Exploring the USA's 2019 National Health Interview Survey

Fouad Yared

Hunter College, Statistics 754 Design and Analysis of Complex Surveys, Fall 2020

Each year since 1957, the United States federal government conducts the National Health Interview Survey (NHIS), which is seen as the main source of health information for the civilian, noninstitutionalized population in the United States. [1] Researchers at the CDC's National Center for Health Statistics (NCHS) design and document changes to the survey while field workers at the U.S. Census Bureau conduct the survey.

Origin and Purpose

The National Health Interview Survey is mandated by the National Health Survey Act of 1956. The Act allowed the Surgeon General to survey, through sampling, the United States population to "determine the extent of illness and disability and related information" and "to develop and test new or improved methods for obtaining current data on illness and disability and related information." [2] Other intended purposes, as detailed in the report by the U.S. Subcommittee on National Morbidity Survey,[2] were to find out how to:

- prioritize public health problems and allocate resources
- better estimate economic loss to businesses due to disability and injury
- better estimate the market for health-related products and services
- better estimate resources needed at hospitals, nursing homes, home care programs, and other places
- develop ideas for further research based on associations within the data (incidence or prevalence of condition compared to demographic, geographic factors)

Currently, the federal government's Department of Health and Human Services uses data from the NHIS to monitor trends. Public health researchers use the data to learn about access to health care and to evaluate federal health programs.

As mandated by the National Health Survey Act of 1956, results of the annual survey are published each year. Published results are aggregations with pseudo-strata and pseudo-PSUs (for clusters). The National Center for Health Statistics "must adhere to Section 308(d) of the Public Health Service Act … which forbids the disclosure of any information that may compromise the confidentiality promised to survey respondents." NHIS data is "strictly confidential, seen only by persons who work on the NHIS … with a need to know, and such information is not disclosed or released to anyone for any other purpose without the consent of the respondent." [1]

Survey Design: Target Population

Since the NHIS is a survey for the civilian, noninstitutionalized population, the survey universe includes those who are household residents or are in "noninstitutional group quarters" (e.g., homeless shelters, rooming houses, and group homes). Individuals not included in the survey universe are those who:

- do not have a fixed household address
- are active-duty military personnel or are civilians who live on military bases
- live in long-term care institutions (e.g., nursing homes, hospitals for the ill or disabled, or wards for abused or neglected children)
- are in correctional facilities (e.g., jails, prisons, juvenile detention, or halfway houses)
- are U.S. Nationals living in foreign countries

It is unclear how resources needed at nursing homes, one of the suggested purposes of the NHIS, are determined as the nursing home population is excluded from this study.

Survey Design: Sampling

The sampling plan begins by referencing population data provided by the U.S. Census every ten years. This data includes both the total number of people who live in a particular area along with the number of people in particular subgroups, based on demographics (e.g., age, sex, race/ethnicity) and other characteristics (e.g., educational attainment, employment, etc). The 2010 U.S. Census data is used for NHIS surveys conducted between 2016 to 2025.

The second step for researchers at the National Center for Health Statistics is to split the United States into 1,689 geographic areas which consist of counties, county equivalents, or groups of counties. The geographic areas are usually contiguous and always appear entirely within one state. The geographic areas are then divided into two strata based on the population density (i.e., urban or rural). Some areas may only have one stratum.

The third step is to find clusters of addresses within each stratum. This process begins by selecting clusters systematically from each stratum (it is unclear whether clusters are sorted randomly or by certain variables prior to systematic sampling as this is not discussed in the 2019 NHIS documentation). Clusters of addresses appear entirely within geographic areas and there are about 2,500 addresses per cluster. The number of clusters chosen, also known as the sampling probability, is proportional to the size of the strata. The exception is that more clusters are chosen in the 10 least populous states and in D.C. to get the minimum number of addresses required. After selecting a particular cluster of addresses, addresses are sampled within clusters (details on how these addresses are chosen are not mentioned in the 2019 NHIS documentation). Commercial address lists are used and supplemented by field enumeration when the "vendor-supplied list did not adequately include all eligible households."[1] It is unclear whether areas with an inadequate address list were sampled as often or more often than other areas, which is one of the difficulties of this survey design.

Lastly, one Sample Adult (age 18 or older) is chosen among the adults and, if there are any children in the household, one Sample Child (age 17 or younger) is also chosen among the children. The 2019 NHIS documentation states a sample adult and sample child are "randomly chosen by the computer", [1] but it is unclear whether convenience/availability and language play a role in choosing the survey participant.

The final sample for the 2019 NHIS yielded 534 clusters of addresses found in 310 geographic areas. There was at least one cluster selected in each state. One of the changes in 2019 was that the clusters of addresses chosen determined which geographic areas were included (in prior years, the geographic areas chosen determined which clusters to include).

A simple and effective sampling technique is to randomly sample a population. The researchers at the National Center for Health Statistics decided not to randomly select people from the eligible population because simple random sampling was said to be "prohibitive" and "too dispersed throughout the nation for cost-effective interviewing." Although generally more imprecise, [3] researchers decided to use the geographic clustering methodology to help "keep survey operations manageable, cost-effective, and timely."

Collecting Data

While researchers at the CDC's National Center for Health Statistics prepare the annual survey, around 750 field workers from the Census Bureau conduct face-to-face interviews with survey participants. (The NHIS is not a part of and is conducted separately from other U.S. Census Bureau surveys.) An advance letter is mailed prior to an interviewer's visit that informs the potential participant about the purpose of the survey, how much time it will require, and that participation is voluntary. The letter is mailed one week prior the start of the interview period (one week before the 1st of the month).

Initial interviews usually occur in-person but may occur by phone if the participant requests, the weather is bad, or the travel distance is too far. Follow-up to complete an unfinished survey is conducted by phone. About a third of Sample Adult and Sample Child interviews were conducted at least partially by phone (34.3% and 31.7%, respectively).

A computer-assisted personal interviewing (CAPI) system is used to ask questions based on previous responses (e.g., questions about why a person doesn't have health insurance will only be asked if the participant stated, in a previous question, that they do not have health insurance). This system also checks and saves the data and makes the process of data collection more efficient.

Administering Surveys

There are up to three surveys administered in a household. First, there is the household roster which asks who lives in the household and their basic demographic information. As mentioned earlier, one "sample adult" is chosen (18 or older) and one "sample child is chosen" (17 or younger, if there are children present in home). The interview questions in the Sample Adult survey and the Sample Child survey vary somewhat based on age and sex. The Sample Adult information comes from the chosen participant unless s/he is physically or mentally unable to do so. The Sample Child information comes from the parent or adult responsible for the health care of the child, which could be the sample adult. Either interview may appear first.

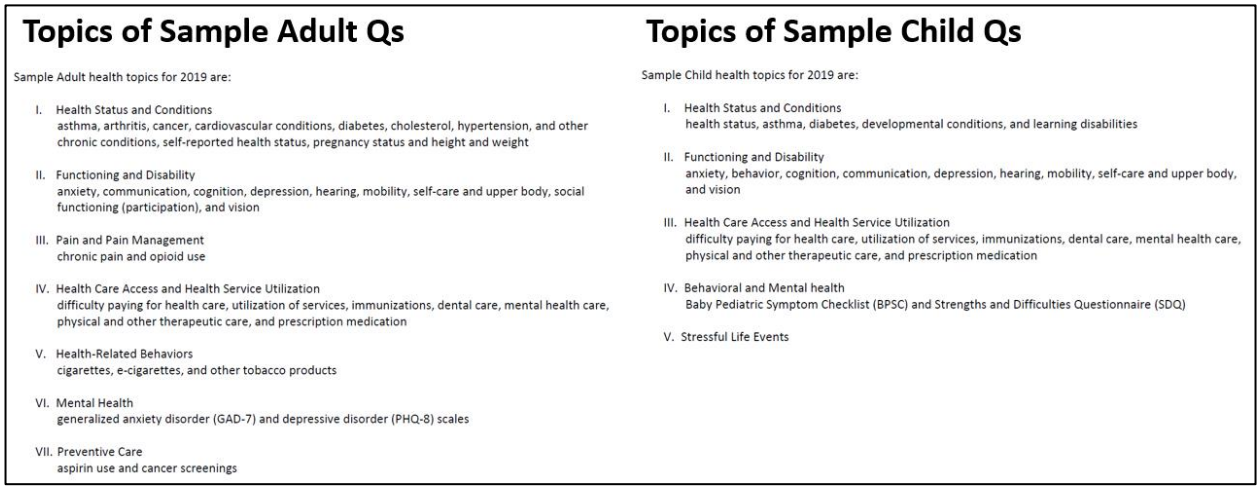
The 2019 NHIS redesign included the removal of the Family Interview. The purpose of the family interview was to detail the relationships of the family members and to collect information on each family member (including whether each household member has health insurance, whether they are employed, what their earnings are, and more). Some information from the Sample Adult surveys may be replicated to the Sample Child surveys. Details below:

- If both the sample adult and sample child belong to the same family, then the family-level questions are only asked to the sample adult.
- If the sample adult refuses or doesn't know the answer to a family-level question, it will appear in the sample child survey if the person responding to the sample child survey is not the sample adult.
- If the sample adult and sample child belong to different families within the same household, both will be asked family-level questions about their families.

It is unclear how much loss of information occurred by removing the Family Interview and reducing the potential differences that may occur within a family. Evaluating the 1997-2018 NHIS documentation and data may provide further information. Difficulties observed in the 2019 redesign were how and why sizable changes were made to the survey for the next 15-20 years. The reason cited for cutting the Family Interview is to "reduce respondent burden."

Question Topics

Questions ranged from determining whether someone ever had various conditions (e.g., asthma, arthritis, or cancer) to asking about day-to-day functioning and mental health, pain management, access to health care, tobacco and e-cigarette use, and preventive care. Topics and subtopics for the Sample Adult and Sample Child surveys can be found in the image below.



Topics of Sample Adult Qs

Sample Adult health topics for 2019 are:

I. Health Status and Conditions
asthma, arthritis, cancer, cardiovascular conditions, diabetes, cholesterol, hypertension, and other chronic conditions, self-reported health status, pregnancy status and height and weight

II. Functioning and Disability
anxiety, communication, cognition, depression, hearing, mobility, self-care and upper body, social functioning (participation), and vision

III. Pain and Pain Management
chronic pain and opioid use

IV. Health Care Access and Health Service Utilization
difficulty paying for health care, utilization of services, immunizations, dental care, mental health care, physical and other therapeutic care, and prescription medication

V. Health-Related Behaviors
cigarettes, e-cigarettes, and other tobacco products

VI. Mental Health
generalized anxiety disorder (GAD-7) and depressive disorder (PHQ-8) scales

VII. Preventive Care
aspirin use and cancer screenings

Topics of Sample Child Qs

Sample Child health topics for 2019 are:

I. Health Status and Conditions
health status, asthma, diabetes, developmental conditions, and learning disabilities

II. Functioning and Disability
anxiety, behavior, cognition, communication, depression, hearing, mobility, self-care and upper body, and vision

III. Health Care Access and Health Service Utilization
difficulty paying for health care, utilization of services, immunizations, dental care, mental health care, physical and other therapeutic care, and prescription medication

IV. Behavioral and Mental health
Baby Pediatric Symptom Checklist (BPSC) and Strengths and Difficulties Questionnaire (SDQ)

V. Stressful Life Events

Most Sample Child topics appear in the Sample Adult topics. Stressful Life Events is one topic that appears only in the Sample Child survey.

The NHIS has four types of questions: annual core, rotating, sponsored, and emerging:

- Annual core questions appear every year. Topics are chronic conditions, functioning and mental health, health insurance, health care access, tobacco use, and demographics.
- Rotating core questions are asked 1 out of 2, 1 out of 3, or 2 out of every 3 years. These ask about services used, allergies, psychological distress, a mental health assessment, industry and occupation, injuries, chronic pain, and preventive services.
- Sponsored content consists of questions another U.S. Federal Agency pays NCHS to ask. Topics were on "cancer control and prevention, immunizations, noncigarette tobacco product use, food security, food program participation, insulin use and arthritis."[1]
- Emerging content changes yearly. Topics in 2019 were on opioid use and managing pain.

Over time, the importance of some topics has changed. Although not originally one of the suggested purposes of the survey, questions around health insurance and health care access are asked annually. Conversely, one of the original purposes for the survey, to better estimate economic loss to businesses due to disability and injury, is asked 2 out of every 3 years.

Sample size, Response Rate

In 2019, 54,231 total eligible households were asked to complete a household roster. Overall, 35,404 households (65.3%) completed the household roster and 33,138 (66.1%) also completed a substantial portion of either the Sample Adult or the Sample Child interview.

Among the 35,404 households that completed a household roster, 35,365 adults and 10,155 children were identified. Of these, 31,997 adults and 9,193 children completed the

survey. The Final Sample Adult response rate is (31,997 / 35,365) = 90.5% which is multiplied by the household response rate 65.3% to get 59.1%. The Final Sample Child response rate is (9,193 / 10,155) = 90.5% * 65.3% = 59.1%, the same as the Sample Adult Response Rate.

Addressing Non-Response

Survey documentation for 2019 notes there are "different response rates among different household and person-level subgroups" [1] but does not mention which targeted methods, if any, were used to reduce non-response and encourage survey completion. Incentives for completing the survey were not mentioned for either the general population or for those in subgroups who do not have high response rates.

Survey Weights

While survey results can be observed with non-weighted frequencies, they do not account for subgroups with lower response rates, which means the results are less representative. Complex surveys such as the NHIS should be evaluated using weights to account for non-response bias. The 2019 weights are the product of multiple parts:

- Base weight: inverse of probability of being included as sample from household
    - With two adults in the household, the inclusion probability would be 1/2
- Adjustment for non-response:
    - Larger weights are given to households and persons with lower response rates. Smaller weights are given to those with higher response rates.
- Response propensities:
    - Households are classified into one of five response propensities by using multilevel logistic regression that uses predictors related to response rates and health outcomes. (Using multilevel logistic regression for weights is an interesting design feature that was not covered in class.)
    - The weight is 1/(median response propensity for that quintile).
        - This value is capped at 2.5
    - Predictors and the cap of 2.5 are reviewed annually.
- Calibrating non-response adjusted weights:
    - Raking is used to make sure the sum of weights for various subgroups is the same. "Non-response adjusted weights are calibrated to U.S. Census Bureau population projections and American Community Survey (ACS) one-year estimates for age, sex, race and ethnicity, educational attainment, Census division, and Metropolitan Statistical Area (MSA) status."

The non-response adjustment for NHIS from 1997-2018 was to create weights based on geographies. Instead of raking, post-stratification was based on age, sex, and race/ethnicity totals. Raking was preferred in 2019 since it could handle more variables.

The public use data file includes both interim (prior to raking) and final (after raking) weights for the Sample Adult and Sample Child surveys. Along with these are pseudo-strata, pseudo-PSUs, and paradata which are predictors for the multilevel logistic regression.

Calculations

Referencing the public use Sample Adult data, the weights of several features were calculated using the survey package in R. Features analyzed include insurance coverage rates, types of coverage, the cost of annual premiums, and the deductibles of private health care plans. Unweighted frequencies for coverage rates are 90.7%, 9.0%, and 0.3% for insured, uninsured, and those who don't know, respectively. As shown in the table on the left, coverage rates ranged from 88% to 11.6% to 0.3% for the insured, uninsured, and those who don't know, respectively. When including weights in the analysis, the share of those insured dropped by 2.7% while the share for those Uninsured increased by 2.6%. Uninsured persons tend to have lower response rates and thus larger weights. Since every adult was asked this question, the sum of weights for the survey population is 250.3 million, which is seen as the U.S. civilian noninstitutionalized population. Percentages were used along witih final weights to ease interpretability.

	Final Weights	Share of Total
Insured (covered)	220,392,011	88.0%
Uninsured (not covered)	29,139,272	11.6%
Don't know	802,899	0.3%
Total	250,334,182	100.0%

The table showing coverage rates by Age Group (below, left) shows there's a tendency for younger people to be more uninsured than older people. The table showing coverage rates by Family Income Group (below, right) shows a tendency for lower income groups to be more uninsured than higher income groups. To reveal more refined drops in coverage, Age and Income groups can be split at different intervals as many individuals who are low-income qualify for their state's Medicaid plan and younger adults may qualify for subsidized health care or health care on their parents insurance plan until age 26. Medicare, public health insurance for those 65 and older that takes covers 80% of expenses, is responsible for the large increase in the insured rate between those age 50-64 and 65-84.

Age Group	Insured	Uninsured	Don't Know
18-29	82.0%	17.4%	0.6%
30-49	84.0%	15.7%	0.3%
50-64	89.3%	10.5%	0.2%
65-84	98.8%	1.0%	0.2%
85+	99.0%	0.8%	0.2%

Family Income Group	Insured	Uninsured	Don't Know
$0 to $34,999	80.6%	19.0%	0.4%
$35,000 to $49,999	82.1%	17.7%	0.2%
$50,000 to $74,999	86.9%	12.7%	0.4%
$75,000 to $99,999	92.9%	6.9%	0.3%
$100,000 or greater	95.4%	4.3%	0.3%

Types of health care plans for adults age 18-64 and 65 or older are shown below. The majority of adults 18-64 had private insurance, followed by Medicaid, and Other coverage. The Uninsured were a sizable portion too. Adults 65 or older usually supplemented their Medicare with Private or Medicare Advantage coverage.

For Adults 18-64

	Final Weights	Share of Total
Private	132,923,518	67.3%
Medicaid and other public	26,720,871	13.5%
Other coverage	8,658,490	4.4%
Uninsured	28,637,929	14.5%
Don't know	692,745	0.4%
Total	197,633,554	100%

For Adults 65 or older

	Final Weights	Share of Total
Private	21,440,131	40.7%
Dual eligible	4,021,734	7.6%
Medicare Advantage	14,727,161	27.9%
Medicare only excluding Medicare Advantage	7,133,482	13.5%
Other coverage	4,677,094	8.9%
Uninsured	501,344	1.0%
Don't know	199,683	0.4%
Total	52,700,629	100.0%

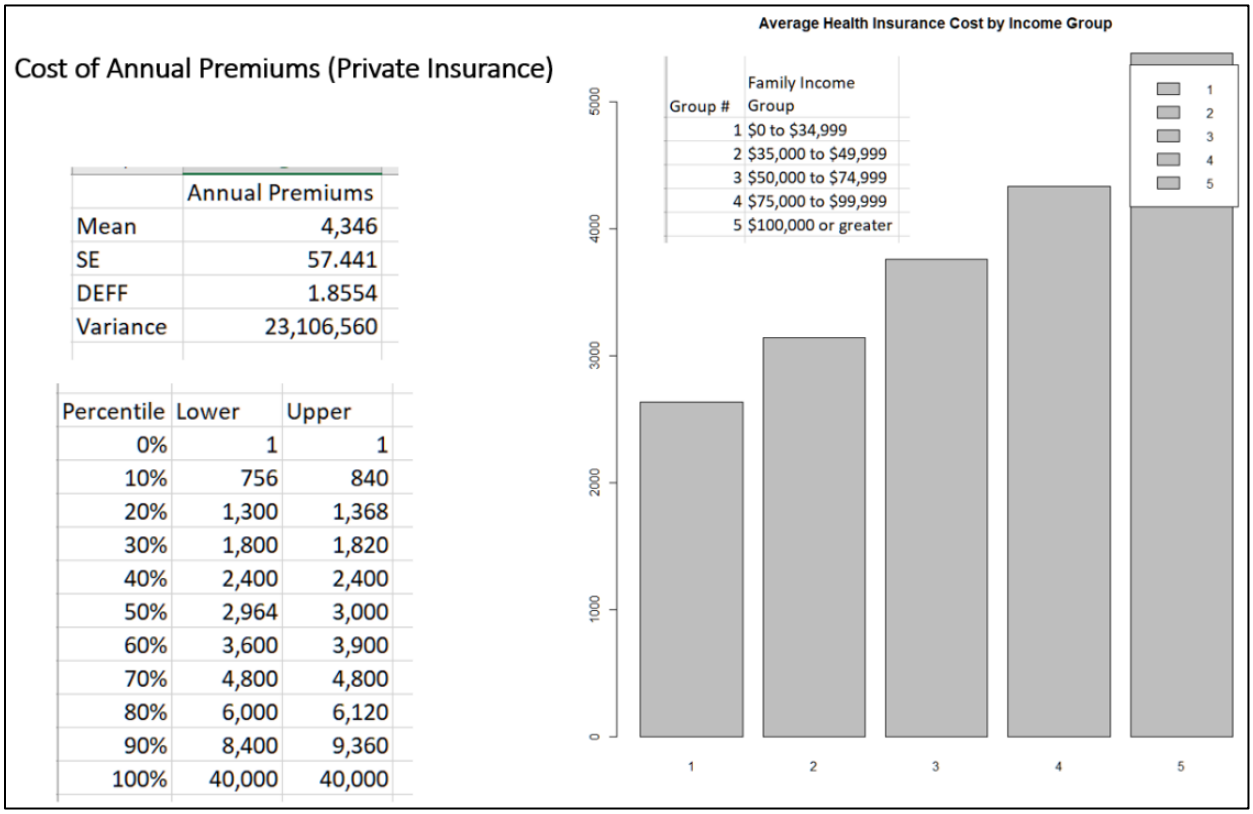
The cost of annual premiums for those with private insurance was analyzed to better understand what people pay. As shown in the table below, the mean of annual premiums is $4,346, the standard error of the mean is 57.441, the design effect is 1.8554, and the variance is over 23 million.

- The design effect is the variance of the stratified clustering design over the variance of the simple random sample when using the same number of observation units. A design effect of 1.85 means that 17,296 persons would need to be interviewed if using a SRS methology (as opposed to 31,997 persons when using a stratified clustering methology.)

Cost of Annual Premiums (Private Insurance)

	Annual Premiums
Mean	4,346
SE	57.441
DEFF	1.8554
Variance	23,106,560

Percentile	Lower	Upper
0%	1	1
10%	756	840
20%	1,300	1,368
30%	1,800	1,820
40%	2,400	2,400
50%	2,964	3,000
60%	3,600	3,900
70%	4,800	4,800
80%	6,000	6,120
90%	8,400	9,360
100%	40,000	40,000



The percentile table above shows how premium costs slowly then dramatically rise when looking at the last 90th to 100th percentiles. The bar plot shows that as family income rises, the average annual premium tends to rise. Families with incomes from $0-$35,000 pay $2,500 on average while families with $100,000 or more pay $5,500 on average.

Among those with a Private Health Care Plan, 74% have a deductible and 26% do not. Of those that have a deductible, 52% have a high deductible health plan and the remaining 48% do not. (A high deductible health plan is based on the number of people in the household. For one person, the threshold is $1,350. For two or more people, it is $2,700.)

	Final Weights	Share of Total
Have a deductible	82,096,503	73.9%
Do not have a deductible	29,068,110	26.1%
Total	111,164,613	100.0%

	Final Weights	Share of Total
Do not have a high deductible health plan	42,910,087	52.3%
Have a high deductible health plan	39,186,417	47.7%
Total	82,096,504	100.0%

References

1. National Center for Health Statistics. National Health Interview Survey, 2019. Public-use data file and documentation. https://www.cdc.gov/nchs/nhis/data-questionnaires-documentation.htm. 2020.
    - Link to public data, documentation: https://www.cdc.gov/nchs/nhis/2019nhis.htm

2. The National Health Survey Act. (1957). *Public Health Reports (1896-1970), 72*(1), 1-4. Retrieved December 8, 2020, from http://www.jstor.org/stable/4589668

3. Lohr, Sharon L. Sampling: Design and Analysis Second Edition. Boca Raton, FL, CRC Press, 2019.