

**Responses to questions from the NCHS Board of Scientific Counselors'
National Health Interview Survey (NHIS) Review Panel¹**

This report contains responses from the Division of Health Interview Statistics (DHIS) at the National Center for Health Statistics (NCHS) to nine questions posed by the NHIS Review Panel convened by the NCHS Board of Scientific Counselors (BSC). The nine questions were provided to DHIS by Virginia Cain, BSC Executive Secretary, on April 9, 2008. The Review Panel indicated that it would like to receive responses to the questions in writing before their site visit to NCHS on June 9-10, 2008. The table below lists the nine questions and the page numbers in this report where the responses to each question begin.

This report frequently cites parts of the April 25, 2008 update of *The National Health Interview Survey Program: Report to the NCHS Board of Scientific Counselors and its NHIS Review Panel*, which will be referred to in this report as “the April 25, 2008 BSC report.”

#	Question	Page
1	Provide a 1-2 page discussion of the issues involved in the upcoming Title 13 - Title 15 changes. Please provide the relevant language for the statutory changes. What are the perceived pros and cons of the main alternative strategies available to NCHS?	3
2	Provide some background on the release of data items that may be sensitive in some ways. For example, have you considered releasing state identifiers? What are the legal/scientific issues involved in releasing geocoded data? How can such data be better accessed by the user community? Have you considered releasing data pooled over several years as a new data product?	5
3	Has the NHIS considered collecting biomarker data?	9
4	Describe the relationships between NHIS and NHANES and SLAITS. How can NHIS collaborate with non-NCHS entities or surveys?	10
5	Describe the process and standards for the review of data requests. What are the criteria used for screening proposals to use restricted files?	12
6	Data use. What does NCHS do to make the data most useable to researchers? How does the staff keep track of data use? Does anyone track NHIS-based publications and/or citations? Can you provide an overview of data use levels and patterns?	13
7	NHIS high scientific impact. Please briefly identify 3-5 major items that we now know because of the NHIS in fairly recent years.	16
8	How does NHIS see its own role? Is it to provide data for government reports or to provide data sets for the scientific community to analyze? What policy questions does it answer? Who in the government uses the data? Provide responses by use of core data or special modules.	19
9	Future Directions. If there were improvements in the budget what are the 2 or 3 areas that you would want to pursue to improve the NHIS in the future? If budgetary improvements do not occur, what priorities of the NHIS are most important to continue?	20

¹ Page 18 contains a color graph, so it is recommended that that page be printed in color.

See also the following Appendices to this report:

Appendix A. NHIS microdata release and access policy (p. 19)

Appendix B. Division of Health Interview Statistics policy on special release of estimates derived from unreleased National Health Interview Survey microdata (p. 22)

Appendix C. Release of state identifiers with National Health Interview Survey microdata (p. 23)

Question 1: *Provide a 1-2 page discussion of the issues involved in the upcoming Title 13 - Title 15 changes. Please provide the relevant language for the statutory changes. What are the perceived pros and cons of the main alternative strategies available to NCHS?*

Response to Question 1:

As noted in the concluding section of the April 25, 2008 BSC report, NCHS must decide whether the NHIS will become a Title 13 survey, remain a Title 15 survey, or move away from using Census as its fielding contractor. The Title 13 and Title 15 options and their implications are discussed in detail in three places in the April 25, 2008 BSC report: (1) Pages 20-21; (2) Page 25; and (3) Appendix D, entitled *U.S. Census Bureau responses to questions raised by Jane F. Gentleman in February-April 2008*, pages 47-54.

NCHS is considering and, for the time being, pursuing, all three options. To pursue the option of becoming a Title 13 survey, NCHS has begun working with the Census Bureau to draft a Memorandum of Understanding (MOU) between Census and NCHS that would allow NCHS to become a Title 13 survey while having much more access to NHIS data than is normally possible under Title 13. The Consumer Expenditure Survey currently has such a MOU with Census. The scenario being considered would have a remote center—like Census’ current Research Data Centers—set up at NCHS in which properly-sworn-in NHIS staff could access “in-house” NHIS microdata. Such an arrangement would not, however, remove the restriction (noted on pages 20-21 of the April 25, 2008 BSC report) imposed by Title 13 that any follow-up surveys to the NHIS that recontact respondents/households must be conducted by the Census Bureau.

NCHS must carefully consider the numerous implications of the NHIS becoming a Title 13 survey under a special MOU. For example, NCHS would have to negotiate a suitably long term agreement with Census for a Title 13 arrangement to be effective. And NCHS must think about what would happen if, sometime after becoming a Title 13 survey, NCHS wanted NHIS to become a Title 15 survey again or to use another contractor than Census. Custody and access to NHIS data collected while NHIS was a Title 13 survey would have to be considered and negotiated.

Another important consideration would be the effect of NHIS becoming a Title 13 survey on the NHIS record linkage program. The NCHS Office of Analysis and Epidemiology has developed a record linkage program designed to link the health and socio-demographic information of NHIS participants with mortality data, Medicare Enrollment and Claims data, and Social Security Benefit History data, providing a crucial longitudinal component to the NHIS; see the subsection on *Other microdata linked to NHIS microdata* on page 15 of the April 25, 2008 BSC report. The linked NHIS files provide the opportunity to conduct a wide variety of public health and health policy research studies as well as critical methodological studies designed to improve questionnaire design. NCHS would likely want to insure that any MOU with Census allowed for these linkages to continue to be conducted at NCHS.

The following table summarizes some of the pros and cons of the NHIS remaining a Title 15 survey versus the NHIS becoming a Title 13 survey. This is based primarily on information provided by Census (see Appendix D in the April 25, 2008 BSC report). The pros and cons of NHIS moving away from Census as its fielding contractor are largely unknown at this point.

Issue	Description	Pro Title 13	Pro Title 15
Sources of addresses for the sampling frame	The sampling frame for a Title 15 survey cannot be developed from Census' Title 13 sources such as Census' Master Address File (MAF) or the American Community Survey (ACS), and a Title 15 survey cannot use any of the area frame address listings produced for Title 13 surveys (although they share certain infrastructure costs).	X	
Extent of the need for listing	Census is working toward a greatly reduced dependence on listing for its Title 13 surveys. Title 15 surveys will either continue to rely on extensive listing or will have to make extensive use of commercial address files.	X	
Quality of addresses	Title 13 surveys have access to Census' MAF, which Census describes as "the most accurate, comprehensive, and updated source of address information available."	X	
Costs of listing	The next sample design developed by Census for and with the surveys it conducts is targeted for implementation in 2013. If listing continued as a major activity for Title 15 surveys but not for Title 13 surveys, NHIS listing costs as a Title 15 survey would increase dramatically. Census expects that if NHIS remains a Title 15 survey, it will have no other survey with which to share the major listing costs.	X	
Costs of commercial address files	Census plans to use commercial address files to improve its MAF. Beginning in 2013, Census expects that commercial address files will become the primary sources of addresses for Title 15 surveys' sampling frames. Title 15 surveys using commercial address files to form the foundation of their sampling frames would have to bear alone the costs of purchasing and using any commercial address files not directly needed by Census. Census expects that if NHIS remains a Title 15 survey, it will have no other survey with which to share those costs.	X	
Need for research on use of commercial address files, and uncertainty of outcomes	Research on the use of commercial address files is being conducted, but much remains to be done to learn how to use such files to replace substantial portions of the listing now done by Title 15 surveys, and the extent to which commercial address files can successfully replace listing is as yet uncertain.	X	
Control of and access to NHIS data collected under Title 13	In order to be able to access, process, edit, link, and analyze NHIS data in a manner similar to that to which NCHS is now accustomed, NCHS and Census would have to develop an MOU that would set up an offsite location (away from Census) at which sworn-in NCHS staff members could carry out those activities. It is not known what Title 13 restrictions would remain.		X

Issue	Description	Pro Title 13	Pro Title 15
Freedom of choice of contractors for follow-up surveys	Under Title 15, Census can give address information to NCHS. Under Title 13, even if an MOU were developed to permit special access to Title 13 NHIS data by NCHS staff, only Census would be permitted to conduct follow-up surveys to NHIS that required the use of identifiers such as addresses. For example, AHRQ would not be able to continue to use a non-Census contractor to field MEPS, and NHIS supplement sponsors would have to use Census if they wanted to conduct follow-up surveys separate from their NHIS supplements.		X
Length of commitment to being a Title 13 survey	As a Title 15 survey, NCHS formally contracts with Census to field the NHIS one year at a time. If NCHS became a Title 13 survey, that arrangement would have to be much longer term for it to be effective, given the special arrangements that would have to be made.		X
Aftermath of being a Title 13 survey	If NCHS became a Title 13 survey and sometime later wanted to terminate that arrangement, some means of continuing access by NCHS to old Title-13-collected NHIS data would have to be arranged.		X

Question 2: *Provide some background on the release of data items that may be sensitive in some ways. For example, have you considered releasing state identifiers? What are the legal/scientific issues involved in releasing geocoded data? How can such data be better accessed by the user community? Have you considered releasing data pooled over several years as a new data product?*

Response to Question 2: The response to Question 2 is provided below in five separate subsections.

Response to: *Provide some background on the release of data items that may be sensitive in some way.*

Release of microdata and estimates is conditional on meeting the requirements of (1) the Public Health Service Act, (2) NCHS policies governing non-disclosure, and (3) DHIS policies on dissemination of data. Microdata that may be sensitive include *identifiable information* that can be used to establish individual or establishment identity, whether directly—by using items such as name, address, or unique identifying number—or indirectly—by linking data about respondents with external information that directly identifies them, and *confidential information*, which is any identifiable information, or information associated with identifiable information, about a person or establishment collected under an assurance that restricts the degree to which the information can be shared with others. It is important to note that information that by itself would not lead to the identity of a respondent, but that could do so if combined with information already released—released by anyone, not just by NCHS—must also be considered confidential.

The rest of this section summarizes confidentiality-related policies derived from the Public Health Service Act, NCHS confidentiality policies, and DHIS dissemination policies. The information below

about the Public Health Service Act and NCHS policies was derived from the NCHS Staff Manual on Confidentiality, accessible (within NCHS) at

<http://inside.nchs.cdc.gov/opbl/confand%20priv/staffmanual2004-06.pdf>.

1. Public Health Service Act

NCHS operates under the authority and restrictions of Section 308(d) of the Public Health Service Act, which provides in summary that no information obtained in the course of an agency's activities may be used for any purpose other than the purpose for which it was supplied, and that such information may not be published or released in a manner in which the information provider (or person described in the information) is identifiable unless the provider has consented.

Section 308(d) of the Public Health Service Act (42 U.S.C. 242m(d)) provides the basic legal requirement for protecting NCHS records:

No information, if an establishment or person supplying the information or described in it is identifiable, obtained in the course of activities undertaken or supported under section 242b, 242k, or 242l of this title may be used for any purpose other than the purpose for which it was supplied unless such establishment or person has consented (as determined under regulations of the Secretary) to its use for such other purpose; and in the case of information obtained in the course of health statistical or epidemiological activities under section 242b or 242k of this title, such information may not be published or released in other form if the particular establishment or person supplying the information or described in it is identifiable unless such establishment or person has consented (as determined under regulations of the Secretary) to its publication or release in other form.

2. NCHS confidentiality policies

Avoiding inadvertent disclosures through release of microdata

Even when all personal identifiers are removed from a microdata file, a large amount of information remains, and this information may identify NCHS respondents to a person who has access to related information from another source. For example, if file descriptors indicate that the respondent is a Ph.D. in the 30–34 years age group who reports his/her race as white and lives in the Northeast section of the country, the respondent is probably not identifiable. If, however, the file indicates that his age is 31, that he is married to a 42-year-old woman who reports her race as Asian, has three children, and lives in Litchfield County, Conn., it is possible that the respondent may now be identified uniquely, and all information in the file about him and his family would be considered disclosed to anyone with access to the file (who could then identify the person from the given set of characteristics or by matching this information to that contained in another file containing the respondent's name). The place of residence, especially when it is not a heavily populated area, is particularly useful in the identification process. Moreover, if it is known that a particular person or establishment has been selected in the sample, chances are much better that the person or establishment can be identified.

It is NCHS policy to release microdata files for purposes of statistical research only and when the risk of disclosure is judged to be extremely low. In order to make such judgments, NCHS has established a Disclosure Review Board (DRB), which considers proposals for public release of microdata files and makes recommendations to the NCHS Confidentiality Officer.

The DRB applies the following rules to all files released by NCHS that contain any information about individual persons or establishments, except where the supplier of information was told, prior to his giving the information, that the information would be made public:

- A. Before any new or revised microdata files are published, they, together with their full documentation, must be approved for publication by the Confidentiality Officer, who will rely upon assistance from the NCHS Disclosure Review Board in reaching decisions.
- B. The file must not contain information about the subject that could facilitate identification and that is not essential for research purposes (e.g., exact date of the subject's birth, excessive detail for occupation, extreme values of income and age, detailed race or ethnicity for small and highly visible groups, and other characteristics that would make an individual or establishment easier to identify). It is recommended that the following be consulted concerning possible techniques that would permit the maximum amount of information to be released consistent with sound principles of statistical disclosure limitation: The NCHS Checklist on Disclosure Potential of Data and Statistical Policy, accessible (within NCHS) at <http://inside.nchs.cdc.gov/opbl/confand%20priv/Disclosure%20checklist%20for%20release%20of%20NCHS%20data.pdf>, and Statistical Policy Working Paper 22 - Report on Statistical Disclosure Limitation Methodology, Report on Statistical Disclosure Limitation Methodology, Office of Information and Regulatory Affairs, Office of Management and Budget, accessible at <http://www.fcs.m.gov/working-papers/wp22.html>.
- C. Geographic places that have fewer than 100,000 people are not to be identified on the file. Depending upon the statistical structure of a file and other circumstances, a higher figure may be employed. It is the responsibility of the program proposing the data release to determine the disclosure risk associated with the proposed minimum size of geographic areas to be identified.
- D. Characteristics of an area are not to appear on the file if they would uniquely identify an area of less than 100,000 people (e.g., a variable describing the size of a metropolitan area in which a respondent was interviewed providing for a category of less than 100,000 in a file where a region is also provided).
- E. Information on the drawing of the sample that might assist in identifying a respondent must not be released outside NCHS. Thus, the identities of primary sampling units (PSUs) are not to be made available outside NCHS except in limited circumstances and as approved by the Confidentiality Officer.

Restricted access to microdata files with identifiable data

Under certain circumstances, access to microdata files containing high levels of detail may be granted to researchers by the NCHS Research Data Center (RDC). Via the RDC, analytical manipulation of elements of confidential files not released to the public is permitted. However, the RDC-accessible files contain no names, addresses, or other direct identifiers. No statistical output can be removed from the RDC without being subjected to statistical disclosure analysis by RDC staff, and access to information within the RDC is highly restricted. The RDC also has a remote access system, and RDC users can arrange to access NCHS data using any one of the nine Census Bureau Research Data Centers. See the RDC Website at <http://www.cdc.gov/nchs/r&d/rdc.htm>.

3. DHIS dissemination policies

DHIS has developed division policies for microdata release and for release of estimates derived from unreleased NHIS microdata; see Appendices A and B, respectively. The DHIS policies are consistent with those of CDC and NCHS and are guided by the principle of making high quality data available as widely as practicable, as soon as possible after data collection, and in as much detail as possible, while maintaining survey participant confidentiality.

See also the response to Question 5 below.

Response to: *For example, have you considered releasing state identifiers?*

DHIS would like to release state identifiers (and has occasionally done so) under appropriate circumstances, i.e., that the sample sizes are large enough to produce sufficiently precise estimates, that coverage is adequate, and that confidentiality is assured. To address these issues, DHIS recently commissioned an internal report on the pros and cons of releasing state identifiers; see Appendix C, entitled *Release of state identifiers with National Health Interview Survey microdata*.

Response to: *How can such data [sensitive data items] be better accessed by the user community?*

The NCHS Research Data Center (RDC) was established in 1998 in response to a growing demand from the research community for access to restricted NCHS data. Through the RDC, researchers not affiliated with NCHS can access such information without compromising confidentiality. To obtain access to restricted data sets, researchers are required to submit a proposal to the RDC that is evaluated based upon the following criteria: availability of RDC resources, the project's support of the mission of NCHS, and the project's feasibility.

There are two primary methods for accessing RDC data: onsite and remote. Onsite researchers have greater access to data through a variety of programming options and immediate RDC staff assistance. Remote access, however, only allows researchers to use SAS to access information. Generally, there is an RDC service fee for accessing restricted data, either on-site or remotely. More information is available on the RDC Website at www.cdc.gov/nchs/r&d/rdc.htm.

One way to improve access to restricted data would be to increase the number of NCHS Research Data Centers across the country, and indeed, NCHS has recently entered into an agreement with the Census Bureau for NCHS to share facilities with Census RDCs across the country. The NCHS RDC is also planning to open another RDC in Atlanta, CDC's home location.

Another way to improve access to restricted data would be for the RDC to offer extensive custom computer programming services (at a cost) to the public. (In fact, this would be a desirable service whether or not the data were restricted.)

See also the response to Question 5 below.

Response to: *What are the scientific issues involved in releasing geocoded data?*

See also: (1) The subsection on *Geocoding* on page 12 in section 3 in the April 25, 2008 BSC report; and (2) Appendix C, entitled *Release of state identifiers with National Health Interview Survey microdata*.

The central scientific issue involved in releasing geocoded NHIS data is that the NHIS sample is in limited geographic areas, which means that the NHIS data cannot necessarily or generally be used to make direct estimates much below the U.S. level. The NHIS sample process begins by selecting a sample of geographic areas (counties, or groups of adjacent counties) from the ~3,100 counties in the U.S. This is done because the NHIS interview is conducted by an interviewer visiting a sample address. The survey has much lower interviewing costs with this type of sample design than if a random sample of addresses from all U.S. counties were selected.

The NHIS public use files always have included Census Region geographic information. The U.S. is partitioned into 4 Census Regions, each consisting of 9 or more states, and each historic NHIS sample design has selected large samples from all 4 Census Regions. Thus, data users can make direct Census-Region-level estimates with all historic NHIS data. Most, but not all, historic NHIS sample designs have selected sample from each U.S. state. Thus, releasing NHIS data with state identifiers would not allow data users to make direct state-level estimates for all states for all historic NHIS data. All historic NHIS sample designs have been limited to less than 1,000 counties, so releasing NHIS data with state and county identifiers would not allow data users to make direct county-level estimates for over two-thirds of all U.S. counties.

Response to: *Have you considered releasing data pooled over several years as a new data product?*

We assume that the question refers to the possibility of pooling microdata for selected variables over several years because, for individual years, data for those variables cannot be released for confidentiality reasons and/or because of low precision for estimates. For example, NHIS descriptive statistics publications cannot show all race categories listed on the questionnaire. In DHIS’ annual publication *Summary Health Statistics for the U.S. Population*, e.g., a footnote on many of the tables explains that “Persons who indicated a single race other than the groups shown are included in the total...but not shown separately because of small sample sizes.” The table shows counts by race of persons who selected “1 race,” but the table must group some race categories together and omit others altogether.

DHIS provides guidance to data users—in documentation and in workshops—about pooling data from public use files, and DHIS analysts frequently pool data in their in-house analyses. But DHIS has not released files of pooled data to allow outside analysis of data that would be restricted if they were not pooled. To pool and release such data as a data product, choices of what to release and careful study of the consequences would be needed, because what is released today affects what can be released subsequently.

Question 3: *Has the NHIS considered collecting biomarker data?*

Response to Question 3:

Known problems with the validity of some NHIS height and weight data have led NHIS staff to wish that funding were available for collecting actual physical measurements of height and weight rather than (or in addition to) collecting self-reported height and weight figures. We understand that

Statistics Canada added such a component to its Canadian Community Health Survey in a previous survey cycle, and we intend to learn more from them about that activity.

Consideration has also been given to collecting other types of biological and environmental specimens such as saliva, window sill swabs, air quality, and tap water, but no sponsors have been identified to fund such options.

Question 4: Describe the relationships between NHIS and NHANES and SLAITS. How can NHIS collaborate with non-NCHS entities or surveys?

Response to Question 4: The response to Question 4 is provided below in three separate subsections.

Response to: Describe the relationships between NHIS and NHANES.

The surveys now known as NHIS and NHANES both came into existence after the passage of the National Health Survey Act in 1956. Initially, both were parts of the "National Health Survey." The "National Health Survey - Health Interview Survey," now known as NHIS, has always been a continuous nationwide survey since it began in July 1957. The "National Health Survey - Health Examination Survey," now known as NHANES, began as a series of separate surveys (or cycles) in November 1959, becoming a continuous survey in 1999.

Both surveys collect health information from a random sample of the U.S. civilian noninstitutionalized population. The NHIS has a much larger sample size than NHANES. The NHIS annual sample size is now approximately 87,500 persons, while the NHANES annual sample size is approximately 5,000 persons. Both surveys include administration of a questionnaire by an interviewer who makes a personal visit to a sample address. NHANES has an additional major component, consisting of direct physical examinations of sample persons, along with health-related clinical and laboratory tests. The physical examinations are conducted in Mobile Examination Centers (MECs), which are physically moved from one NHANES sample area to the next one. Use of MECs requires extensive geographic clustering of the NHANES sample.

The samples for the two surveys always have been distinct. Sample design researchers for NHANES III (1988-94) considered the option of using a sample of either persons or addresses previously interviewed for NHIS, but this option was rejected as not being as efficient as an independent sample for NHANES III. The sampling process for several early cycles of NHANES either used the NHIS universe of primary sampling units (PSU) for sampling, or selected a subsample of PSUs already selected for NHIS, but NHANES always selected a sample distinct from NHIS within PSUs actually sampled for both NHIS and NHANES.

NCHS has decided to do some brainstorming and investigate the possibility of somehow combining NHIS and NHANES in the future, so as to use funds more efficiently. For example, NHANES administers an interview before administering the MEC portion of the survey, so combining that NHANES interview and the NHIS interview could be considered. Much thought would have to go into developing a survey that could meet the distinct needs of both NHANES and NHIS.

Response to: *Describe the relationships between NHIS and SLAITS.*

The State and Local Area Integrated Telephone Survey (SLAITS) is physically located in the Division of Health Interview Statistics, but its staff and survey activities have been mostly independent of the NHIS. Areas where there has been collaboration include the addition of cell phone usage questions on the NHIS to assess random digit dialing (RDD) survey undercoverage. Extensive analyses of the resulting NHIS data have been conducted and disseminated by SLAITS staff. Results of these studies are especially relevant to SLAITS and to the National Immunization Survey (NIS), which is also a telephone survey and which serves as the SLAITS sample platform.

Another collaboration between NHIS and SLAITS was the addition of screening questions to the NHIS to identify children with special health care needs and then compare those results to results from SLAITS.

Also, the NHIS has included questions on immunization that were, in some years, accompanied by a follow-up immunization-provider record check component. The resulting findings are compared to estimates from the National Immunization Survey to examine undercoverage that results from excluding cell-phone-only households and non-phone households from NIS, and to examine nonresponse, since the NHIS has higher response rates than telephone surveys.

Other collaborations have been under consideration, including adding other questionnaire batteries from the telephone surveys to the NHIS to continue to assess the impact of undercoverage and nonresponse in RDD surveys.

Another potentially important area of collaboration that has not been explored in-depth would be to use SLAITS to enhance the NHIS by increasing the NHIS sample size in specific geographic areas such as states. SLAITS could include, for example, an abbreviated NHIS questionnaire using state-level RDD samples. This would produce the frequently requested estimates for state health data, with national nonresponse/undercoverage adjustments derived from the NHIS. These dual mode, dual frame surveys could have many useful applications. Similarly, it might be possible to use the SLAITS phone interview capacity to reduce the costs of NHIS in-person interviewing by conducting a portion of the NHIS questionnaire by phone, although extensive research and testing would be required to determine if such an approach would be feasible and cost-effective. Also, SLAITS could be used to conduct follow-ups to the NHIS by phone.

Response to: *How can NHIS collaborate with non-NCHS entities or surveys?*

NHIS can and does collaborate with non-NCHS entities and surveys. Some of those collaborative activities are as follows:

- Collaboration with sponsors and potential sponsors of NHIS supplements. See the subsection on *Supplements* on page 7 of the April 25, 2008 BSC report.
- Collaboration with the Agency for Healthcare Research and Quality (AHRQ), whose MEPS survey is a follow-up survey to NHIS. See the subsection on *Integration of the NHIS with the MEPS* on page 15 of the April 25, 2008 BSC report.
- Meetings with the NIH NHIS Interest Group. Since 2005, DHIS staff members and representatives of various NIH Institutes have met periodically at NIH to share information about the status and plans of the NHIS, procedures for developing NHIS

supplements, and the needs of supplement sponsors. NIH Institutes are encouraged to combine resources to sponsor supplements.

- Meetings with other CDC agencies in Atlanta, as we seek to establish a CDC NHIS Interest Group.
- Participating in joint briefing sessions with AHRQ and the producers (Bureau of Labor Statistics and Census) of the Current Population Survey, to provide each other with health insurance estimates just prior to their release.
- Supporting DHHS’ Healthy People Program by fielding NHIS questions—some without funding from outside NCHS—to monitor Healthy People objectives. A table listing over 110 Healthy People objectives for which NHIS data are used to measure progress is available on request. The table shows for which objectives and in which years from 1998 to 2008 core questions and supplement questions were fielded. See some examples in the response to Question 7.
- Participating in research evaluating survey estimates of Medicaid enrollment based on administrative data. A collaboration among the State Health Access Data Assistance Center of the University of Minnesota (SHADAC), NCHS, the Office of the Assistant Secretary for Planning and Evaluation (ASPE), Census, and the Centers for Medicare and Medicaid Services (CMS), partially funded by the Robert Wood Johnson Foundation and ASPE, compared estimates of Medicaid enrollment obtained from the Current Population Survey (CPS), the NHIS, and MEPS. Results indicated that NHIS estimates of Medicaid enrollment were closer to the estimates from the CMS administrative files than estimates obtained from CPS.
- Supporting international projects to develop common batteries of questions that can be used in multiple surveys. For example, as an experiment, NHIS will soon be fielding two similar sets of disability questions—one set from the Current Population Survey and another from the American Community Survey—that are the subject of international interest.
- Responding to emergent Departmental interests. For example, ASPE is working with DHIS staff to develop NHIS questions on Health Information Technology that are of interest to ASPE.
- Collaborating with the California Health Interview Survey by serving on its advisory groups, participating in its benchmarking projects, providing NHIS data, etc.
- Providing assistance to the University of Minnesota’s Integrated Health Interview Series (IHIS) project; see the subsection on *The Integrated Health Interview Series* on page 23 of the April 25, 2008 BSC report.
- Interacting extensively with Statistics Canada; see the subsection on *Fostering a close and productive international relationship* on page 23 of the April 25, 2008 BSC report.

Question 5: *Describe the process and standards for the review of data requests. What are the criteria used for screening proposals to use restricted files?*

Response to Question 5:

We assume that “data requests” here means requests for access to restricted microdata.

The DHIS policy on microdata release is provided in Appendix A. NHIS microdata access mechanisms include:

- NHIS Quality Assurance/Quality Control Collaborator Agreements enabling collaborators early access to *portions* of unreleased or non-public files.
- NHIS Data Support Agreements designed to obtain outside expertise in data collection or processing.
- NHIS Special Use Data Agreements to provide a limited non-public special dataset that cannot be released publicly.
- Approved applications for use of the NCHS Research Data Center.
- NCHS Designated Agent Authority, a venue based upon the Confidential Information Protection and Statistical Efficiency Act of 2002 (CIPSEA), allowing access for non-NCHS employees under very specific conditions.

The DHIS policy identifies, for each of the mechanisms mentioned above, who the data recipients may be, when access is granted, and what the agreement process is. The criteria for access are:

- Whether the respondents’ informed consent to the NHIS provided for the data to be used for purposes proposed by the requestor, and for persons other than NCHS staff.
- The need for such data (i.e., indication that the user could not accomplish the analysis with more generally available releases of the data).
- The ability of the requestor to provide adequate safeguards as defined by DHIS.

The Research Data Center’s criteria for evaluating proposals to use it include:

- Appropriate use of the data, e.g., use of the data in accordance with the informed consent procedures associated with the collection of the data.
- Suitability of the data for the proposed use.
- General scientific and technical feasibility of the proposed project (to the extent that that can be determined without an intense scientific review).
- Availability of RDC resources.
- Risk of disclosure of restricted information.
- Accordance with the mission of the NCHS “...to provide statistical information that will guide actions and policies to improve the health of the American people.”

One example of DHIS providing special access to restricted NHIS data is the Special Use Data Agreement Between NCHS and UCLA Center for Health Policy Research Regarding Provision of Access to NHIS Data. In 2006, under this agreement, NCHS provided UCLA with two variables containing record identifiers corresponding to all records in the 2003 and 2004 public use NHIS microdata files with data from California households. Access to those identifiers by the Center for Health Policy Research supported an ongoing research project for benchmarking California Health Interview Survey (CHIS) data by comparing estimates from CHIS with estimates from NHIS.

See also the response to Question 2 above.

Question 6: *Data use. What does NCHS do to make the data most useable to researchers? How does the staff keep track of data use? Does anyone track NHIS-based publications and/or citations? Can you provide an overview of data use levels and patterns?*

Response to Question 6: The response to Question 6 is provided below in three separate subsections.

Response to: *Data use. What does NCHS do to make the data most useable to researchers?*

DHIS strives to make NHIS data available as widely as possible and in as “user-friendly” a manner as possible by processing, checking, editing, and imputing data; by checking new data completely through two testing cycles; by creating additional derived variables (“recodes”) for the public use microdata files to simplify analysis; by provision of extensive documentation of the data; by release of numerous analytic and methodological reports; by supporting the DHIS data requests program, and by supporting the NHIS ListServ. In addition, DHIS conducts training sessions for NHIS users at professional venues such as American Public Health Association meetings, AcademyHealth meetings, and the NCHS Data Users Conference. Also, free day-long NHIS user workshops are held at NCHS on the day after each Data Users Conference, and at other venues by special arrangement. For example, a NHIS Users Workshop will be held at the University of Maryland’s new College Park School of Public Health in the fall of 2008. NHIS experts also participate occasionally in the NCHS University Visitation Program, which funds visits by experts to speak at universities.

NHIS releases of microdata files also include the questionnaire; instructions for interviewers; documentation on collection, editing, processing of the data; and frequencies of every data item on the microdata file. In addition, sample computer programs are provided that can be used to manipulate the data using SAS, Stata, or SPSS. Complex sample design variance estimation guidance is provided, including sample fragments of variance estimation computer code for SUDAAN, SAS survey procedures, Stata, SPSS, R, and VPLX.

Microdata files are currently released less than 6 months after the end of the each data collection year. Imputed income files are released shortly after the annual microdata release, and paradata files linkable to the health-related files are now released annually. Public use NHIS microdata files are available on the NHIS Website for survey years 1966-2006, and files for survey years 1963-1965 are being prepared for release. Currently, three Summary Health Statistics reports containing descriptive statistics and highlights for a wide variety of health and demographic characteristics are released annually on the Internet within 3 months of the release of the microdata files.

DHIS provides two quarterly reports and one semi-annual report through the NHIS Early Release Program; see the subsection on *Timely release of estimates through the NHIS Early Release Program* on page 15 of the April 25, 2008 BSC report.

DHIS recently began releasing annual files of paradata; see the subsection on *Pioneering release of paradata* on page 14 of the April 25, 2008 BSC report.

An appointed team of staff members (the DATA Request Triage team, or DART team) responds to requests from outside DHIS for NHIS information, data, and analyses. Generally, the DART team does not perform extensive custom runs for individuals, unless the requestor is a senior member of the government. Reports generated from the NHIS data request database initiated in July 2007 show the following distribution of requests by type of requestor: 39% government; 37% academia or research; 7% non-profit organizations; 6% for-profit businesses; and 11% other types. See also the subsection on the *Data requests program* on page 17 of the April 25, 2008 BSC report.

DHIS assists staff members of the Integrated Health Interview Series (IHIS) in producing a harmonized subset of NHIS data and documentation to facilitate trend analysis; see the subsection on *The Integrated Health Interview Series* on page 23 of the April 25, 2008 BSC report.

Response to: *How does the staff keep track of data use? Does anyone track NHIS-based publications and/or citations?*

DHIS does not have a fully-developed program to keep track of NHIS data use, publication use, or citations. We know and hear from many sources who many of our users are and how they use our data and publications, but it is likely that our information is uneven and full of gaps. It is easier now than it used to be to track use of NHIS data and publications because we have the Internet and sophisticated search engines. However, not all users of NHIS data and publications cite their source, and even when they do, not all citations are trackable. For example, the use of NHIS data by teachers in their classes cannot be tracked, and graduate students who use NHIS data in their theses may not publish their theses. Also, under CDC's "one voice" policy, the NCHS logo is no longer used on NCHS publications, and citations of publications about NHIS often mention CDC but not NCHS. In addition, users who download public use microdata from the NHIS Website are not required to register before doing so, and the current CDC Web statistics system does not collect data on downloads of individual microdata files.

The NCHS Library has modern software to facilitate tracking of use of NHIS. Every week, the Library provides DHIS with an ISI Web of Knowledge Alert listing new articles citing the NHIS. For example, DHIS was told about an article entitled *Racial differences in functional capacity among patients with heart disease: analysis of the National Health Interview Survey* appeared in the Journal of the American College of Cardiology in March 2008. When an abstract and subject matter might be of particular relevance to a DHIS staff member, the Knowledge Alert is passed to them as well as being stored in an electronic file. When a full copy of an article is desired, the NCHS library has access to most original sources and can obtain a copy of the article for DHIS.

NCHS' small Public Affairs Office provides monthly listings of news stories that mention NCHS. For example, one item in the April 2008 report (which contained synopses of 129 news stories) described an item on *The News At 10* (WBFF-TV, FOX, in Baltimore) as follows: "Parents are loving their children to death. According to the National Center for Health Statistics 16% of U.S. children 6 and older are overweight and 34% of them are at risk." But it is not clear from that synopsis (and perhaps not even from the full broadcast) whether the estimates being quoted are from NHIS data.

The relatively new NHIS data requests data base now keeps track of incoming requests for NHIS information, estimates, and data. These data can be analyzed to study user needs. Staff members who respond to data requests encourage users to send DHIS copies of any of their articles that use NHIS data, but the users rarely do so. See also the subsection on the *Data requests program* on page 17 of the April 25, 2008 BSC report, and see the response to Question 6.

DHIS learned about NHIS user accomplishments when it held the NHIS Poster Contest to commemorate the 50th anniversary of the NHIS in 2007. Contest entrants were asked to submit a description of how they used NHIS data or analyses successfully to solve a research problem, to evaluate a policy, to teach a class, to write a report, etc. As it turned out, each of the 10 winning posters describes a body of important work using the NHIS.

Other feedback from NHIS users comes to DHIS via the NCHS Data Users Conference, the NHIS ListServ, NHIS, and user workshops.

Response to: *Can you provide an overview of data use levels and patterns?*

We are unable to do that at this time.

Question 7: *NHIS high scientific impact. Please briefly identify 3-5 major items that we now know because of the NHIS in fairly recent years.*

Response to Question 7:

Health insurance coverage

The NHIS provides the most comprehensive national and regional information on health insurance status of the noninstitutionalized civilian population of any general federal survey, and it is the only survey with a wide variety of health status, risk behaviors, health care access, and utilization measures, thus providing a rich assortment of multivariate information. Although the CPS provides the most widely cited estimates of uninsurance, it is not primarily a health, health insurance, or health care survey; its primary purpose is to provide employment and income data. The American Community Survey (ACS) has a much larger sample size but does not ask as detailed questions as the CPS. Thus, the NHIS, while not providing state-level estimates, is the only survey with sufficient data for examining disparities in race/ethnicity subgroups at the national level. The NHIS insurance data enable tracking changes in health insurance coverage of children and adults by type of health insurance plan and numerous socio-demographic characteristics associated with health status, health care access, and utilization.

NHIS staff members continuously monitor health insurance programs and policies in the United States so that the NHIS can address emerging health insurance issues. The versatility and depth of the health insurance section of the NHIS enables policymakers and researchers to examine trends in health coverage and emerging health access issues. Following the implementation of the State Children’s Health Insurance Program (SCHIP) in 1997, the NHIS has been able to monitor the impact of SCHIP over time. There has been a decrease in uninsurance for children, especially among children who have family incomes that are less than 200% of the Federal poverty level, the very children that the SCHIP program was designed to cover. More recently, national attention to consumer-directed health care has increased following the enactment of the Medicare Prescription Drug Improvement and Modernization Act of 2003 (P.L. 108-173), which established tax-advantaged health savings accounts. Three additional questions were added to the health insurance section of the NHIS to monitor enrollment in consumer-directed health care among persons with private health insurance. In March 2008, DHIS released preliminary rates of enrollment in high deductible health plans (HDHPs); of being in a plan with high deductibles coupled with health savings accounts, also known as consumer-directed health plans (CDHPs); and of being in a family with a flexible spending account (FSA) for medical expenses not otherwise covered. The NHIS estimates are among the first to be released for national level use of this type of health insurance coverage. Based on data from the January - September 2007 NHIS, 17.5% of persons under 65 years of age with private health insurance were enrolled in a HDHP, 4.5% were enrolled in a CDHP, and 16.9% were in a family with a FSA for medical expenses. Adults aged 18-64 with more than a high school diploma were more likely to be covered by a HDHP, more likely

to be covered by a CDHP and more likely to be in a family with an FSA for medical expenses than those who had only a high school diploma or were not high school graduates. Research is under way to analyze these new data more thoroughly.

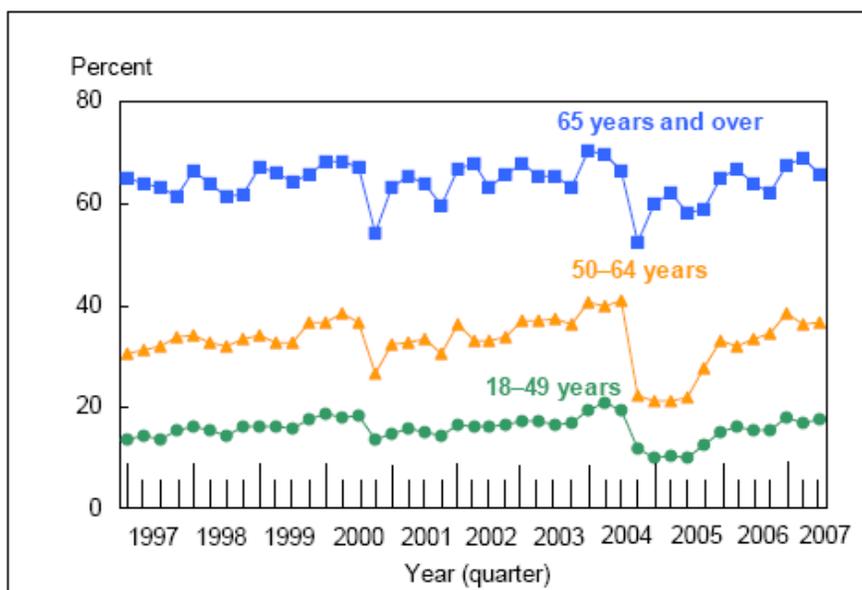
DHIS was pleased to learn earlier this month that the Congressional Research Service is updating the Green Book,² which is produced for the U.S. House Committee on Ways and Means. In the past, the CPS has been used to produce a table in the book that examines Medicaid coverage by various poverty and age groupings. DHIS was told that the producers of the Green Book recently became aware that the NHIS does a better job of collecting Medicaid data than the CPS, so they have decided to use the NHIS as the data source for the table. DHIS staff provided them with the needed estimates.

Flu vaccination

The NHIS provides data on influenza vaccination for adults. In two separate questions, adults of all ages are asked if they received an influenza vaccination by injection (shot) or by intranasal spray during the past 12 months. Respondents who indicate that they received a vaccination are asked in which month and year the vaccination was received. NHIS data are used to update and inform CDC's Advisory Committee on Immunization Practices of current estimates of influenza vaccination among various target groups for vaccination, such as older adults, children, high-risk adults, health care workers, and household contacts of individuals in target groups. NHIS allows tracking the receipt of influenza immunization and the impact of shortage or delays in vaccine delivery on the immunization of the U.S. population. Estimates of influenza vaccination are presented in the NHIS Early Release Program's quarterly report on key health indicators. The following graph from a recent NHIS Early Release shows time trends in flu vaccination. Large dips in vaccination rates due to vaccine shortages are evident at the ends of 2000 and 2004, and, to a lesser extent, at the end of 2001.

² "The Green Book" is the informal name for a publication entitled *Background Material and Data on the Programs within the Jurisdiction of the Committee on Ways and Means*. The Green Book provides updated data on major programs within the Committee's jurisdiction, as well as related programs and issues. Compiled by the Ways and Means staff with the assistance of the Congressional Research Service and various Federal agencies, since its first publication in 1981 the Green Book has become an important reference guide for legislators, program administrators, scholars, and interested citizens. The Green Book is publicly available.

Figure 4.1. Percentage of adults aged 18 years and over who had received an influenza vaccination during the past 12 months, by age group and quarter: United States, 1997–September 2007



Use of complementary and alternative medicine

The 2002 and 2007 NHIS included lengthy supplements on Complementary and Alternative Medicine (CAM), sponsored by NIH’s National Center for Complementary and Alternative Medicine. Questions were asked about use of acupuncture, folk medicine, chiropractic care, massage, special diets, relaxation techniques, faith healing, etc. Combined with the core health data routinely collected on the NHIS, these supplements provided a wealth of information never before available on who uses CAM, the purposes for which CAM is used, the types of CAM used, and what else is used besides CAM. We know that people often turn to CAM because conventional medicine is too costly and that physicians are suggesting that their patients try CAM. The NHIS is also the only source of national information on the use of CAM by children. DHIS analysts and NCCAM staff have jointly produced several publications based on analysis of the CAM data.

Cancer control

Lengthy Cancer Control Supplements were on the 2000 and 2005 NHIS, and another is planned for 2010. In some years between those years, shorter Cancer Control Supplements are fielded. These supplements are sponsored by NIH’s National Cancer Institute (NCI) and CDC’s National Center for Chronic Disease Prevention and Health Promotion (NCCDPHP).

A poster presented by NCI at the NHIS 50th anniversary commemorative conference in June 2007 indicates the value placed by the supplement sponsors on the knowledge gained from NHIS data. The title of the poster was “40 Years of Progress in Monitoring Cancer Control, NHIS & NCI.” The following is an excerpt from the poster:

“The National Cancer Institute (NCI) and the National Center for Health Statistics (NCHS)...have collaborated on groundbreaking ways to use [the NHIS] to monitor trends

and patterns, answer key research questions, and provoke further scientific inquiry into cancer prevention and control. Selected examples demonstrate how cancer control and core data have

- evolved over time to meet the needs of the cancer surveillance community;
- monitored three longstanding areas critical to national surveillance: (1) tobacco control, (2) cancer screening, and (3) some aspects of diet and nutrition;
- led to significant analytical and methodological advances.”

Numerous graphs and tables on the poster show some of the results of analyzing the supplement data. A description of questions on the 2005 supplement says that “The need for a short assessment of diet [led] to the development of a five-factor screener to assess fruits and vegetables, fiber, added sugar, calcium, and dairy food intake. Questions [were] added to monitor use of federal tobacco telephone quit lines.”

Tracking progress toward Healthy People 2010 objectives

As noted above in the response to Question 4, the NHIS is used by DHHS’ Healthy People Program to monitor Healthy People objectives. A very few of the many Healthy People 2010 objectives that are being monitored by the 2008 NHIS are as follows:

Focus Area	Objective Number	Short Title
Access to quality health services	01-01	Persons with health insurance (aged under 65 years)
Arthritis, osteoporosis, and chronic back conditions	02-02	Activity limitations due to arthritis (age adjusted, aged 18 years and over)
Cancer	03-11b	Pap tests – Received within past 3 years (age adjusted, aged 18 years and over)
Diabetes	05-03	Prevalence of diabetes (age adjusted per 1,000 standard population)
Disability and secondary conditions	06-08	Employment parity – Adults with disabilities (aged 18 to 64 years)
Physical activity and fitness	22-01	No leisure-time physical activity (age adjusted, aged 18 years and over)

Question 8: *How does NHIS see its own role? Is it to provide data for government reports or to provide data sets for the scientific community to analyze? What policy questions does it answer? Who in the government uses the data? Provide responses by use of core data or special modules.*

Response to Question 8:

For many years NCHS used as its motto "...monitoring the Nation's health..." That is an excellent, concise summary of what the NHIS is best used for. Its annual administration, consistency, and coverage, both geographic and topical, make it a valuable resource to a wide range of data users. We are frequently asked what policy questions it answers, but a more appropriate question might be *How does the NHIS measure the impact of health policy on the nation?* The NHIS can detect fluctuations in insurance coverage, access to medical care, and serious psychological distress, for example, on a quarterly or annual basis for many population subgroups.

NHIS data are widely cited in government research and by the larger scientific community. Our government collaborators include other offices in CDC and many institutes and offices at NIH. Not only have they funded the inclusion of NHIS questions annually or periodically, but they rely on the NHIS to provide the foundation for further research. Lists of agencies and the NHIS supplements that they have sponsored since 1973 are available at

<http://www.cdc.gov/nchs/about/major/nhis/co-sponsors.htm>.

DHIS also responds to requests for data from the Department level (sometimes with startup funding from ASPE) by including questions on emerging areas of interest, such as consumer-directed health plans and use of health information technology. Also, a few times a year, Congress requests statistics on specific topics for use in developing legislation. Since 2005, DHIS has recently provided NHIS population estimates to Congressional staffers on asthma, COPD, disability, health insurance, food allergies, mammography, physical activity, and smoking.

In a nutshell, we want the NHIS data to be useful, and we want the NHIS data to be used. Our role is to facilitate that.

Question 9: *Future Directions. If there were improvements in the budget what are the 2 or 3 areas that you would want to pursue to improve the NHIS in the future? If budgetary improvements do not occur, what priorities of the NHIS are most important to continue?*

Response to Question 9:

Our wish list if the budget improves:

- Restore the sample size to 100,000 persons (or more). The 2007 sample size was less than 76,000. With an even larger sample than 100,000, we could, for example, produce more and more precise state estimates and estimates for other subgroups.
- Implement a new sample design with more flexibility, so that the size and cost of the survey could be responsive to fluctuations in funding, and so that the design could be updated more often than every 10 years.
- Increase the time and ability of NHIS staff to conduct more in-depth analysis of NHIS data.
- Conduct more methodological studies using the NHIS (such as the study of the effectiveness of asking only for the last 4 digits of people's Social Security numbers).
- Develop an NHIS analysis capability for the NHIS Website, with proper variance estimation. (NCHS' Office of Information Technology has plans for a project that

would provide this capability in a centralized manner for virtually all NCHS data systems.)

Most important to keep without budget improvements:

- Keep the essential survey infrastructure, so that when the budget improves, the survey can be restored with the least possible disruption and cost. That includes DHIS' valued, experienced staff.
- Keep the rich NHIS multivariate data (demographic, SES, health status, etc.).
- Keep continuity of data over time.
- Continue innovation that adds relatively little cost, such as the new semi-annual report on telephone usage and the new annual release of paradata.

Appendix A. NHIS microdata release and access policy

This policy addresses when, to whom, and in what form the Division of Health Interview Statistics (DHIS) disseminates National Health Interview Survey (NHIS) microdata, and this policy also outlines dissemination procedures. The term *microdata* refers to a data file containing information collected as part of the NHIS in which each record provides information at the unit level of data collection (e.g., individual persons, households, or events). The term *dissemination* refers to any mechanism by which microdata are made available to users.

This DHIS policy is consistent with policies of the Centers for Disease Control and Prevention (CDC) and the National Center for Health Statistics (NCHS), including the guiding principle of making high quality data available

- as widely as practicable
- as soon as possible after data collection
- in as much detail as possible

while maintaining survey participant confidentiality.

The CDC/ATSDR policy on releasing and sharing data may be accessed at:

www.cdc.gov/od/foia/policies/sharing.htm;

the NCHS data release policy may be accessed at www.cdc.gov/nchs/about/policy/reldata.htm.

NHIS Microdata Release Policy

Public use microdata files are available to anyone. NHIS public use microdata releases are made available on an annual basis, with the primary public use file release scheduled no later than December of the year following data collection [and lately in the following June]. Due to the voluminous nature of NHIS data and the large amount of post-collection data processing, release of all microdata may not occur at one point in time. It is estimated that microdata requiring special coding or processing will be released within six months following release of the primary public use file. Data requiring complex imputation, such as income data, will be released on an ad hoc basis, since they are subject to manpower limitations. Data release will occur after data cleaning, editing, documentation and Disclosure Review Board (DRB) clearance have been completed. Public use microdata releases from 1997-2002 NHIS are available either from the NHIS Internet site (www.cdc.gov/nchs/nhis.htm) or on ASCII CD (E-mail: nchsquery@cdc.gov). Public use microdata 1969-1986 and for some years from 1987-1997 are available on ASCII CD.

DHIS Policy on Obtaining Access to Data that will not be Publicly Released:

Several mechanisms may be used for access to non-public microdata. In some strictly limited circumstances, microdata that are not released publicly may be provided by DHIS through a special data use agreement that provides for NHIS oversight of the use of the data. No data that cannot be or have not been publicly released will be made available outside NCHS without a data user's signed written agreement to provide such safeguards as are necessary. The agreement must be countersigned by the Director, DHIS, or designee and the Director, NCHS, or designee. The circumstances of such agreements are limited by 1) whether the respondents' informed consent for the NHIS provided for the data to be used by the recipient; 2) the need for such data (i.e., that the user could not accomplish the analysis with more generally available releases of the data); and 3) the ability of the recipient to provide adequate safeguards as defined by DHIS.

Each mechanism has specific access criteria and associated procedures. All users of data, irrespective of the mechanism, must complete a confidentiality non-disclosure form [attachment not provided in the present report] and see the NCHS Confidentiality Training Video.

1. NHIS QA/QC Collaborator agreements:

- a. Data recipients: Current NHIS Collaborators, as outlined in collaborator agreement [attachment not provided in the present report].
- b. When access is granted: at any stage post-collection, although normally at Beta testing stage, typically three months prior to public release. These files may contain weighted or unweighted data, and may be a subset of the NHIS file.
- c. Agreement process: Except in very unusual cases initiated by the NHIS program, this request is made during the survey component planning or funding process, and consists of a signed inter-agency collaborator QA/QC dataset agreement [attachment not provided in the present report] noting restrictions on QA/QC dataset sharing, analysis and publication prior to public release.

2. NHIS Data Support Agreements:

- a. Data recipients: Identified experts under signed agreement [attachment not provided in the present report] to assist in data collection or processing.
- b. When access is granted: During survey planning, data collection or editing.
- c. Agreement process: These agreements are initiated by the NHIS program.
- d. NHIS Special Use Data Agreements: Under special circumstances NCHS enters into an agreement [attachment not provided in the present report] to provide a limited non-public special dataset that cannot be released publicly. The user must agree to all conditions of data use for this data. These files may contain weighted or unweighted data, and may be a subset of the file.
- e. Data recipients:
 - i. Current NHIS collaborators
 - ii. CDC employees, contractors or consultants on-site at NCHS – e.g., non-DHIS employees
 - iii. Any researcher: data that have been approved for release by the Disclosure Review Board under a special use agreement
- f. When access is granted: Ongoing. Timing of access may be limited by manpower resources to prepare a dataset; agreements to access the data cannot exceed a one-year period.
- g. Agreement process: Requestors must submit a request for a Special Use Data Agreement to the Director, DHIS. These agreements shall be reviewed for approval by the Director, DHIS or his/her designate, the NCHS Confidentiality Officer, and the NCHS Director or his/her designate.

3. NCHS Research Data Center applications: Examples of applications include requests to match NHIS data to external data sources; to analyze lower-level geography or indirect identifiers; for access to non-public-release data that are the basis of published analyses of NHIS data (e.g., published analyses based detailed race/ethnic or design data).

- a. Data recipients: Any researcher
- b. When access is granted: Ongoing. RDC proposals involving NHIS data are first reviewed and approved by the Director, DHIS or his/her designee and then reviewed by

the RDC and the NCHS Confidentiality Officer. Timing of access may be limited by manpower resources to prepare a dataset.

- c. Process: Submission of a research proposal to the RDC (information available at <http://www.cdc.gov/nchs/r&d/rdc.htm>).
4. NCHS Designated Agent Authority: Interim Policy for Implementing Designated Agent Authority at the National Center for Health Statistics

The National Center for Health Statistics has long promoted the widest possible dissemination of data for research and analysis. Access to detailed data, however, is restricted through legislation designed to protect the confidentiality of NCHS data. The original NCHS confidentiality legislation, Section 308(d) of the Public Health Service Act, provides the basic legal requirement for protecting the Center's records. The newer Confidential Information Protection and Statistical Efficiency Act of 2002 (CIPSEA) provides similar protection for survey respondents with stronger enforcement through the use of fines and penalties. However, CIPSEA also provides new avenues of data access if certain requirements are met and responsibility accepted by researchers. This new authority allows for bringing non-employee researchers "inside" the agency umbrella through designated agent authority to allow access (under certain specific conditions) to data not otherwise releasable to the public.

The Interim Policy for Implementing Designated Agent Authority at the National Center for Health Statistics [attachment not provided in the present report] outlines the process by which, under CIPSEA, researchers outside of NCHS can apply to become designated agents and access NCHS non-public-use data, under carefully-controlled conditions, for statistical purposes only. Data from both NCHS-initiated surveys and from Vital Statistics records obtained from States are covered by the legislation. Researchers seeking access to NHIS microdata should review the policy, paying particular attention to:

- a. The set of standards for research that all potential researchers must meet.
- b. The categories of designated agents permissible by the policy.
- c. The delineation of the level of disclosure risk of the data required for the research.
- d. Description of the initial implementation steps.
- e. Description of the process for requesting designated agent status and NCHS Approval Process.

Appendix B. Division of Health Interview Statistics policy on special release of estimates derived from unreleased National Health Interview Survey microdata

The Division of Health Interview Statistics (DHIS), National Center for Health Statistics (NCHS), releases estimates derived from unreleased National Health Interview Survey (NHIS) microdata under strictly limited circumstances. The policy regarding release of such estimates by DHIS to others is described below. Estimates covered by this policy include both estimates derived from microdata that have *not yet* been released as well as estimates from microdata that *will not* be released for reasons of confidentiality. This policy does not cover microdata files.

DHIS Policy

The special release of estimates derived from unreleased microdata is limited to the following purposes: (a) quality assurance activities; (b) urgent policy- or legislative-related requests; (c) important emerging DHIS issues; and (d) MMWR publication of national CDC program reports that are both time sensitive and of high value with respect to public health. MMWR publications covered by this policy include: (1) Recommendations of the Advisory Committee on Immunization Practices; (2) National Influenza Vaccination Awareness Week; and (3) Tobacco Use among U.S. Adults.

Requests for special release of estimates should be addressed, in writing, to the Director of the Division of Health Interview Statistics. Requests should include justification of the need for these estimates as well as sufficient details to assess the time and level of effort required by DHIS staff to produce them. Approval of requests meeting the special release criteria is at the sole discretion of the DHIS Director. SAS programs for generating the estimates may be required from the requestor if approval of the request is granted.

Special release of estimates derived from unreleased microdata also requires review and clearance by the DHIS Director or his/her designee to insure nondisclosure of confidential information. Estimates are to be used only by the person(s) identified in the original request and only for the purpose identified in the request.

Also, separate clearance by the DHIS Director or his/her designee must be obtained prior to publishing or otherwise releasing the estimates or results of further research based on them. The source of the microdata must be identified with any products using the estimates.

Appendix C. Release of state identifiers with National Health Interview Survey microdata

Lead authors: Susan Jack, NCHS Division of Health Interview Statistics; and Van Parsons, NCHS Office of Research and Methodology

This report summarizes and reviews issues, risks, pros, cons, and opinions about releasing state identifiers with National Health Interview Survey data.

Users of the NHIS public-use microdata files frequently express a need for the inclusion of geographical locations on the public microdata. The current exclusion from public use microdata files of geographical identifiers below the regional level is based on NCHS' policies on confidentiality/disclosure risks and required statistical accuracy/reliability levels. In this report we focus upon factors that must be considered before the releasing state-level identifiers. While focusing on a state unit, much of the discussion is also applicable to other geographical units, e.g., metro areas and counties. This report is based on discussions with staff in the Division of Health Interview Statistics (DHIS) and the Office of Research and Methodology (ORM), and members of the NCHS Disclosure Review Board (DRB). A technical discussion of the aspects of protecting confidentiality and NHIS survey structures is well beyond the scope of this report, and the content that follows should only be considered as a broad overview.

I. Current and Past NHIS Public Release Policy and Methods

A. Geographical Release

NHIS is a Title 15 survey, which means that NCHS agrees not to permit re-disclosure of the NHIS data in a manner that would permit identification of the data of an individual respondent. Advances in computer hardware and algorithms and the availability of auxiliary external databases have made prevention of respondent disclosure more difficult than in the past. For the current 2006-2014 NHIS design, the NCHS policy is to release geographical identifiers no finer than the Census regional level (Northeast, South, Midwest and West). Previously, until 2001, the NHIS sample was also identified by MSA size, and some of the largest MSA regions were identified by name. For the 1985-1994 NHIS, state identifiers were also released on special request.

B. Geographical Disclosure Risk

The true NHIS sample is geographically clustered at the state, county and Census block levels. Furthermore, to reduce survey costs, the same sets of clusters tend to be in sample over the life of the survey design. Even without explicit cluster identification, the socio-economic-demographic (SED) data within a cluster, possibly accumulated over time, may in fact explicitly identify the cluster, thus putting the sample respondents in that cluster at risk of identification. However, the clustering must be preserved in the data in some form that still represents the true complex survey design structure, and thus allowing analysts to use appropriate analytical software, e.g., SUDAAN.

C. Methods used to Mask Variables

Since 1997, NCHS has been releasing sets of masked pseudo-Strata and pseudo-PSUs that are based in part on a method that blurs geography. Using a probabilistic method, original clusters, possibly in different geographical areas, are collapsed into larger clusters, thus reducing information about the true geographical locations of the respondents. For national-level domains, this method produces standard

errors that are of similar magnitude to the standard errors produced by original design structures. These pseudo-strata and pseudo-PSUs are consistently defined over the life of the survey design, which allows correct data clustering for analyses of combined years of NHIS data.

All other NHIS variables are assessed for disclosure risk by the NCHS Disclosure Review Board, and appropriate coarsening or deletion action is taken when necessary.

II. Specific Issues in Releasing State Identifiers on NHIS Microdata

The Title 15 authority of the NHIS allows NCHS some discretion in defining data release policy. If NCHS does decide to release NHIS state identifiers at the microdata level, then each state's microdata can be thought of as a mini-NHIS, but in a somewhat incompletely processed form with respect to confidentiality safeguards, the weights, providing an efficient design structure, and providing user-support. Such a release would not be considered a responsible action by NCHS unless the problematic issues described below were resolved.

A. Issues: Inclusion of State Identifiers Increases Disclosure Risk

1. Existing variables may be defined too finely within a state

As knowledge of a respondent's location of possible residency increases, the risk of disclosure increases. Reported variable levels like race/ethnicity, rare conditions, and unusual family structures may be considered as having little disclosure risk at the national level, but when these same variables are restricted to a specific state, available auxiliary state information may put the respondent at risk. For example, race/ethnicity status of "American Indian Alaska Native" and family structure of triplets as children are now released with the current NHIS, but such variable levels might pose unacceptable risk at the state level.

2. Older databases may incur greater disclosure risks

Prior to 1997, NHIS microdata had little masking of the sampled clusters other than nondescript labels, and the respondent data were subject to less masking, e.g., ICD codes could express rare conditions. Prior to 2002, large MSAs were explicitly named, and variables distinguishing MSA size were available. Any post-addition of state identifiers could lead to finer-level geographical identification and possibly to identification of individuals with rare conditions in such areas. Furthermore, the consistency of definition of public-use design variables—e.g., strata and PSUs—over the life of the survey design allows inadvertent geographical disclosure for any one year to be extended over past and future years.

The oldest NHIS databases, say prior to 1995, are not linkable to later NHIS designs. The risks and consequences of respondent identification for older data sets have not been studied.

3. State identifiers may compromise the protection from the blurring of geography.

Many of the public-use pseudo-PSUs contain data from more than one state (see I.C). In particular, this is true of 4 of the 5 largest states. Most likely, only "larger" states would be considered for explicit identification, and for these states some public-use PSUs will be decomposable into an identified state and also into an unidentified smaller area. This unidentified area may have been

originally combined outside of its true state with the sole purpose of avoiding disclosure. This defeats the original intention of geographical masking. This is an issue for all NHIS data from 1997 to the present.

4. The MEPS-NHIS linkage must be assessed for additional disclosure risks

Since 1995 the Medical Expenditure Panel Survey, MEPS, has been linked at the household level with the NHIS. (See page 15 in the April 25, 2008 BSC report.) Public-use data linkages between both surveys have been available. State identification on the NHIS may possibly compromise the MEPS disclosure protections, and the additional information linked from MEPS may now also compromise the NHIS.

B. Issues: Methodologies for State Estimation

It is assumed that most analyses of state data will use the traditional NHIS design-based approach of weighted estimates and standard errors computed using a supplied design structure with complex-survey analysis software like SUDAAN. State design information is needed.

1. State weights

The existing survey weights are adjusted to national-level controls, but state-based analyses may benefit from a creation of state-based controls. If NCHS provided state weights, it would need a customized operation for each state released. Tasks include defining appropriate weighting classes by state and obtaining accurate controls from a source on a periodic schedule. These operations might require substantial NCHS resources.

2. State design

The pseudo strata and PSUs on the public use microdata files are constructed with national estimates in mind. These national strata and PSUs, when applied at the state level, may result in inefficient variance estimation, i.e., variance estimators may be more unstable than those variances produced by the true design. Efficient variance estimation at the state level may require a deliberate un-blurring of geography and the use of true design clusters. Most likely, this finer information could not be released.

3. States having small samples

While a state's annual sample may be adequate for reliable estimation over the entire state, most state-based analyses focus on domains within the state. Frequently, these targeted domains become very thin for an annual sample, and analysis requires accumulation of years of data. Furthermore, since the NHIS sample is highly clustered, accumulating data over time may not improve stability of the variance estimators. Overlooking the confidentiality concerns, most likely only about 10 states have the sample and design structures amenable to producing design-based state variance estimation.

4. Model-based methods

While analysts who use traditional design-based methods are somewhat constrained by the quality of the weights, design information and sample sizes, many analysts use model-based and model-assisted methods to enhance the design and also to "borrow strength" across all the available data. The

restrictions just mentioned in II.B.1-3 above are not as critical to model-based approaches to survey data analysis.

III. Possible Ways to Release State Identifiers: Pros and Cons

Currently, microdata with state identifiers are only available through the highly restrictive NCHS Research Data Center (RDC). Some suggestions follow for public release of state microdata, along with pros and cons motivated by the material in sections I and II.

A. Release *Linkable* State Microdata Files

Conceptually, a set of state identifiers (possibly limited in scope) that can be merged onto any of the annual data bases is released. There are numerous restrictions that can be attached to such a release:

1. Make state identifiers available to the public without user restrictions.
2. Make state identifiers available to only trusted clients and with some written agreement. This arrangement does not require an RDC worksite.
3. Supply identifiers for only large states for which the additional disclosure risk discussed in II.A is low and that have “large enough” annual samples on several socio-economic-demographic domains for reliable estimators (See II.B).
4. Supply a larger number of states, but only under restriction 2 above.
5. Supply no additional design information, as discussed in II.B.
6. Supply additional design information, as discussed in II.B.

Pros:

NCHS would be responsive to public-data needs. Data under release restriction 2 could greatly reduce any additional public disclosure risk factor. If efficient state designs under release restriction 6 are produced, there could be some standardization of proposed state methods.

Cons:

All the release strategies impose additional disclosure risks to the NHIS data beyond the current release policies. Disclosure problems II.A.2 and II.A.3 are troublesome, especially under release restriction 1. A thorough study of the issues of II would require major NCHS resources.

Any partial release of state microdata may result in stronger pressure from those states whose data are not released. A data release under restriction 5 may result in unsound data analyses.

B. Release *Non-linkable* State Microdata Files to the Public

Here, original data may be further coarsened, noise may be added, data subsampled, or synthetic data added. Using these masking procedures would mitigate the state inclusion disclosure risks discussed in II A. These procedures would be performed in such a way as to preserve analytic capabilities somewhat consistent with original unperturbed NHIS microdata.

Pro:

Such databases may be successful for simple descriptive-types of analyses involving usual socio-economic-demographic domains. Such databases would allow analysts an idea of state analytic potential before committing resources to use the real data at the RDC.

Cons:

Users of these data may question the quality of its statistical inferences.

Substantial costs to NCHS in determining the specific risks discussed in II.A and in determining appropriate masking procedures that retain data information. Any provided design information (see II.B) would also need careful consideration to avoid linkage, but still capture the features of the complex survey.