<th>License</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AM Software</td>
<td>free</td>
<td>American Inst. for Research</td>
</tr>
<tr>
<td>Bascula</td>
<td>$</td>
<td>Statistics Netherlands</td>
</tr>
<tr>
<td>CENVAR</td>
<td>free</td>
<td>U.S. Census Bureau</td>
</tr>
<tr>
<td>CLUSTERS</td>
<td>free</td>
<td>University of Essex</td>
</tr>
<tr>
<td>Epi Info</td>
<td>free</td>
<td>Centers for Disease Control</td>
</tr>
<tr>
<td>GES</td>
<td>$</td>
<td>Statistics Canada</td>
</tr>
<tr>
<td>IVEware</td>
<td>free</td>
<td>University of Michigan</td>
</tr>
<tr>
<td>PCCARP</td>
<td>$</td>
<td>Iowa State University</td>
</tr>
<tr>
<td>R survey</td>
<td>free</td>
<td><a href="http://www.r-project.org">www.r-project.org</a></td>
</tr>
<tr>
<td>SAS/STAT</td>
<td>$</td>
<td>SAS Institute</td>
</tr>
<tr>
<td>SPSS</td>
<td>$</td>
<td>SPSS</td>
</tr>
<tr>
<td>Stata</td>
<td>$</td>
<td>Stata Corporation</td>
</tr>
<tr>
<td>SUDAAN</td>
<td>$</td>
<td>Research Triangle Institute</td>
</tr>
<tr>
<td>VPLX</td>
<td>free</td>
<td>U.S. Census Bureau</td>
</tr>
<tr>
<td>WesVar</td>
<td>free</td>
<td>Westat, Inc.</td>
</tr>
</tbody>
</table>
Variance estimation guidance at NHIS methods page - 1963 to 2012

SUDAAN, Stata, R survey, SAS/STAT survey procedures, SPSS, VPLX: Sample code provided for use with NHIS data

SAS/STAT, SPSS: Guidance provided to avoid problems with missing DOMAIN/SUBPOP variables in analyses of NHIS data
Summary of new features in 2006-15 NHIS design

Asian-Americans oversampled

Slight drop in sample size (flat budget assumption during sample redesign planning stage)

Sample adult selection: black, Hispanic, Asian persons aged 65+ have increased selection chance
2011: NHIS began to receive supplemental funding

Supplemental funding provided for new NHIS questions, and for increasing NHIS sample

Focus was to enhance NHIS's ability to assess the effect of the Patient Protection and Affordable Care Act, enacted in 2010
NHIS sample augmentation

Augmentation in 32 states and the District of Columbia

Purpose of targeted augmentation was to increase sample sizes in less populous states

2011: 13% sample increase

2012: 21% sample increase
2006-2015 NHIS survey design did not plan for augmentation

Plans were made for sample cuts in case of budget shortfalls, but not for sample increases

Initial augmentation sources: sample cut in previous years, reserve sample assigned for years beyond 2015
Another augmentation source beginning in 2012: "subsampled-out sample"

Census Bureau drew a large initial sample for the current NHIS sample design, then did multiple stages of subsampling.

We obtained some new sample by bringing back some areas subsampled out at the first stage.
We needed to find a new augmentation source for 2013 and beyond

Initial augmentation sources (sample cut in previous years, reserve sample assigned for years beyond 2015) exhausted by the end of 2012

Subsampled-out sample source not enough by itself to meet augmentation targets
Potential new augmentation source: new PSUs

All previous augmentation occurred in existing PSUs

Some less populous states had a small number of PSUs

Some strata only had one sample PSU selected initially
2013 NHIS sample augmentation using new PSUs

Research that allowed us to expand to new PSUs: see Moriarity and Parsons, 2013 Joint Statistical Meetings proceedings (forthcoming)

We needed to select sample of addresses in the new PSUs
Sample address source within new PSUs: commercial address list

Changes in Census Bureau survey support systems to 2010 Census geography ruled out traditional NHIS field listing method

Acquired complete commercial address list for areas in new PSUs

Selected a sample of ~1000 blocks
Quality of commercial address list was unknown

We compared address counts from the commercial list by block to 2010 Census counts

We selected a ~50% subsample of initial sample for field listing

Future evaluation planned
Future assessment of commercial address list quality

Evaluation in blocks with both address list and field listing

Comparison of survey outcome in address list areas versus nearby field listing areas: percentage out of scope (non-residential, uninhabitable), etc.
NHIS sample address source, 1985-2015

Sample addresses for the 1985-94, 1995-2005, and current NHIS sample designs have come primarily from field listing.

The Census Bureau can share NHIS sample addresses with NCHS.

NCHS uses the addresses for other surveys, record linkage, etc.
2016 NHIS sample design: change in sample address source

Other demographic surveys conducted by the Census Bureau moving toward using the Master Address File (MAF) as the main source of sample addresses

NCHS has been sharing costs for NHIS field listing; NCHS cannot afford the full cost of listing
2016 NHIS sample design address source: address list(s)

NCHS does not want to use MAF addresses for NHIS because MAF addresses are confidential, they cannot be shared

NHIS will use one or more commercial address lists as main sample address source, beginning in 2016
Issues to address when using a commercial address list

Accuracy, geocoding quality, coverage of areas like college dormitories, etc.

How to do periodic updates

How to select annual samples that do not overlap with previous ones
2016 NHIS sample design will still include some field listing

Rural route/highway contract route/P.O. Box addresses not suitable for personal visit interviewing

Situations like a large apartment buildings with a single mail delivery point: listing likely needed at least part of the time
Summary

The NHIS sample design is complex, with unequal probability sampling, clustering, etc.

The NHIS sample augmentation that began in 2011 has focused on increasing sample sizes in less populous states.

Major change for 2016 NHIS sample design: most sample addresses will come from commercial list(s)