

THE U.S. NATIONAL HEALTH INTERVIEW SURVEY

Jane F. Gentleman¹

ABSTRACT

The National Health Interview Survey (NHIS) is one of the major health data programs in the United States. Through the NHIS, information concerning the health of the U.S. civilian noninstitutionalized population is collected in in-person household interviews throughout the country. The NHIS has been in continuous operation since 1957. The data are collected during 50 weeks of the year and are released annually. Each week's sample is representative. The survey is conducted by the National Center for Health Statistics (NCHS), which is located in the Washington, D.C. area and is a component of the Centers for Disease Control and Prevention (CDC). The U.S. Bureau of the Census, under contract to NCHS, has been the primary data collector for the NHIS since the survey's inception. Annually, the survey obtains information from about 41,000 households containing about 107,000 people. Topics covered by the NHIS include household composition, health status and limitation of activity, injury, health care access and utilization, health insurance, socio-demographic characteristics, income and assets, conditions related to functional limitations and activity limitations, health behaviors, immunization, and other topics covered by supplementary questions that vary from year to year. This paper describes key aspects of the design and content of the NHIS, recent major changes to the survey, future plans for the survey, and means of access to NHIS data.

KEY WORDS: Health status; Health policy; Longitudinal survey data; National Health Interview Survey.

RÉSUMÉ

L'Enquête nationale sur la santé aux États-Unis (National Health Interview Survey, NHIS) est un des plus importants programmes de données sur la santé aux États-Unis. L'information concernant la santé de la population civile hors institutions des États-Unis y est recueillie au moyen d'entrevues à domicile à travers le pays. Le NHIS est en opération depuis 1957. Les données sont recueillies durant 50 semaines et sont publiées annuellement. L'échantillon de chaque semaine est représentatif. Le sondage est réalisé par le National Center for Health Statistics (NCHS). Cet organisme, situé dans les environs de Washington (D.C.), est une branche du Centers for Disease Control and Prevention (CDC) d'Atlanta. Depuis les tous débuts du sondage, le U.S. Census Bureau est responsable de la collecte des données. Chaque année, le sondage obtient l'information d'environ 41 000 ménages comprenant un total d'environ 107 000 personnes. Les sujets couverts par le NHIS sont la composition des ménages, l'état de santé des individus et la limitation dans leurs activités, les blessures, l'accès aux soins de santé, l'utilisation des soins de santé, les assurances-santé, les caractéristiques socio-démographiques, les revenus et les actifs personnels, les conditions reliées aux limitations fonctionnelles et aux limitations dans les activités, les comportements des gens pouvant affecter leur santé, l'immunisation et bien d'autres sujets abordés par des questions supplémentaires qui varient d'année en année. Cet exposé décrira les aspects essentiels dans la conception et le contenu du NHIS, des changements d'envergure récents à l'enquête, les plans pour l'avenir de l'enquête, et l'accès aux données du NHIS.

MOTS CLÉS : L'état de santé des individus; la politique sur la santé; l'analyse des données provenant d'enquêtes longitudinales; l'Enquête nationale sur la santé aux États-Unis.

1. THE BACKGROUND AND NATURE OF THE NATIONAL HEALTH INTERVIEW SURVEY

The National Health Interview Survey (NHIS) is a multi-purpose, cross-sectional health survey conducted

by the National Center for Health Statistics (NCHS), which is part of the U.S. Centers for Disease Control and Prevention (CDC). The NHIS is the principal source of information on the health of the civilian, noninstitutionalized household population of the U.S.

¹ Jane F. Gentleman, Division of Health Interview Statistics, National Center for Health Statistics, 6525 Belcrest Road, Hyattsville, Maryland, U.S.A., 20782, JGentleman@cdc.gov.

NHIS data are used to provide a snapshot of the health of the U.S. population, track national health objectives, evaluate health policies, and conduct public health and other research. The survey has been conducted continuously since its beginning in 1957. Data are released on an annual basis. The U.S. Bureau of the Census, under a contractual agreement, is the data collection agent for the NHIS. The data are collected through a personal household interview by Census interviewers.

The sampling frame for the NHIS is redesigned every ten years to better measure the changing U.S. population and to meet new survey objectives. The new design is implemented five years after each decennial census. Thus, the current design was implemented in 1995 and will be used through 2004. About 41,000 households containing about 107,000 persons are in the NHIS sample each year. The NHIS's probability sample has a multistage, stratified, cluster design and uses an area frame. It was designed to over-sample black and Hispanic subpopulations in order to increase the precision of estimates for those populations. Plans for the next sample design, to be implemented in 2005, include over-sampling these same two subpopulations, and perhaps others.

Households receive a letter in advance of the interview explaining the purpose of the survey, that participation is voluntary, and that confidentiality will be protected. The first contact made by the interviewer is at the home, and the telephone is sometimes used if necessary to continue the interview at a later time.

2. THE NATIONAL HEALTH INTERVIEW SURVEY SINCE 1997

2.1 Changes to the questionnaire in 1997

The NHIS questionnaire underwent major changes that were implemented in 1997. The redesigned questionnaire has a Basic Module that remains essentially unchanged from year to year, with more in-depth questions and/or questions on new topics added as supplements each year, as needed. The Basic Module contains three major components: the Family Core, the Sample Adult Core, and the Sample Child Core. The Family Core component collects information on everyone in the family. Questions that can reasonably be answered by proxy are included in this section of the survey. The sample for the Family Core can serve as a sampling frame for additional surveys. For example, the Medical Expenditure Panel Survey, which is conducted by the Agency for Health Care Policy and Research, collects additional data from some of the NHIS respondents about health care

use, health care expenses, and health insurance coverage. Also, the National Survey of Family Growth has re-used a portion of the NHIS sample. Questions from the Sample Adult Core are administered to a randomly selected adult in each family, and a knowledgeable adult is asked questions from the Sample Child Core about a randomly-selected child under 18.

Another component of the NHIS Basic Module is the Child Immunization Questionnaire, a set of questions sponsored by CDC's National Immunization Program that ask about children's vaccinations. The interviewer asks the respondent (usually a parent) to refer to the children's shot records if they are available, and information about dates, numbers, and types of shots is recorded. The parent is also asked for permission to contact the provider(s) of the children's shots to obtain vaccination records. The provider data subsequently obtained (by mail-in, mail-back questionnaire) permits combination and reconciliation of the two types of information, resulting in increased accuracy. The NHIS data and immunization provider data are also used to adjust estimates from the National Immunization Survey (NIS) to account for non-telephone households not covered by the NIS, which is a telephone survey.

To administer the NHIS questionnaire, the interviewer reads questions from the screen of a laptop computer and types in the responses. This is done in a family situation. Once general information about the household and its occupants has been obtained, the NHIS Family Core is the first component to be administered. A knowledgeable adult provides responses about the whole family, but other family members may be present and may assist in answering questions.

The random selection of a single adult for the Sample Adult Core and a single child for the Sample Child Core is a new feature introduced in 1997. The use of this procedure has advantages and drawbacks. Administering these questions to one adult shortens the total time spent responding to the survey (because the questions are asked of only one person, not of the whole family) and increases the accuracy of responses (because the questions are asked about that one person, i.e., proxy responses are not accepted). However, if the computer program happens to select a Sample Adult who is not at home at the time, the interviewer must return to the household a second time (or perhaps complete that part of the survey by telephone), which increases the total elapsed time from the beginning to the end of the survey, and increases the chances of not completing the survey. Another drawback is that

estimates calculated from Sample Adult data may be based on a smaller sample size because those questions are asked of approximately 41,000 Sample Adults instead of approximately 101,000 household members.

Many significant changes were made to the NHIS questions in 1997, including changes to the wording, reference period, context (surrounding questions), and positioning of questions, the proxy rules, and the type and number of possible answers. One objective of the questionnaire redesign was to shorten the length of the survey, which now takes approximately 65 minutes on average to administer, including supplements. The previous redesign of the NHIS questionnaire had been in 1982; fine tuning of existing questions or more substantial changes were needed after 15 years to maintain relevance and adapt to changes. Demographic and societal changes, changes in the types of health problems occurring, changes in the health care system, and new needs of health policy makers all made it imperative to update the NHIS questionnaire. For example, survey questions needed to be adapted or added to provide additional information about blended families, HIV/AIDS, and HMO membership.

Changes to the NHIS body weight question provide a simple example of the increased accuracy afforded by non-proxy NHIS questions. In 1996, the responding adult was asked "About how much does --- weigh without shoes?" about him/herself and the other members of the household. In 1997, this question became, "About how much do you weigh without shoes?" and was asked of the Sample Adult.

A medication use question is one example of a change in the time reference period covered by a question. A 1994 NHIS question asked, "Does --- NOW have a physical, mental, or emotional problem for which they regularly take prescription medicine?" In 1997, the question became "Does {name of Sample Child} now have a problem for which {he/she} has regularly taken prescription medication for at least three months?" Clearly, rates of taking medication will be higher when based on the older version of the question, and the two types of rates are not comparable. This illustrates the discontinuities that necessarily ensue when questionnaires are redesigned.

Since 1976, NHIS respondents have been able to identify themselves as being of more than one race. Relatively recently, the categories used on the NHIS to indicate race/ethnicity changed in response to new government-wide guidelines. An individual may now

select from a list of 15 possibilities (or identify another race not on the list).

The NHIS has traditionally collected information on health conditions in two ways: by asking directly whether the respondent had been diagnosed with specific conditions, and by ascertaining indirectly which conditions were the causes of behaviors related to ill health, such as contacting a physician's office, staying in bed, and/or cutting down on normal activities. The latter approach yields lower rates for a condition, because the condition will be mentioned by the respondent only if it triggered contact with the health care system or a reduction in the individual's ability to function. Thus, the latter rates tend to exclude conditions with negligible impact. Also, results from the latter approach are confounded by differences in socioeconomic status, because persons of lower socioeconomic status may be less able to stay home from work, to limit their activities, or to seek medical care, so when they do not demonstrate those required behaviors, their health conditions are not counted. For further discussion of this issue, see Madans (1997). The two types of rates have their respective valid uses. For example, NCHS regularly publishes prevalence rates derived from direct questions about the presence of chronic conditions (see, for example, Collins (1997, 2001), Adams et al. (1999), and Benson et al. (1998)), and Hing and Bloom (1990) published rates of chronic conditions for persons with functional dependency.

Because of the profound changes to the NHIS questionnaire in 1997, analysts should be especially careful in comparing pre-1997 results to results from 1997 and later. The changes are fully described in NHIS documentation. NHIS public use data files and documentation are available free of charge on the NCHS web site at www.cdc.gov/nchs/.

2.2 Changes to the instrument in 1997

Another major change in 1997 was the change from a paper-and-pencil questionnaire to a computer assisted personal interview (CAPI) instrument. The use of CAPI eliminates one step in the electronic capture of responses; instead of the responses being written down or checked off by the interviewer during a paper-and-pencil interview and then later captured electronically, the data are electronically captured during the interview. The CAPI software automatically guides the interviewer through the maze of possible questions, displaying them on the computer screen for the interviewer to read aloud. The survey instrument ensures that the correct path, which depends not only on the socio-demographic characteristics of the

respondent, but also on the responses to the health questions, is taken. The responses are typed in, and some responses are edited on the spot, during the interview; for example, the CAPI program can check that responses are within reasonable ranges and are consistent. This permits corroboration by the respondent, who is still present at the time of such editing. When the time comes during the interview for an adult respondent to be selected from among the family members, the computer program selects the adult, ensuring randomness and adherence to the survey protocol. A randomly selected child is similarly identified. The questionnaire can be displayed on the computer screen in either English or Spanish, changing languages on command by the interviewer. On-line help is available for the interviewer and/or the respondent. Notes made by the interviewer can be captured in the file and attended to later, at Census or NCHS. The path of the interview, including the time spent on each question, backups, corrections, etc., can be recorded for future research and analysis. For example, this allows trouble spots in the questionnaire to be identified and studied.

2.3 NHIS supplements since 1997

The 1997 NHIS did not contain a supplement because of the extensive number of changes in the survey that had to be attended to. In 1998, a supplement collected data to track progress toward meeting the objectives of the Healthy People 2000 program, which were to prevent disease and improve health. In 1999, detailed questions about specific chronic health conditions were added to the core. A cancer supplement sponsored by the National Cancer Institute and by CDC is on the 2000 NHIS, beginning its fourth and last quarter in the field at the time of this writing. The year 2001 will see two supplements on the NHIS, one containing questions on children's mental health (sponsored by the National Institute of Mental Health) and one containing new questions to begin tracking progress toward meeting the objectives of the Healthy People 2010 program. A different set of Healthy People 2010 questions is slated to appear on the 2002 NHIS. The Healthy People 2010 questions on the 2001 and 2002 NHIS will be repeated in NHIS supplements twice more during the decade.

The Healthy People 2010 program is a comprehensive, nation-wide health promotion and disease prevention agenda. About 470 objectives have been identified to serve as a road map for improving the health of all people in the U.S. Each objective has a target for improvements to be achieved by 2010. This decade's effort is built on similar initiatives that were pursued over the past two decades. The Healthy People

program has two overarching goals: 1) to increase quality and years of healthy life, and 2) to eliminate health disparities.

NCHS is responsible for coordinating this major effort. Data to monitor progress come from NCHS and other data sources, with NCHS being a major data provider. Over seven Federal Government Departments participate in the project, including Health and Human Services (which contains CDC and NCHS), Commerce, Education, Justice, Labor, Transportation, and the Environmental Protection Agency.

Tobacco Objectives are one example of Healthy People program objectives. These include "Reduce cigarette smoking to a prevalence of no more than 15 percent among people aged 18 and older." Related, more specific objectives include "Reduce cigarette smoking to a prevalence of no more than 20 percent among people with a high school education or less aged 20 and older," "Reduce cigarette smoking to a prevalence of no more than 20 percent among blue-collar workers aged 18 and older," "Reduce cigarette smoking to a prevalence of no more than 15 percent among Hispanics aged 18 and older," and "Reduce cigarette smoking to a prevalence of no more than 18 percent among blacks aged 18 and older." The combination of data from questions in the Healthy People NHIS supplements and the existing rich multivariate NHIS data makes it possible to track progress toward meeting these and many other objectives. For example, Healthy People program data show that cigarette smoking prevalence rates for people aged 18 and over were 29%, 25%, 26%, 27%, 25%, 26%, and 25% in 1987, 1990, 1991, 1992, 1993, 1994, and 1995, respectively.

Further details about the Healthy People program and data are available on the NCHS web site at www.cdc.gov/nchs/.

3. FUTURE PLANS FOR THE NATIONAL HEALTH INTERVIEW SURVEY

Extensive planning and research has already begun to prepare for the new NHIS sample design, to be used in 2005-2014. Plans are also under way to upgrade the software used in the NHIS instrument from CASES software in a DOS environment to Blaise software in a Windows environment. This will provide enhanced data editing capability and will generally be more author-friendly and interviewer-friendly. When respondents' signatures are needed to provide consent, it will also enable NHIS to use devices to collect them electronically, doing away with the present

cumbersome and expensive system of using paper forms, tracking the forms if they are not submitted on time by the interviewer to the Census Bureau, and storing the forms. The new instrument will also be able to use audio-CASI (audio computer assisted self-interviewing) technology to afford more privacy to respondents when especially personal questions are asked. With audio-CASI, the respondent self-administers the survey using a laptop computer, listening to questions on audio headphones and entering responses into the computer. The response categories are usually displayed on the screen, and the questions may or may not be displayed. (See Couper et al., 1998.)

Also in the NHIS future are relational databases to store in-house and public use data files, and a metadata repository to organize the multiple and complex versions of NHIS documentation. These will facilitate editing, imputing, dissemination, and access to NHIS data.

4. THE NCHS RESEARCH DATA CENTER

NCHS has a legal requirement to ensure that no individual can be identified by using publicly-released survey data. This clearly makes it impossible to release all of the survey data. In response to the need for more detailed data than can be released, NCHS has created a Research Data Center that provides analysts with controlled access to much of the in-house microdata.

NCHS provides two modes of special access via its Research Data Center (RDC)—on-site access and remote access. In both cases, potential RDC users submit a detailed proposal to the RDC, indicating the purpose of analysis to be undertaken, specifying the software, methods, and data to be used, and describing the output files. If the proposed analysis is generally sound and can be undertaken without breaching confidentiality, the proposal is approved and the analyst arranges either to visit the RDC or to submit a computer program to be run at NCHS. Restrictions are different for remote access than for on-site access.

A fee is charged for use of the RDC, and users sign a confidentiality agreement. Identifiers such as names and social security numbers are removed from NCHS data used in the RDC, and strict disclosure limits are enforced. User-written computer programs and their output are scanned and screened by RDC staff, both manually and using special software. Complex systems are required to permit this screening process to be highly automated. For more details, see the RDC web site at www.cdc.gov/nchs/r&d/rdc.htm.

REFERENCES

- Adams, P.F., Hendershot, G.E., and Marano, M.A. (1999). *Current estimates from the National Health Interview Survey, 1996*. National Center for Health Statistics. Vital Health Stat 10(200).
- Benson, V., and Marano, M.S. (1998). *Current estimates from the National Health Interview Survey, 1995*. National Center for Health Statistics. Vital Health Stat 10(199). 1998.
- Collins, J.G. (1997). *Prevalence of selected chronic conditions: United States, 1990-1992*. National Center for Health Statistics. Vital Health Stat 10(194).
- Collins, J.G. (2001). *Prevalence of selected chronic conditions: United States, 1993-1995*. National Center for Health Statistics. Vital Health Stat 10(194). In press.
- Couper, M.P., Baker, R.P., Bethlehem, J., Clark, C.Z.F., Martin, J., Nicholls, W.L. II, and O'Reilly, J.M. (1998). *Computer assisted survey information collection*. New York: Wiley.
- Hing, E., and Bloom, B. (1990). *Long-term care for the functionally dependent elderly*. National Center for Health Statistics. Vital Health Stat 13(104).
- Madans, J. (1997). The measurement of health status in the United States. Presented at the Symposium on Health and Mortality, Brussels, Belgium, November 1997.