Vital and Health Statistics

Design and Estimation for the National Health Interview Survey, 1985-94

Series 2:
Data Evaluation and Methods Research
No. 110

A description of the redesign of the National Health Interview Survey, a national probability sample survey of the civilian noninstitutionalized population of the United States.

U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES
Public Health Service
Centers for Disease Control
National Center for Health Statistics

Hyattsville, Maryland
August 1989
DHHS Publication No. (PHS) 89-1384

Copyright information

Suggested citation

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

Library of Congress Cataloging-in-Publication Data

Design and estimation for the National Health Interview Survey. 1985-94.

(Series 2, Data evaluation and methods research ; no. 110) (DHHS publication ; no. (PHS) 89-1384)
By James T. Massey and others.
Bibliography: p.
Supt. of Docs. no.: HE 20.6209:2/110
1. National Health Interview Survey (U.S.)
2. Health surveys-United States. I. Massey, James T.
II. National Center for Health Statistics (U.S.)
III. Series: Vital and health statistics. Series 2,
Data evaluation and methods research ; no. 110.
IV. Series: DHHS publication ; no. (PHS) 89-1384.
[DNLM: 1. National Health Interview Survey (U.S.)
2. Data Collection. 3. Health Surveys-United States.
4. Research Design. W2 A N148vb no. 110]
RA409.U45 no. 110 362.1'0723 s 88-600448
[RA408.5] [362.1'0723]
ISBN 0-8406-0406-8

National Center for Health Statistics

Manning Feinleib, M.D., Dr.P.H., Director

Robert A. Israel, ***Deputy Director***

Jacob J. Feldman, Ph.D., ***Associate Director for Analysis and Epidemiology***

Gail F. Fisher, Ph.D., ***Associate Director for Planning and Extramural Programs***

Peter L. Hurley, ***Associate Director for Vital and Health Statistics Systems***

Stephen E. Nieberding, ***Associate Director for Management***

Charles J. Rothwell, ***Associate Director for Data Processing and Services***

Monroe G. Sirken, Ph.D., ***Associate Director for Research and Methodology***

Office of Research and Methodology

Monroe G. Sirken, Ph.D., ***Associate Director***

Kenneth W. Harris, ***Special Assistant for Program Coordination and Statistical Standards***

Lester R. Curtin, Ph.D., ***Chief; Statistical Methods Staff***

James T. Massey, Ph.D., ***Chief Survey Design Staff***

Andrew A. White, Ph.D., ***Chief; Statistical Technology Staff***

Design and Estimation for the National Health Interview Survey, 1985-94

Preface

This report presents a detailed description of the research and selection of a sample design for the National Health Interview Survey (NHIS) for the period 1985 through 1994. The NHIS is one of the major surveys of the National Center for Health Statistics (NCHS), Centers for Disease Control. Through the NHIS, information concerning the health of the U.S. civilian noninstitutionalized population is collected in household interviews throughout the United States. The NHIS has been in continuous operation since 1957, and its sample design has been reevaluated and modified following each of the last three decennial censuses of the U.S. population.

The overall responsibility for the development of the redesign following the 1980 decennial census was carried out by the Survey Design Staff in the Office of Research and Methodology (ORM) at NCHS. The Statistical Methods Division of the U.S. Bureau of the Census had the major responsibility for implementing the design and selecting the sample for the NHIS once the general design specifications were defined. The Statistical Methods Staff in ORM had the responsibility for developing the estimation procedures for the NHIS. The data requirements for the NHIS for the period 1985-94 were jointly developed by the Division of Health Interview Statistics, the Office of Analysis and Epidemiology, and the Office of Research and Methodology at NCHS. The total redesign effort was conducted from 1979 through 1984.

Redesigning a survey as complex as the National Health Interview Survey was a major undertaking and many persons contributed to this effort. Some of the persons who contributed to the redesign are cited in the reports and memoranda referenced at the end of the report. A number of persons deserve a special thanks for their contribution. At the National Center for Health Statistics the authors of this report would like to acknowledge the contributions of Monroe Sirken, Owen Thornberry, Jack Feldman, Earl Bryant, Robert Fuchsberg, Clinton Burnbaum, Robert Casady, and Andrew White. At the U.S. Bureau of the Census, persons deserving special recognition include Charles D. Jones, Gary Shapiro, Rameswar P. Chakrabarty, Chester Ponikowski, Paul Hsen, and Richard Hartgen.

This report is organized into four chapters so that researchers and users of the NHIS data can refer to different parts of the report for different levels of detail about the NHIS design. Chapter I presents a general overview of the NHIS and its sample design. Chapter II describes the redesign process and the new directions to be taken by the NHIS during 198.5-94. Major research findings are also included in chapter II. Chapter III provides a more detailed description of the sample design and how the sample was selected. Chapter IV presents a description of the estimators used in the NHIS for analyzing and summarizing the results from the survey.

Contents

Symbols

- - -	Data not available
. . .	Category not applicable
	Quantity zero
0.0	Quantity more than zero but less than 0.05
Z	Quantity more than zero but less than 500 where numbers are rounded to thousands
*	Figure does not meet standard of reliability or precision (more than 30-percent relative standard error)
#	Figure suppressed to comply with confidentiality requirements

Chapter I. Overview of the National Health Interview Survey and its sample design

by James T. Massey, Ph.D., Office of Research and Methodology, National Center for Health Statistics

Description of the National Health Interview Survey

One of the oldest national health surveys, the National Health Interview Survey (NHIS) has been in continuous operation since July 1957. As one of the major data collection programs of the National Center for Health Statistics (NCHS), the NHIS is a principal source of information on the health of the civilian noninstitutionalized population of the United States.

The survey covers the civilian noninstitutionalized population of the United States living at the time of the interview. Because of technical and logistical problems, several segments of the population are not included in the sample or in the estimates from the survey. Persons excluded are patients in long-term care facilities, persons on active duty with the U.S. Armed Forces (although their dependents are included), and U.S. nationals living in foreign countries.

The NHIS data are obtained through personal interviews with household members. Interviews are conducted each week throughout the year in a probability sample of households. The interviewing is performed by a permanent staff of interviewers employed by the U.S. Bureau of the Census. Data collected over the period of a year form the basis for the development of annual estimates of the health characteristics of the population and for the analysis of trends in those characteristics.

All adult members of the household 17 years of age and over who are at home at the time of the interview are invited to participate and to respond for themselves. For children and for adults not at home during the interview, information is provided by a responsible adult (19 years of age and over) residing in the household. Between 65 and 70 percent of the adults 17 years of age and over are self-respondents. Generally, a random subsample of adult household members is selected to respond for themselves to additional questions on current health topics that vary from year to year.

Depending on the household size and the nature and extent of health conditions of household members, the length of interview ranges between 20 and 90 minutes. On average, the interviews require about 50 minutes in the household.

The questionnaire consists of two basic parts: (1) a set of basic health and demographic items, and (2) one or more sets of questions on current health topics. The basic items constitute approximately 50 percent of the questionnaire and are repeated each year. These items provide continuous information on basic health variables. Questions on current health topics facilitate a response to changing needs for data and coverage of a wide variety of issues. This combination yields a unique national health data base.

The questionnaire includes the following types of basic health and demographic questions:

- Demographic characteristics of household members, including age, sex, race, education, and family income.
- Disability days, including restricted-activity and bed-disability days, and work- and school-loss days occurring during the 2-week period prior to the week of interview, as well as bed days during the last 12 months.
- Physician visits occurring during the same 2-week period, the interval since the last physician visit, and the number of visits during the last 12 months.
- Acute and chronic conditions responsible for these days and visits.
- Long-term limitation of activity resulting from chronic disease or impairment and the chronic conditions associated with the disability.
- Short-stay hospitalization data, including the number of hospital episodes during the past year and the number of days for each stay.

In addition, each of six representative subsamples of households is asked to respond to questions about one of six lists of selected chronic conditions. These six lists are a part of the basic NHIS.

Questions on special health topics change from year to year in response to current interest and need for data. The 1983 questionnaire contained questions on alcohol, dental care, physician services, and health insurance. The health topic questionnaire in 1984 was devoted entirely to issues of aging; in 1985 it covered health promotion and disease prevention; and in 1986 it included questions on health insurance, dental health, vitamin and mineral intake, longest job worked, and functional limitations. The 1987 NHIS included an extensive questionnaire on cancer risk factors and questions on child adoption.

There are five different reference periods used for the basic NHIS items--a 2-week period, a 12-month period, a 13-month period, "now," and "ever." Other reference periods are sometimes used for the special health topics. The reference period used for a particular item depends on the difficulty of reporting the item and the frequency of occurrence of the item.

The households selected for interview each week are a probability sample representative of the target population. Data are collected from approximately 49,000 households including about 132,000 persons in a calendar year. Participation is voluntary; confidentiality of responses is guaranteed. The annual response rate of NHIS is over 95 percent of the eligible households in the sample. The nonresponse is divided equally between refusals and households where no eligible respondent could be found at home after repeated calls.

Statistics from the NHIS are published in NCHS reports and in professional journals. A major strength of the NHIS is its ability to display the basic health characteristics by many demographic and socioeconomic characteristics. Most of the descriptive statistics published yearly appear in NCHS *Vital and Health Statistics* Series 10 reports. The survey procedures and a description of the statistics collected are described in two NCHS reports (NCHS, 1975, 1985).

NHIS sample design

From the beginning of the NHIS the U.S. Bureau of the Census has had a primary responsibility for the sample design and data collection for the NHIS. Historically, the NHIS sample design has been linked with the sample design of the Current Population Survey (CPS), a labor force survey also conducted by the U.S. Bureau of the Census. Prior to 1985, the sampling frame for the CPS was used to select a smaller sample for the NHIS. Although the primary locations for the NHIS were a subset of the primary locations for the CPS sample, the surveys were always conducted in different households within the primary locations. Descriptions of the sample designs have been published for the NHIS (NCHS, 1958, 1974, 1985) and the CPS (U.S. Bureau of the Census, 1963, 1978).

The sample design for the NHIS has been revised following each decennial census of the population. The revisions following the 1960 and 1970 censuses of the population were made primarily to update the sampling frame and to improve the efficiency of the design by using more current information. The redesign following the 1980 decennial census of the population was a more comprehensive redesign than the previous redesigns. For this redesign, implemented in 1985, a thorough evaluation was made of all of the basic design features of the NHIS and a number of new features were also investigated.

Conceptually, the sampling plan for the NHIS has remained the same since 1957 and follows a stratified multistage probability design which permits a continuous sampling of the civilian noninstitutionalized population of the United States. The sample is designed in such a way that the sample of households interviewed each week is representative of the target population, and that weekly samples are additive over time. The yearly samples represent a cross-sectional sample of the population and the weekly subsamples are selected from a yearly sample without replacement. For the NHIS a single interview is conducted in each household. The additive feature of the design permits both continuous measurement of characteristics of samples and more detailed analysis of less common characteristics and smaller categories of health related items.

The overall sample is designed so that statistics can be provided for each of the four major geographic regions of the United States and for selected places of residences within the United States.

The basic characteristics that define a sample design, referred to as design parameters, are shown in figure 1 for the current (1985-94) and the preceding (1973-84) NHIS designs. The parameters for the 1973-84 design are similar to the parameters for all of the previous NHIS designs. A number of factors are considered in the selection of parameters for a sample design. Some of the most important factors include the analytic objectives and data requirements for the survey; the reliability of the estimates produced by a design for a given set of parameters; the cost of the survey, including administrative and travel costs; and the size of interviewer assignments. For a survey such as the NHIS, which has a number of analytical objectives, it is not an easy task to determine the best set of design parameters. The best set of parameters for one survey objective may not be the best set of parameters for another survey objective. The final set of parameters selected defines a design that comes closest to satisfying the total set of survey objectives and operational requirements.

The first stage of the NHIS sample design consists of selecting a sample of primary sampling units (PSU's) from approximately 1,900 geographically defined PSU's. A PSU consists of a county, a small group of contiguous counties, or a metropolitan statistical area (MSA). The PSU's collectively cover the 50 States and the District of Columbia. The 1,900 PSU's are stratified (grouped) using socioeconomic and demographic variables and then selected with a probability proportional to their population size (pps) within a stratum. Prior to 1985 the number of PSU's selected for the NHIS sample ranged from 356 to 376. In 1985 the number of PSU's was reduced to 198. Because the PSU's are selected pps, the largest PSU's in the United States are selected into the sample with certainty (probability of 1) and are referred to as self-representing (SR) PSU's. The PSU's that are not selected with certainty are referred to as non-self-representing (NSR) PSU's. The stratification and selection of the PSU's is described in chapter III. The dramatic reduction in the number of PSU's for the NHIS design is primarily a reflection of the increase in interviewer and travel costs since 1973. In order to maintain the same

Design parameter	1985-94 design	1973-84 design
Sampling frame	Geographical areas Building permits	List of addresses Geographical areas Building permits
Primary sampling units (PSU's)	1983 MSA's[1], single counties, contiguous counties	1970 SMSA's[2], single counties, contiguous counties
Total	198	376
Self-representing (SR)	52	156
Non-self-representing (NSR)	146	220
Cluster size (housing units)		
List of addresses	NA	4 compact
Geographical areas	8 noncompact	4 noncompact
Building permits	4 noncompact	4 noncompact
Sample size (yearly)	61,400 occupied and unoccupied housing units	51,000 occupied and unoccupied housing units
Expected interviewed households (yearly)	49,000	40,000
Expected number of persons in survey	132,000	108,000
PSU's per NSR stratum	2	1
National subsamples of PSU's (panels)	4	Not defined
Subdomain and minority estimation	Poststratification by age-sex-race; PSU stratification for Hispanics; oversample for black population within PSU's	Poststratification by age-sex-race
Number of inter-viewers	140	120

[1]Metropolitan statistical area as defined by U.S. Office of Management and Budget.
[2]Standard metropolitan statistical area as defined by U.S. Office of Management and Budget.

Figure 1. Sample design parameters for 1973-84 and 1986-94 National Health Interview Survey designs

overall level of precision for NHIS statistics, the increase in interviewer costs required a sample of fewer PSU's with more households per PSU.

The distribution of the NHIS PSU's across the United States (based on the 1980 decennial census) is shown in figure 2. Only 79 of the 198 PSU's were not included in the 1970-based sample used for the NHIS from 1973 through 1984. The remaining PSU's were either completely or partially contained in the 1970-based sample design. Approximately 76 percent of the 1980-based sample was selected from geographical areas selected for the 1970-based sample.

Within PSU's, an iterative program is used to select a sample of geographically defined area segments. The sample segments are systematically selected so they are distributed throughout each of the PSU's. The segments are subdivided into clusters which contain a small number of housing units. The housing units may be grouped closely together (compact) or spread over a small geographical area (noncompact). Historically the cluster size for the NHIS has ranged from four to nine households. For the 1985-94 design the expected cluster size is eight housing units for the geographically defined area clusters and four housing units for building permit clusters of housing units constructed after the 1980 census (see chapter III for a more detailed description). Basic demographic and health information is collected on all eligible persons in the occupied housing units within the sample clusters. The average number of households interviewed each year within a PSU is approximately 500 for self-representing PSU's and 156 for non-self-representing PSU's.

The total yearly number of interviewed households for the 1985-94 design is expected to be approximately 49,000. This represents approximately a 20-percent increase in the sample size from the 1973-84 design. By traveling to fewer PSU's the interviewers can collect information in more households for approximately the same cost. The design research conducted by the U.S. Bureau of the Census indicated that both the data collection costs and the sampling precision for the NHIS should be similar for the 1973-84 and 1985-94 designs.

Several of the design parameters for the NHIS were modified beginning in 1985. The use of a list sampling frame constructed from address information collected in the decennial census was dropped for the new design in order for NCHS to be able to obtain identifiable household information for followback surveys. Information from the 1980 census is confidential and could not be provided to NCHS. An area sampling frame was used to replace the list sampling frame. The number of non-self-representing (NSR) PSU's selected per stratum was changed from one to two. This change was made to more accurately estimate the precision of the NHIS statistics. The third change that was made for the new design was the formation of national subsamples of PSU's. Four subsamples of PSU's, referred to as panels, were created to provide greater design flexibility to reduce the NHIS sample size from year to year if necessary, to link other NCHS surveys to some or all of the NHIS PSU's, and to be able to easily subsample the NHIS households for possible subsequent surveys. Each of the four panels is a representative sample of the United States. The final and most significant change made in the new design involved the oversampling of the black population. In a sample that is selected strictly proportionally to population size, the number of black persons is often not adequate for producing detailed demographic subdomain statistics for the black population or for making comparisons with the white population. In the new design, the black population was oversampled to improve NHIS's ability to produce more precise statistics. The oversampling was accomplished by sampling the more predominantly black geographic areas within PSU's at a higher rate than other areas. The number of black persons surveyed each year is expected to increase by approximately 40 percent over that in the old design. A similar procedure was evaluated for the Hispanic population, but only a very small improvement in precision could be made without significantly increasing the cost of the NHIS.

The panels of PSU's described above were used to reduce the NHIS sample in 1985 and 1986. The 1985 NHIS sample included three of the four panels and the 1986 NHIS sample included only two of the four panels due to budget reductions. In 1987 the full four-panel design was used.

References

National Center for Health Statistics. 1958. The statistical design of the Health Household-Interview Survey. By staff of the U.S. National Health Survey and the Bureau of the Census. *Health Statistics*. PHS Pub. No. 584-A2. Public Health Service. Washington: U.S. Government Printing Office.

National Center for Health Statistics, J. A. Bean. 1974. Estimate and sampling variance in the Health Interview Survey. *Vital and Health Statistics.* Series 2, No. 38. DHEW Pub. No. (HRA) 74-1288. Health Resources Administration. Washington: U.S. Government Printing Office.

National Center for Health Statistics. 1975. *Health Interview Survey Procedures, 1957-1974.* DHEW Pub. No. (HRA) 75-1311. Health Resources Administration. Washington: U.S. Government Printing Office.

National Center for Health Statistics, M. G. Kovar and G. S. Poe. 1985. The National Health Interview Survey design, 1973-84, and procedures, 1975-83. *Vital and Health Statistics.* Series 1, No. 18. DHHS Pub. No. (PHS) 85-1320. Public Health Service. Washington: U.S. Government Printing Office.

U.S. Bureau of the Census. 1963. The Current Population Survey, a report on methodology. *Technical Paper No. 7*. Washington: U.S. Government Printing Office.

U.S. Bureau of the Census. 1978. The Current Population Survey, design and methodology. *Technical Paper No. 40*. Washington: U.S. Government Printing Office.

Chapter II. Redesigning the National Health Interview Survey

by James T. Massey, Ph.D., Office of Research and Methodology, National Center for Health Statistics

Objectives

The first step in designing any survey is to define the analytical objectives for the survey. In redesigning a survey that has been in continuous operation for more than 25 years, the task of defining objectives involves the reassessment of old objectives, both implicit and explicit, defining new objectives, and merging and establishing priorities for a single set of objectives. Plewes (1980) states that "possibly the only task more challenging to the survey statistician than initially designing a large-scale socioeconomic survey is that of redesigning one." Because the redesign involves an ongoing program, many resources have already been invested in the existing design, and a shift in objectives often puts the old objectives in conflict with the new objectives.

To develop a set of objectives for the National Health Interview Survey (NHIS), the Division of Health Interview Statistics (DHIS) was asked to specify its goals and data requirements for the 1985-94 decade. An analytical committee within the National Center for Health Statistics (NCHS) was also established to provide input from other offices and divisions within NCHS. Many of the final recommendations from these two groups incorporated recommendations from an earlier Technical Consultant Panel of the National Committee on Vital and Health Statistics (NCHS, 1980). A summary of the objectives for the redesign of the NHIS is given below.

- To continue to produce descriptive statistics about the health and health-related parameters of the civilian noninstitutionalized population of the United States, and to monitor change in these variables over time. The basic variables of interest are extent and nature of illness, both acute (disability days, work loss, and so forth) and chronic (limitation of activity); extent and nature of disability (acute and chronic); incidence and prevalence of acute and chronic morbidity; utilization of health-care services; health-care expenditures; utilization of health facilities and health resources; and other health-related variables.
- To put more emphasis on measures of change in the NHIS. Research should be conducted to identify alternative sample designs that would provide more precise measures of change and not seriously reduce the precision of the descriptive statistics that are currently published by DHIS. The investigation should include an evaluation of the integration of a longitudinal component with the traditional cross-sectional design.
- To improve the precision of the NHIS statistics for special small domains, such as black and Hispanic, of the population.
- To improve the efficiency and analytical utility of NCHS's combined population survey program. Research should be conducted to investigate the coordination and linkage of the NHIS with NCHS's other population and establishment surveys.
- To investigate methods that would provide the NHIS greater flexibility to oversample subgroups of the population, oversample geographic areas of the United States, conduct follow-back surveys, and respond more rapidly in the collection and production of statistics on special health topics.
- To make the NHIS sample design more cost efficient while maintaining the same overall level of precision.
- To incorporate health and health-related variables in the implementation of the redesign of the NHIS. Two specific areas that should be considered are the stratification of the primary sampling units and the differential sampling of the population.

These broad objectives were used as the primary criteria for redesigning the NHIS.

Model used for redesign

Although the NHIS has always been a multipurpose survey which collected information on a variety of health topics, the expanded objectives for 1985-94 presented NCHS with the overwhelming task of selecting a single sample design that would satisfy all or most of these objectives. In order to translate these objectives into an operational design strategy, a model approach was used to help evaluate different design alternatives.

The model used for the redesign was first developed by Sirken and Royston (1976) to study the design effects in retrospective mortality surveys. The model was further refined and applied to the redesign of NCHS's population-based surveys.

Conceptually, the use of a model to define the design process is straightforward. The first step is to identify the major design features related to the survey design. A design feature is defined as a distinctive part of the survey design that is manipulable. Examples of design features are the questionnaire, the sample design parameters (such as the stratification variables and number of primary sampling units), the sampling frame, and the mode of data collection.

Once the design features have been enumerated, the possible design options associated with each feature are identified. One of the possible options is the null option, which implies the elimination of the design feature. Examples of the design options for the design feature "mode of data collection" are face-to-face interviews, telephone interviews, and self-administered questionnaires. In some instances a design option can be eliminated for consideration in a survey because of its known properties. In the NHIS a mail questionnaire would never be used (except possibly for a small supplemental topic) because of the complexity of the NHIS questionnaire and therefore represents an infeasible design option.

After the design options have been defined for each of the design features, design-option sets are formed. A design-option set is composed of a single design option from each of the design features. A design-option set represents a possible alternative survey design. All possible design-option sets are then enumerated. The number of design-option sets is the product of the number of design options for each feature.

Following enumeration, the design-option sets are systematically evaluated in two phases. During the first phase the design options within a design-option set are examined for compatibility with each other. If any two design options are incompatible, the design-option set is deleted. The second phase of the evaluation involves the application of a set of criteria to identify the optimum design-option set. The optimum design-option set represents the survey design to be implemented. A single criterion might be the minimum sampling variance or mean squared error for a fixed cost associated with each of the design-option sets. The variance or mean squared error might represent an average value over a number of variables of interest in the survey. Another criterion might be an evaluation of how well each of the design-option sets satisfies the major objectives of the survey with respect to cost, timeliness of results, and burden on the respondent.

In applying this model to the NHIS redesign, a number of simplifying assumptions were made, and the step-wise progression of evaluating design options and design-option sets was modified. For the NHIS five major design features were identified as best for meeting one or more of the objectives of the NHIS. The five major design features are shown in figure 1 along with two design options that were considered for each feature. This represents a simplification of the model as well as of the number of options actually considered. The first option shown for each of the design features represents one or more of the new options that were evaluated, and the second option shown for each feature represents the option used under the old design. The number of design-option sets that can be formed from these design options is 2^5 or 32 design-option sets. A brief description of the five design features and their options is given below, and research on the design features is presented in the next section.

From the inception of the NHIS in 1957 the U.S. Bureau of the Census has been the data collection agent for NCHS. Since the early 1970's a combined list-and-area frame has been used for selecting the NHIS sample. The list frame obtained from the decennial census was supplemented by an area frame in areas with incomplete or insufficient address information and by a building-permit list frame for housing units constructed since the decennial census. Under the confidentiality provisions in the authorizing legislation of the U.S. Bureau of the Census, it is not allowed to release any information collected in the decennial census to other Federal statistical agencies. It cannot, therefore, release any of the identifiable person or household information to NCHS in areas where the census list frame is used. Without the identifiable information NCHS would not be able to analyze its results in geographical detail, conduct independent follow-back surveys, use the NHIS sample as a frame for other NCHS surveys, or analytically link its information to other data sets through the classification of geographical areas. It is obvious that

Design feature	*Design options*
Sampling frame	1. Area frame based entirely on geographical areas. 2. List-and-area frame based on 1980 decennial census information (1973-84 design].
Multiple-survey linkage	1. Single integrated survey design for the National Health Interview Survey, the National Medical Care Utilization and Expenditure Survey. the National Survey of Family Growth, and possibly for the National Health and Nutrition Examination survey. 2. Independent designs for each of the Center's population surveys (1973-84 design strategy).
Mode of data collection	1. Dual mode of data collection using both face-to-face and telephone interviews. 2. Face-to-face mode of data collection (1973-84 design).
Oversampling small subdomains of the population.	1. Greater probabilities of selecting persons from small domains of special interest (differential sampling). 2. Same probability of selecting each person in the target population (1973-84 design).
Rotation of NHIS sample	1. Multiple interviews, 1 year apart, at each dwelling unit for a subsample of dwelling units. 2. Single interview at each dwelling unit (1973-84 cross-sectional design).

Figure 1. Major design features and design options considered for the redesign of the National Health Interview *Survey*

these capabilities are directly related to the future objectives of the NHIS.

One way to avoid the problems of confidentiality is to replace the list frame with an area sampling frame that is not dependent on the 1980 decennial census. The disadvantage of using an area frame is the additional cost required to construct the frame and the somewhat larger costs associated with maintaining the sampling frame over its lo-year period. The sampling frame for the NHIS represents the first design feature evaluated for the NHIS redesign.

The second design feature investigated was the use of the NHIS sample as a sampling frame for the other NCHS population surveys to provide a link between the NCHS surveys. By using the National Health Interview Survey (NHIS) sample as a sampling frame for the National Survey of Family Growth (NSFG), the National Medical Care Utilization and Expenditure Survey (NMCUES), or the National Health and Nutrition Examination Survey (NHANES), NCHS could

- link the surveys analytically and possibly avoid some duplication of data collection,
- use the NHIS information to oversample sociodemographic subdomains of the population for the linked surveys more efficiently, or
- use the NHIS information to oversample persons with major expenditures in the previous year for the NMCUES.

A possible disadvantage of using the same design for all of its population surveys is the inability to have an optimum design for each survey.

Since 1978, NCHS has invested considerable resources in the study of conducting health interviews by telephone. In the redesign the use of both face-to-face and telephone interviews was considered for the NHIS as a third new design feature. The main advantages of the telephone interview are its lower cost and the closer supervision of interviewers. Before a dual mode of data collection could be used in the NHIS, however, a number of methodological questions had to be addressed. The telephone interview issues are discussed in a paper by Massey, Marquis, and Tortora (1982). The key issues related to a dual telephone and face-to-face design for the NHIS include telephone coverage, telephone response rate, costs of a dual-frame survey, and adaptability of the NHIS interview to a telephone interview.

The 1973-84 NHIS design gave every person in the target population an approximately equal chance of being selected into the NHIS. The oversampling of persons from small subdomains of the population was considered as a way to improve the precision of the estimates for the small subdomains. The most important small domains of special interest in the NHIS are the black, the Hispanic, the aged, and the low income populations. The NHIS small sample sizes for these subdomains have often prohibited detailed demographic analysis. The key methodological issue related to this fourth design feature was the increase in the precision for small subdomain statistics versus the decrease in the precision for the total population and other subdomain statistics.

In order to improve its ability to measure health trends, NCHS asked the U.S. Bureau of the Census to evaluate the partial rotation of the NHIS sample from one year to the next as a fifth design feature. Under the 1973-84 cross-sectional design, a new sample of housing units is interviewed each week. The weekly samples are accumulated over time to produce quarterly, yearly, and multiyear estimates. Under the rotation feature, a proportion of the NHIS yearly sample of housing units would be revisited the following year. By using a proportion of the same housing units in consecutive years, estimates from one year to the next year would be more highly correlated and smaller changes could be detected from one year to the next. Households interviewed in the first year would also be available for followup topics the next year. The disadvantage of using the same housing units from one year to the next year is the loss in the precision of estimates that combine the weekly samples over a multiple-year period. For questions that are repeated in the NHIS from one year to the next, the weekly samples can be combined over long periods of time to produce detailed descriptive statistics for small subdomains of the population.

The approach used to research and select a survey design with the five design features described above deviated somewhat from the conceptual model. Evaluation of all possible design-option sets was not attempted; instead, the new proposed options for the five features were studied independently and compared with the features of the 1973-84 design. Whenever the independent research investigations yielded conclusive results, a single option was selected for a design feature. Following this step, the compatibility between the remaining design options for different features was evaluated. The final design for the NHIS for 1985-94 was the set of design options that were most compatible and that best satisfied the multiple objectives for the NHIS.

Research results for major design features

The research investigations conducted for the five major design features are described in this section. The research was conducted by NCHS staff, U.S. Bureau of the Census staff, and private contractors to NCHS.

Sampling frame

The capability to link the NCHS population surveys and to conduct longitudinal followup surveys of the NHIS respondents was believed critical to meeting NHIS long-range goals. Establishing these capabilities required the creation of a sampling frame that was independent of the decennial census. To address this issue the U.S. Bureau of the Census was asked to examine the costs of creating and maintaining an independent area sampling frame versus the costs associated with its combined list-and-area frame linked to the decennial census. At the same time NCHS conducted some investigations of the cost savings that could be achieved by using the NHIS sample as a sampling frame for other NCHS population surveys.

The sampling frame used for the 1973-84 NHIS was created by combining three different sampling frames. Approximately two-thirds of the sampling frame was constructed using address listings from the 1970 decennial census of the population (list frame), about 30 percent of the sampling frame was a geographical area frame for nonlisted areas, and the remainder of the sampling frame was constructed from building permits issued since the decennial census (new-construction frame). Returning to an area sampling frame such as the one used for the NHIS during the 1950's and 1960's would require several major additional costs.

The cost study conducted by the U.S. Bureau of the Census indicated that the extra redesign research, the office preparation of segment maps, the listing of households within area segments, and the selection of sample households would cost approximately $700,000 to $1.1 million (1980 dollars) more than comparable operations for the list frame (Jones, 1981; NCHS, 1981). This range of costs is a function of the cluster size adopted for the survey. Because NCHS anticipated that a larger cluster size (requiring fewer segments to list) would be used for an area design, the lower estimate of an additional $712,000 was assumed for the transition to an area design. A more complete discussion of cluster size is presented in chapter III. One factor contributing to the large additional cost for an area sampling frame was the reduction in cost-sharing with other agencies. Most of the major surveys conducted by the U.S. Bureau of the Census for other agencies use the same sampling frame, and many of the costs associated with the construction of the sampling frame are shared.

The second major cost increase for an area design involves the updating and maintenance of the sampling frame for the decade 1985-94. The initial cost estimates from the U.S. Bureau of the Census for the maintenance of the sampling frame were very large because of the field work required to update an area frame on a continuing basis. Also, as the area segments began to vary in size over the decade because of growth and reduction in size, the precision of NCHS statistics would be adversely affected. After further investigation both problems were solved by using a new-construction frame of building permits to update the sampling frame. The new-construction frame was originally used by the U.S. Bureau of the Census with the transition to a list frame. The older area sampling frames had been updated using new segments listings. As a new-construction frame was already in operation, it could easily be adapted to update an area sampling frame as well as a list frame. The use of the new-construction frame reduced the maintenance cost for the area sampling frame and eliminated the change in area segment sizes over time. The revised estimate for frame maintenance from 1985 through 1994 was approximately $666,000. Thus, the total additional cost of having an area frame for the NHIS would be approximately $1.4 million (1980 dollars).

To determine the potential for offsetting the additional cost of the area frame design, NCHS conducted a small-scale feasibility study of linking the NSFG and the NMCUES to the NHIS. In addition to the analytical desirability of linking the NCHS population-based surveys, it was hypothesized that a considerable amount of money could be saved under a common design for NHIS, NSFG, and NMCUES instead of independent designs for the three surveys. The feasibility study also included the ability of the NHIS sample to meet the sample size requirements for the other two surveys.

In evaluating the adequacy of the NHIS sample as a sampling frame for the NSFG and the NMCUES, the comparability of the designs had to be taken into account. The NHIS yearly sample for 1973-84 consisted of approximately 40,000 interviewed households selected in 376 different locations (primary sampling units or PSU's). The NMCUES sample size requirement is 10,000 households in approximately 100 PSU's. For the NSFG approximately 14,000 sample persons are selected in 100 PSU's. The target population for the NSFG, however, is restricted to women of childbearing ages (15-44 years) with a heavy oversampling of the black female population within this age group. The 1980 NSFG had to screen 55,000 households to produce the required sample of 14,000 sample persons.

Even with a requirement to oversample households with large health expenditures, it was determined that the NHIS yearly sample could produce a more than adequate sample for the NMCUES. Before the NSFG sample size requirements were investigated, two other NHIS design options were assumed. Both were considered likely options for the new NHIS design. The first option assumed that the number of PSU's in the NHIS was 200 instead of 376, and the second option assumed that the black population would be oversampled in the NHIS. The number of persons screened for the NSFG subdomains using each of the three NHIS design options (Moore, 1981; NCHS, 1981) is shown in table A. The 1973-84 NHIS would screen only one-half

Table A. Number of persons screened for the National Survey of Family Growth (NSFG) subdomains for 3 National Health Interview Survey (NHIS) design options

Survey design	Black women, ever married, 15–44 years of age	Black women, single, 15–44 years of age	Women other than black, ever married, 15-44 years of age	Women other than black, single, 15-44 years of age
Cycle III, NSFG.	3,600	2,601	5.401	2,401
100 of 376 PSU's[1] from NHIS	1,286	948	5,854	2,490
100 of 200 PSU's[1] from NHIS	1,968	I,450	8,757	3,726
100 of 200 PSU's from NHIS plus oversampling of black population	2,600	I,960	7,421	3,168

[1]PSU's = primary sampling units.

of the required number of black women of childbearing age over a l-year period. By oversampling black persons and assuming a 200-PSU design for the NHIS, the NHIS could screen approximately two-thirds of the desired number of black women 15-44 years of age. Because the NHIS is a continuous weekly survey of the United States, it is not necessary to restrict the screening for the NSFG to a l-year period. Although it was hoped it would not be necessary to screen for more than 1 year, the NSFG sample size requirements could be satisfied by accumulating the NHIS sample for 1½ years. Another alternative, which would shorten the screening period for the NHIS, would be to use an NHIS sample jointly with an independent sample. From this analysis it was concluded that the NHIS sample was large enough to serve as a sampling frame for the NMCUES and the NSFG.

After the determination that the surveys could be linked with respect to their designs, a cost analysis was undertaken. By using the NHIS as a sampling frame for the NMCUES and oversampling persons with chronic activity limitations, the current level of precision for NMCUES could be achieved with a sample that was smaller by 8 to 13 percent. This translates into a cost savings between 4 and 7 percent. For a single NMCUES costing approximately $15 million, a savings between $600,000 and $1.5 million could be achieved. If the NMCUES were repeated twice during the lo-year design period a savings between $1.2 and $2.1 million could be achieved. For the NSFG the major benefit of using the NHIS as a sampling frame is the ability to screen for women of childbearing age by race. Using the design assumptions in table A it appears that the screening for the NSFG could be reduced by 60 to 90 percent. A conservative estimate of the screening costs in the NSFG is one-fourth of the total cost of the survey, or $850,000. Thus the total estimated savings that would result from using the NHIS as a screener for the NSFG is between $500,000 and $750,000. If the NSFG were conducted twice in a decade, the savings would be between $1 million and $1.5 million.

If the NMCUES and NSFG are each conducted twice during the lo-year NHIS design period, a savings between $2.2 million and $3.7 million could be realized by using the NHIS as a sampling frame. These savings would more than offset the additional cost of $1.4 million required for an area sampling frame. These savings do not include the savings that would result from not having to design the NMCUES and the NSFG separately. Much of the initial field work for these surveys could be significantly reduced.

After reviewing the results of these analyses, NCHS made a decision to switch to an area frame design and made a commitment to the concept of linking its population surveys. Even without the potential cost savings of integrating its population surveys, the need for an area frame independent of the decennial census was crucial for meeting many of the NHIS long-range objectives.

One final note with regard to the U.S. Bureau of the Census confidentiality requirements: Even though the use of an area frame would release NCHS from the U.S. Bureau of the Census confidentiality provisions, NCHS confidentiality provisions under the Public Health Service Act still apply to all data collected. The information collected by NCHS is voluntary, strictly confidential, and used for research purposes only.

Multiple survey linkage

From the preceding discussions, it is clear that the linkage of the NCHS surveys is dependent on the use of the area frame design for the NHIS. Following the decision to use an area frame for NHIS and the commitment to the concept of linked surveys, NCHS initiated an integrated survey design (ISD) research program (Sirken and Greenberg, 1983). The intent of this program was to evaluate the feasibility of integrating the NCHS survey designs and to test, develop, and implement the design linkages that would best help NCHS carry out its mission. The ISD research would investigate linkages between both the population surveys and the establishment surveys. NCHS population surveys include the National Health Interview Survey (NHIS), the National Health and Nutrition Examination Survey (NHANES), the National Survey of Family Growth (NSFG), and the National Medical Care Utilization and Expenditure Survey (NMCUES). NCHS establishment surveys include the National Hospital Discharge Survey (NHDS), the National Nursing Home Survey (NNHS), and the National Ambulatory Medical Care Survey (NAMCS). The ISD research would also investigate follow-back linkages for these surveys such as the potential for telephone follow-back surveys. The goal was to increase the analytical potential of the total NCHS survey program as well as to reduce its cost.

The first priority of the ISD research was to investigate the linkages between the population surveys, especially those surveys on which the redesign of the NHIS would have the most impact. To date, research on the survey linkages between the NHIS and the NSFG and between the NHIS and the NMCUES has been completed. Research is currently underway to evaluate the potential of linking the NHIS and the NHANES and on the use of the NHIS PSU's for the NHDS. Planning has begun for a National Health Care Survey (NHCS) that would link all of the establishment surveys together into a single survey. The use of the NHIS PSU's for the NHCS will also be studied. And finally, a research study has been conducted to evaluate the feasibility of conducting telephone follow-back surveys from the NHIS.

All of these surveys will not be linked in the same way. The linkages that can be considered between any two surveys depend on the sampling unit definitions at each stage of sampling for the two surveys. For linkages between the population surveys and the establishment surveys, the use of a common set of geographical PSU's is currently being considered. For linkages between population surveys, the use of the same or a subset of the same PSU's, housing units, households (related family members), and individuals is being evaluated. The discussions in this report will be limited to the research conducted on the linkages between the NHIS, the NSFG, and the NMCUES.

Although the preliminary investigation of the linkages between the NHIS, the NSFG, and the NMCUES indicated a large potential savings from linking these surveys, more detailed research studies were needed to confirm the cost savings and to address a number of other methodological issues. Included among the other issues were the effect on response rates for the linked surveys, the impact on the precision of published statistics, the appropriate sampling unit for linkage, the method of making the initial follow-back contact, the length of time between the NHIS interview and follow-back interviews, necessary NHIS information to be collected and processed for the linkage, and the interface between the U.S. Bureau of the Census and the data collection agents for the NSFG and the NMCUES.

The research on these issues was conducted in two phases. During the first phase, private contractors conducted investigations of these issues using available information to evaluate a number of alternative options and designs. Based on the recommendations from this first phase of research studies, pilot studies were conducted during a second phase to develop and validate the linkage and data collection procedures in the field. Reports from these studies have been or soon will be published in NCHS Series 2 methodological reports (NCHS, 1985, 1987a, 1987b, 1987c). Brief summaries of the major findings from these studies are given below.

The first research study (NCHS, 1985)—the effects of integrating the NHIS and the NSFG—was conducted for NCHS by the Westat Corporation. Seven different linked designs along with the design for the third cycle of the NSFG were compared using the following criteria to evaluate the alternatives:

- Sample size necessary to achieve fixed and identical levels of precision for all alternatives.
- Alternative field and interviewing methods available for the sampling procedures.
- Cost of implementing the designs.
- Anticipated response rates.
- Time schedules.
- Potential administrative or operating problems.

The linked sample designs varied by the number of PSU's, sampling unit definition, method of interview assignments, and degree of subsampling NHIS households without eligible women. The major findings from the study were as follows:

- All of the alternative linked designs required significantly smaller sample sizes than the Cycle III NSFG model to achieve the sample precision. The range was from 10,400 to 11,500 persons compared with 14,000 persons (no allowance for extra nonresponse). This translates into direct cost savings of more than 30 percent for some of the linked designs.
- The NHIS could provide an adequate sample for the NSFG.
- Response rates should remain approximately the same for the linked design versus an independent design. Small losses in response will occur because of nonresponse in the NHIS and the use of an individual sampling unit definition (failure to locate or interview persons who had moved). This result must be confirmed in a field test.
- The administrative and field procedures vary considerably for the alternative designs. The advantages and disadvantages were presented for each design. NCHS must weight the various factors in selecting a final design.
- There were no insurmountable operational problems associated with any of the alternative linked designs.

Following the research study, a field trial was conducted by Westat to address some of the operational issues (NCHS, 1987a). Three factors that were investigated include (1) sampling unit definition-housing unit versus sample women, (2) mode of initial contact-telephone versus personal visits, and (3) length of time between the NHIS interview and the NSFG interview-various lengths from 1 to 15 months. The dependent variables for the study were response rate, level of effort, and cost. The field trial was conducted in 10 of the NHIS PSU's with a sample size of 1,315 NHIS households. The major findings from the field trial were as follows:

- Although the difference between the overall response rate for the housing unit sample and the selected person sample (83.5 percent versus 82.1 percent) was not statistically significant, it appears that somewhat higher response rates would be achieved with a housing unit definition.
- There is no single best linkage strategy. The tradeoff between increased response rates and reductions in costs makes the best sampling unit definition a subjective decision.
- The mode of initial contact did not appear to affect response rates for either type of sample unit. The mode of initial contact did, however, have an important effect on the overall level of effort and associated costs. The ratio of personal visits per completed interview was significantly lower for cases assigned to a telephone mode of initial contact than for cases initially contacted in person.
- There were no clear findings with respect to the effects of elapsed time on response rates. It appears that the major impact of elapsed time would be seen in the level of tracking necessary for the selected person sample.

These findings do not lead to the determination of the optimum design for the linked NSFG. They do, however, provide valuable information for selecting a final design.

The second research study (NCHS, 1987b)—the effects of integrating the NHIS and the NMCUES—was conducted for NCHS by the Research Triangle Institute (RTI). The following four designs were compared:

- An unlinked NMCUES design.
- A NMCUES design linked to the NHIS dwelling units.

- A NMCUES design linked to the household members within the NHIS dwelling units.
- A linked household design that oversampled heavy users of health care services.

For the linked designs, the number of PSU's and total sample size were each assigned two levels. The criteria used to compare the design were sample size required for fixed precision, costs, and time needed to aggregate the NHIS sample. The results of the study indicated that very little cost savings could be achieved with the first two linked designs. If, however, the NHIS could be used to identify and oversample important subdomains, such as persons with large expenditures or high utilizers of health care, the direct costs for the NMCUES could be reduced by as much as 20 percent for the same level of precision. Because of the relative sample sizes of the NHIS and the NMCUES, only a few months would normally be required to accumulate a NHIS sample large enough for the NMCUES.

The field trial for the linkage between the NHIS and the National Medical Expenditure Survey (NMES) was conducted by the Westat Corporation (NCHS, 1987c). NMES represents the next cycle of the national health expenditure surveys following NMCUES. The field trial evaluated alternative strategies for defining the sampling unit-household and housing units-and the mode of initial household contact-telephone and in-person. The criteria used in the evaluation were response rates, level of effort, and costs. The sample size for the field trial was 600 households in eight NHIS PSU's. The findings of the field trials were as follows:

- The overall response rates for the household sample and the housing unit samples were not significantly different (86.0 percent versus 88.4 percent). The direction of the difference, however, was in the expected direction and was similar to results from the NSFG field trial.
- The response rate associated with the in-person mode of initial contact for the NMES did appear to be higher than the response rate for initial telephone contacts (90.1 percent versus 84.4 percent). This result was not statistically significant. The telephone mode of initial contact for the NMES did significantly reduce the overall level of effort and costs associated with a completed interview.
- None of the experimental treatments in the field trial appeared to interact with any of the demographic subdomains of interest for the NMES.
- The response rates for telephone and in-person interviews in the second round of interviews of the field trial were similar. The quality of the data collected on the telephone, however, as measured by the rates of item nonresponse for selected questionnaire items, was somewhat poorer than that for the in-person views.

These results, like the NSFG linkage field results, do not lend themselves to a single best design strategy. They do provide information for selecting a design.

The sum results of the ISD research confirm the preliminary linkage study evaluations with respect to feasibility and costs. It is clear that an integrated design approach for NCHS surveys is both feasible and cost effective.

Mode of data collection

Since its inception in 1957 the NHIS has been conducted using face-to-face household interviews. The U.S. Bureau of the Census interviewers have utilized the telephone in the NHIS on occasions to collect information in followup interviews from household members who were not at home at the time of the initial face-to-face NHIS interview. The followup interviews were usually required to collect information on yearly supplemental topics requiring self-response. For the redesign of the NHIS, use of a dual-frame design was evaluated to determine whether both telephone and face-to-face interviews could be used for the initial NHIS household interview.

Since 1978 NCHS has been actively involved in telephone survey research in an effort to improve the efficiencies of its surveys. Much of the research has involved the NHIS and has included the comparability of telephone and face-to-face interviews and some of the implications of using a dual-frame, telephone and face-to-face, design for the NHIS (Casady, Snowden, and Sirken, 1981; NCHS, 1987d; Sirken and Casady, 1982). The research results from the various studies all indicated potential advantages of conducting some of the NHIS interviews by telephone.

In 1982 a joint NCHS and U.S. Bureau of the Census Telephone Survey Task Force was established to propose recommendations for the research and development of a dual-frame NHIS. The task force was charged with identifying the key methodological issues that should be researched for a dual-frame design and for preparing a plan for addressing these issues and implementing a dual-frame design by 1986. In March 1983 the task force made the following recommendations for the development of the dual-frame NHIS (Joint NCHS-Bureau of the Census Telephone Survey Task Force, 1983).

- Explore improvements in total dual-frame survey design and estimation methods through theory development, modeling, and simulation.
- Develop a random digit dialing (RDD) sample design for the NHIS using existing methods.
- Adapt the NHIS questionnaire and field procedures for a telephone interview and conduct an RDD feasibility study of the telephone questionnaire.
- Develop a computer-assisted telephone interview (CATI) for the NHIS.
- Estimate the primary sources of nonsampling errors for telephone and face-to-face interviewing procedures.
- Develop new, rapid coding and data-processing procedures for the NHIS data-collection activities.
- Conduct a dual-frame feasibility study.

In order to accomplish these objectives, it was recommended that both an operational telephone system and a long-range research program be established. The operational system would be used to conduct the feasibility studies required for the implementation of a dual-frame design. The RDD feasibility study would address methodological issues related to an NHIS telephone interview. The principal issues to be tested included the estimation of the telephone survey response rate, telephone data-collection costs, the comparability of data across modes, respondent rule comparability, and operational differences between modes. If these issues were satisfactorily resolved, a second feasibility study would then be conducted to address the methodological issues related to a dual-frame NHIS design. The task force recommendations were presented to a joint agency steering committee which accepted the recommendations but gave a higher priority to the immediate development of an operational telephone system.

In the fall of 1983 the U.S. Bureau of the Census established a telephone interviewing system for conducting the NHIS-RDD feasibility study. The data collection for this study was from January through May of 1984. The following nine specific objectives were defined for the feasibility study.

- Test the feasibility of conducting the entire core component of the NHIS questionnaire by telephone.
- Estimate the response rate for the telephone component of a dual-frame NHIS.
- Estimate the costs for conducting the telephone component as part of a dual-frame design.
- Evaluate alternative questionnaire structures in terms of length and effect on estimates.
- Identify operational problems associated with administering the NHIS by telephone.
- Develop and evaluate procedures for identifying and handling special places, such as group quarters and hotels, over the telephone.
- Conduct preliminary development and testing of estimation procedures, including nonresponse and post-stratification adjustments.
- Test procedures for the assignment, management, and completion of samples for producing valid estimates.
- Evaluate the operational feasibility and effect on response rates of using a most knowledgeable respondent rule.

The major findings from the study were as follows:

- For the NHIS, RDD response rates of 85 percent (for an average interview length of about 50 minutes) are feasible. However, interviewer and field staff experience are critical factors in achieving such rates. Over the duration of the study, response rates increased by 12 percentage points, and item nonresponse levels were substantially reduced.
- The length of the NHIS interview and nature of the questions did not appear to cause any operational problems.
- The person-by-person questionnaire version resulted in higher reporting of health events than did the family version. Response rates for the two questionnaires were essentially the same.
- The total time (including all interviewer activities such as dialings and callbacks) per interviewed household for this study ranged by replicate from an average of 70 minutes (earlier replicates) to 49 minutes (later replicates) for an overall replicate average of 61 minutes.
- In more than 90 percent of the completed cases, the most knowledgeable respondent (as identified by the phone answerer) was reached on the first household contact. For cases requiring callbacks to reach the most knowledgeable respondent, the refusal rate was three times greater than the average.
- The automated call scheduler performed efficiently in the later replicates where the number of unresolved cases dropped considerably. A longer interview period (say 4 weeks) could have increased the average response rate from 2 to 4 percentage points by reducing the number of unresolved cases.
- There was a problem with identifying special places (using the NHIS definitions) in the RDD survey. Roughly 30 percent were classified as nonresidential units, and 20 percent were classified as residential units other than special places.
- Cost estimates from the RDD feasibility study cannot be used to estimate the telephone costs for a continually operating dual-frame telephone and face-to-face survey. Many of the costs associated with the feasibility study were development costs for a telephone survey system and costs associated with a one-time survey.

A detailed description of the feasibility study and an analysis of the results are presented in a joint report by the U.S. Bureau of the Census and NCHS (1985) and in several papers presented at American Statistical Association meetings (Chapman and Roman, 1985; Hogue and Chapman, 1984; Tompkins and Massey, 1986).

On the whole the results of the NHIS-RDD feasibility study were positive, although there were some indications that a dual-mode NHIS would not be as cost effective as originally thought. The administrative costs of operating two modes of data collection simultaneously significantly reduce the cost savings produced by telephone interviews. Following the completion of the feasibility study, a decision was made by NCHS not to change the NHIS to a dual-mode survey. This decision was based primarily on the incompatibility of the RDD telephone component with other design features being considered for the NHIS (see later discussions in this chapter).

Oversampling small subdomains

One of the objectives that the NHIS has always emphasized has been publication of detailed descriptive health statistics for the U.S. population. This is the rationale for having a cross-sectional sample design that can be accumulated from week to week and from year to year. There are, however, a number of important small subdomains of the U.S. population for which detailed statistics cannot be routinely published. These subdomains are extremely important in understanding health differences within the population; they include the following groups: Black, Hispanic, the aged, and low income. Each of these subdomains represents between 5 and 12 percent of the U.S. population, although there is some overlap between the subdomains. For the redesign of the NHIS, NCHS asked the U.S. Bureau of the Census to investigate ways in which the precision of the estimates for these four subdomains could be improved without too adversely affecting the precision of other NHIS estimates or without adding significantly to the cost of the NHIS.

The description and results of the research conducted by the U.S. Bureau of the Census are presented in a series of internal census memoranda (Mazur, 1983; Olsen, 1981, 1982; Ponikowski, 1982; Ponikowski and Mazur, 1983; Tadros, 1982a, 1982b) and in Chakrabarty (1985). The Statistical Methods Division (SMD) within the U.S. Bureau of the Census considered four different approaches for oversampling the four special subdomains to improve the precision of the related statistics. The four methods were snowball sampling, screening of additional households, oversampling at the first stage of sampling (PSU's), and oversampling at the second stage of sampling (clusters of housing units). The first two methods were discarded very early because of the complexities and costs associated with the procedures. Except for the investigation of Hispanic statistics, oversampling at the second stage of sampling was the primary method considered for improving the precision of estimates for the special subdomains. This method is described in a paper by Waksberg (1973) and involves the stratification of the secondary sampling units (SSU's) within the PSU's. For a given subdomain two strata of SSU's are formed within each PSU. One stratum contains a very high proportion of the subdomain of interest; the second stratum does not.

Different sampling rates are used within the two strata to oversample and improve the precision of the subdomain of interest. Information obtained in the decennial census would be used for the stratification. The effectiveness of this approach depends on how clustered the subdomain of interest is within the PSU's, that is, on how well the stratification can separate the subdomain of interest from other persons within the PSU. Using this method of stratification the following results were obtained by SMD.

- Oversampling at the second stage of sample selection within PSU's was not an effective way to improve the precision of the statistics for the aged or for persons with low income. These subdomains, especially the aged, are not highly clustered within PSU's, and only very small gains in precision could be achieved. Significantly improving the precision of statistics for these subdomains would require other methods that would be costly.
- The reliability of estimates for black persons could be improved considerably (8-25 percent) by optimally oversampling black persons in PSU's with large black populations while constraining the variance increases for the total population to no more than 2 to 3 percent.
- The reliability of the estimates for Hispanic persons could be improved only 3 to 5 percent by optimally oversampling Hispanic persons in PSU's with large Hispanic populations while constraining the variances increases for the total population to less than 5 percent.

The last two results reflect the fact that the black population tends to be more clustered within PSU's than does the Hispanic population. For the stratification method to be effective in improving the precision of estimates for a subdomain in the NHIS redesign, however, the subdomain population must stay highly clustered over the decade 1985-94. Based on a U.S. Bureau of the Census analysis, it was found that black households in the 1970 census contained 89 percent black households 6 years later. This percentage was approximately 65 percent for Hispanics. Thus, it appears that the stratification method of oversampling subdomains would be most effective for black persons for the duration of the NHIS redesign.

The latter result led SMD to investigate the possibility of adding additional PSU's in the five southwestern States containing the largest Hispanic populations. The total number of PSU's was constrained to be constant. The results of this analysis indicated that significant improvements for the precision of Hispanic estimates could be made in the Southwest, but that only modest improvements (about 5 percent) could be made at a national level. These improvements were partially offset by a decline in the precision for other subdomains in the Southwest and in the total U.S. population.

Rotation of the NHIS sample

One of the important new objectives for the redesign of the NHIS was to improve the survey's ability to evaluate the dynamics of change. The dynamics of change refer to the distribution of changes that occur among individuals. For example, how often do individuals go back and forth from a well state of health to a sick state of health over a given period of time? In order to obtain better measures of change a new design feature was proposed. The new proposed feature was the incorporation of a rotational sample into the design. This would involve selecting a subsample of the NHIS housing units interviewed during one calendar year and reinterviewing them in subsequent years as part of the subsequent year's sample.

The previous NHIS sample designs retained the same PSU's from year to year but selected new households in adjacent clusters to be interviewed each year. The rotational designs would take advantage of the correlations between PSU's and clusters from one year to the next to improve NHIS's ability to measure change, whereas the previous designs only take advantage of the PSU year-to-year correlations.

SMD investigated a number of alternative rotational patterns that involved reinterviewing sample persons 1 or 2 years after their initial interview. The patterns most thoroughly investigated were those that reinterviewed subsets of the sample persons in consecutive years. The criteria used to compare the rotational design's ability to measure change with the old design's ability to measure change were the variance of change from one year to the next, the variance of a 2-year combined mean, and the variance of a single-year mean. After a model was developed for estimating these variances, a sensitivity analysis was performed by varying the correlation between PSU means from year to year, the correlation between clusters within PSU's from year to year, and the yearly between-PSU component of variance.

The results of these analyses for two rotational patterns are shown in table B. More complete description of methods and results are documented in internal SMD memoranda (Hsen, 1981, 1982a, 1982b) and an NCHS report (1982a). The two rotational designs shown in table B would reselect one-fourth and one-half of the NHIS sample housing units from one year to the next. It can be seen that under the right set of circumstances improvements of as much as 30 percent can be made in estimating year-to-year change using the rotational designs. This is accompanied by a loss in precision for the 2-year means. For most health statistics an increase in the cluster year-to-year correlations of as much as or more than 0.25 for the rotational designs is unlikely.

Two other issues related to the rotational designs were raised by NCHS and SMD. The first issue involved the sampling unit to use for the rotational designs. In the SMD analysis it was assumed that housing units would be selected and surveyed from one year to the next. This would not be a satisfactory sampling unit definition for analyzing the individual dynamics of change. In order to evaluate the dynamics of change the same sample of persons must be interviewed from one year to the next. Although this would be feasible, a considerable amount of tracking and cost would be required to follow sample persons from one year to the next. No further analysis was conducted, but NCHS has analyzed this issue for the NMCUES.

The second issue raised was the possibility of conducting the NHIS reinterviews in subsequent years by telephone for households with telephones. This would significantly reduce the data-collection costs for the NHIS over the decade. This issue was to be reevaluated after the telephone mode of data collection was tested for the NHIS and determined to be feasible.

Table B. Percent change in variances for 2 rotational designs relative to 1973–84 National Health Interview Survey (NHIS) design assuming a 25- and a 50-percent increase in the cluster year-to-year correlation (ρ)

	Percent of sample rotated from year-to-year			
	One-fourth		*One-half*	
Estimate	ρ = 0.25	ρ = 0.5	ρ = 0.25	ρ = 0.5
Year-to-year change	-7	-19	-13	-31
2-year mean.	+4	+6	+8	+14
1-year mean.	-1	-4	-1	-5

NOTE: The year-lo-year correlation between clusters for the 1973–84 design is assumed to be negligible.

Interaction between design features

After researching and comparing the new options independently for each of the five design features with the 1973-84 design options, the redesign model called for the formation of design-option sets using combinations of the design options. This procedure was slightly modified by first taking a look at all of the two-way interactions between the new options for the design features. For this analysis the area frame option was combined with the linked survey option, as they are so dependent upon each other. This left the six two-way interactions shown in figure 2. The final design of the NHIS was dependent on the compatibility between the various design options. A subjective rating and a brief synopsis of the compatibility between each of the new options are also given in figure 2.

The design options that are least compatible are the linked population survey option with the RDD telephone option and the linked survey option with the sample rotation option. For the first interaction, the NSFG and the initial interview in the NMCUES both require that the interview be done in-person. Following up RDD telephone interviews with a face-to-face interview for the linked surveys would be extremely costly and inefficient due to the lack of geographical clustering of housing units in RDD surveys. The optimum designs for the NSFG and the NMCUES are highly clustered designs at both the first and second stages of sample selection. An RDD sample is not feasible for either of these surveys. The rotation design feature is slightly more compatible with the linked survey feature, although the compatibility is still rated as poor. This is because an overlapping sample for the NHIS from year to year will reduce the total number of eligible households for the NSFG and the NMCUES. Research results show that the NHIS sample is already strained to meet the sample design requirements for the NSFG.

The compatibility between all of the other new options is acceptable, and, in fact, the oversampling of black persons will greatly enhance the linkage between the NHIS, the NSFG, and the NMCUES. The compatibility of the three- and four-way interactions between the new design options requires that all of the two-way interactions be compatible.

Conclusions

Based on the research results presented in the last section and the long-range goals of NCHS, the decision was to make the NHIS sample the cornerstone of the NCHS Integrated Survey Design (ISD) program. This meant switching to an area frame for the NHIS and beginning to establish the survey linkages between the NHIS, the NSFG, and the NMCUES. This also opened up the possibility of conducting follow-back or longitudinal panel surveys for the NHIS supplemental topics.

After the area frame and linkage options had been selected, it was recommended that black persons be oversampled in the NHIS for the 1985-94 design period

Interacting options	Design compatibility
Linkage X telephone (NHIS)	*Poor:* NSFG and NMCUES currently require face-to-face interview.
Linkage X rotation	*Poor:* NHIS reinterviews reduce yearly sample for NSFG and NMCUES.
Linkage X oversampling	*Good:* Oversampling of black persons increases yearly sample for NSFG and NMCUES.
Telephone X rotation.	Good: NHIS reinterviews can be conducted by telephone at lower cost.
Telephone X oversampling	*Fair:* Screening for special domains using RDD procedures is relatively inexpensive.
Rotation X oversampling	*Good:* No apparent adverse effects from interaction.

NOTE: NHIS = National Health Interview Survey; NSFG = National Survey of Family Growth; NMCUES = National Medical Care Utilization and Expenditure Survey; RDD = Random digit dialing.

Figure 2. Two-way interactions for design options considered for the National Health interview Survey redesign

(NCHS, 1982b). It was further recommended that all other special subdomains not be oversampled. The research showed that any significant improvements in precision of 25 percent or more were not easy to achieve and would be offset by decreases in the precision of estimates for other subdomains. This recommendation was adopted for the NHIS.

The rotation of the NHIS sample was not adopted for the NHIS because of its incompatibility with the linking of the NCHS population surveys. It was also reasoned that follow-back interviews could be conducted for the NHIS on an ad hoc basis and that a continuous rotating sample was not required. The suggestion of conducting follow-back interviews by telephone had considerable appeal and will be used by the NHIS in the future.

The use of a dual-frame telephone and face-to-face survey was not adopted for the 1985 redesign because of its incompatibility with an ISD. It was recommended that the Division of Health Interview Statistics investigate other uses of RDD telephone surveys. The recommendation was accepted.

Although the conceptual model used to investigate the design features and design options was considerably modified in practice, the model proved extremely valuable in the planning and organizing of the research for the redesign of the NHIS. Because of the many complexities involved in designing or redesigning a major national survey, the task of coordinating all of the research activities for the NHIS redesign was enormous. Using a model for the redesign effort made the task more manageable.

Once all of the major design features had been selected for the NHIS, the specific parameters for the NHIS design had to be determined. This process is described in chapter III.

References

Casady, R. J., C. B. Snowden, and M. G. Sirken. 1981. A study of dual frame estimators for the National Health Interview Survey. *Proceedings of the Survey Research Methods Section,* American Statistical Association, pp. 444-447.

Chakrabarty, R. P. 1985. Redesign of the National Health Interview Survey. *Proceedings of the Survey Research Methods Section,* American Statistical Association, pp. 252–257.

Chapman, D. W., and A. M. Roman. 1985. An investigation of substitution for an RDD survey. *Proceedings of the Survey Research Methods Section,* American Statistical Association, pp. 269-274.

Hogue, C. R., and D. W. Chapman. 1984. An investigation of PSU cutoff points for a random digit dialing survey. *Proceedings of the Survey Research Methods Section,* American Statistical Association, pp. 286-289.

Hsen, P. L. 1981. HIS Redesign: Alternative Rotation Patterns. Internal memorandum from Rameswar P. Chakrabarty to Gary M. Shapiro. July 30. U.S. Bureau of the Census. Washington, D.C.

Hsen, P. L. 1982a. HIS Redesign: Alternative Approach of Variance Estimation for the Rotation Patterns. Internal memorandum from Rameswar P. Chakrabarty to Gary M. Shapiro. Feb. 3. U.S. Bureau of the Census. Washington, D.C.

Hsen, P. L. 1982b. HIS Redesign: Comparison of Variance Estimation for the Rotation Patterns Assuming Different Percentage of Between PSU Variance Component. Internal memorandum from Paul L. Hsen to Rameswar P. Chakrabarty. Aug. 19. U.S. Bureau of the Census. Washington, D.C.

Joint NCHS-Bureau of the Census Telephone Survey Task Force. 1983. Research and Development Recommendations Leading to a Dual Frame National Health Interview Survey. Unpublished report.

Jones, C. D. 1981. HIS Redesign-Level One Research: Cost Comparisons for Sample Design Alternatives. Unpublished letter to Monroe Sirken, NCHS. May 26.

Massey, J. T., K. H. Marquis, and R. D. Tortora. 1982. Methodological issues related to telephone surveys by Federal agencies. *Proceedings of the Social Statistics Section,* American Statistical Association, pp. 63-72.

Massey, J. T., and M. G. Sirken. 1982. A model for designing NCHS's population-based surveys. *Proceedings of the Social Statistics Section,* American Statistical Association, pp. 285-290.

Mazur, C. 1983. HIS Redesign: Differential Sampling to Achieve a Reduction in Black Subgroup Variances. Internal memorandum from Rameswar P. Chakrabarty to Gary M. Shapiro. June 15. U.S. Bureau of the Census. Washington, D.C.

Moore, T. F. 1981. HIS Redesign: HIS as Frame for FGS and MCUES. Internal memorandum from Rameswar P. Chakrabarty to Gary M. Shapiro. Jan. 14. U.S. Bureau of the Census. Washington, D.C.

National Center for Health Statistics. 1980. *National Health Interview Survey, Report of the National Committee on Vital and Health Statistics.* DHHS Pub. No. (PHS) 81-1160. Public Health Service. Hyattsville, Md.

National Center for Health Statistics. 1981. Discussion of the Major Design Features for the Redesign of the NHIS. Unpublished memorandum. Mar. 20. Office of Research and Methodology. Hyattsville, Md.

National Center for Health Statistics. 1982a. Recommendations and Proposed Schedule for the Redesign of NCHS's Population-Based Surveys. Unpublished report. Survey Design Task Force. Hyattsville, Md.

National Center for Health Statistics. 1982b. Recommendations for the Redesign of the NHIS. Unpublished memorandum to Monroe G. Sirken. Oct. 7. NHIS Sample Design Task Force. Hyattsville, Md.

National Center for Health Statistics, J. Waksberg and D. R. Northrup. 1985. Integration of sample design for the National Survey of Family Growth, Cycle IV, with the National Health Interview Survey. *Vital and Health Statistics. Series 2, No. 96.* DHHS Pub. No. (PHS) 86-1370. Public Health Service. Washington: U.S. Government Printing Office.

National Center for Health Statistics, N. A. Mathiowetz, D. Northrup, S. Sperry, and J. Waksberg. 1987a. Linking the National Survey of Family Growth with the National Health Interview Survey. *Vital and Health Statistics.* Series 2, *No*. 103. DHHS Pub. No. (PHS) 87-1377. Public Health Service. Washington: U.S. Government Printing Office.

National Center for Health Statistics, B. G. Cox, R. E. Folsom, and T. G. Virag. 1987b. Design alternatives for integrating the National Medical Expenditure Survey with the National Health Interview Survey. *Vital and Health Statistics.* Series 2, No. 101. DHHS Pub. No. (PHS) 87-1375. Public Health Service. Washington: U.S. Government Printing Office.

National Center for Health Statistics, N. A. Mathiowetz and E. P. Ward. 1987c. Linking the National Medical Expenditure Survey with the National Health Interview Survey. *Vital and Health Statistics*. Series 2, No. 102. DHHS Pub. No. (PHS) 87-1376. Public Health Service. Washington: U.S. Government Printing Office.

National Center for Health Statistics, O. T. Thornberry, Jr., et al. 1987d. An experimental comparison of Telephone and Personal Health Interview Surveys. *Vital and Health Statistics. Series 2, No.* 106. DHHS Pub. No. (PHS) 87-1380. Public Health Service. Washington: U.S. Government Printing Office.

Olsen, C. 1981. HIS Redesign: Improvement of Small Domain Estimates. Internal memorandum from Christina Olsen to Rameswar P. Chakrabarty. Jan. 14. U.S. Bureau of the Census. Washington, D.C.

Olsen, C. 1982. HIS Redesign: Differential Sampling to Achieve a Reduction in Demographic Subgroup Variances. Internal memorandum from Rameswar P. Chakrabarty to Gary M. Shapiro. May 3. U.S. Bureau of the Census. Washington, D.C.

Plewes, T. J. 1980. Planning for the Current Population redesign effort. *Review of Public Data* Use, 8, pp. 141-145.

Ponikowski, C. 1982. HIS Redesign: Results of Research on Oversampling of Hispanics in the Southeastern United States. Internal memorandum from Rameswar P. Chakrabarty to Gary M. Shapiro. Nov. 16. U.S. Bureau of the Census. Washington, D.C.

Ponikowski, C., and C. Mazur. 1983. HIS Redesign: Results of Research To Improve Spanish Estimates Using Different Measures of Size. Internal memorandum from Rameswar P. Chakrabarty to Gary M. Shapiro. Aug. 12. U.S. Bureau of the Census. Washington, D.C.

Sirken, M. G., and R. J. Casady. 1982. Nonresponse in dual frame surveys based on area/list and telephone surveys. *Proceedings of the Survey Research Methods Section,* American Statistical Association, pp. 151-153.

Sirken, M. G., and M. S. Greenberg. 1983. Redesign and Integration of a Population-Based Health Survey Program. Contributed paper, Proceedings of 44th Session of the International Statistical Institute. Madrid.

Sirken, M. G., and P. N. Royston. 1976. Design effects in retrospective mortality surveys. *Proceedings of the Social Statistics Section,* American Statistical Association, pp. 773-777.

Tadros, W. 1982a. HIS Redesign: Differential Sampling to Achieve a Reduction in Demographic Subgroup Variances, II. Internal memorandum from Rameswar P. Chakrabarty to Gary M. Shapiro. Sept. 22. U.S. Bureau of the Census. Washington, D.C.

Tadros, W. 1982b. HIS Redesign: Estimating National Variance Changes if Black or Spanish Persons Are Oversampled. Internal memorandum from Rameswar P. Chakrabarty to Gary M. Shapiro. Oct. 8. U.S. Bureau of the Census. Washington, D.C.

Tompkins, L., and J. T. Massey. 1986. Using a most knowledgeable respondent rule in a household telephone survey. *Proceedings of the Survey Research Methods Section,* American Statistical Association, pp. 281-286.

U.S. Bureau of the Census and National Center for Health Statistics. 1985. The Results of the 1984 NHIS/RDD Feasibility Study: Final Report. Unpublished report.

Waksberg, J. 1973. The effect of stratification with differential sampling rates on attributes of subsets of the population. *Proceedings of the Social Statistics Section,* American Statistical Association, pp. 429-434.

Chapter III. The 1985–94 NHIS sample design

by Thomas F. Moore and William Tadros,
Statistical Methods Division, U.S. Bureau of the Census

The U.S. Bureau of the Census, in particular the Statistical Methods Division (SMD), conducted research on the parameter options for the major design features of the 1985-94 National Health Interview Survey (NHIS) as described in chapter II. SMD conducted its research over a period from early 1979 to mid-1983 when the final choice of the new design was made. This research resulted in the changes to the design parameters displayed in figure 1 of chapter I.

The research program included work in both the operational aspects of the NHIS and its sample design. This section centers on the sample design research; a description of the research activities used to determine design parameters and the procedure used to select the NHIS sample follows.

Research for redesign of NHIS sample

In the past, the NHIS surveys conducted by the U.S. Bureau of the Census (hereafter referred to as Census Bureau) were closely linked to the Current Population Survey (CPS) in order to minimize costs. Consequently, the NHIS had design features similar to the CPS (U.S. Bureau of the Census, 1978). Traditionally, the Census Bureau redesigns its major demographic surveys following the decennial Census of Population, and early in 1979 a large-scale research project was begun for the NHIS. The NHIS did not use CPS interviewers and, accordingly, did not need to be restricted to CPS primary sampling units (PSU's). Also, the redesigned survey would use an area sampling frame within PSU's rather than the multiframe design of the CPS. For these reasons, the NHIS was designed independently of the CPS.

The primary objective for the NHIS redesign was to select a design that maintained the precision of the NHIS at the level of the 1973-84 design at the lowest possible cost and was compatible with the analytical objectives of the survey. The NHIS redesign research included the following:

- Reevaluating the PSU definitions.
- Determining the optimum number of PSU's for the sample.
- Determining the optimum size of the secondary sampling unit (SSU).
- Determining the total number of sample households.
- Evaluating variables for stratifying PSU's and the number of sample PSU's to be selected per stratum.
- Selecting a method for forming representative subsamples of PSU's.

Primary sampling units

Several options existed regarding how PSU's should be defined, including whether to continue to use multicounty PSU's, or to use single county or subcounty level PSU's. The obvious benefit of smaller PSU's would be a decrease in travel within sample PSU's. It was also possible that the between-PSU variance would be reduced. The use of single county or subcounty PSU's was investigated by the Census Bureau, using labor force data from the 1980 census. The findings from the research indicated that cost savings would probably be small and that variances might actually increase (Kostanich, 1982). Multicounty PSU's were judged to be more efficient and were retained for the redesign.

In the 1973-84 NHIS design, the sample was spread over 376 sample PSU's and over 52 weeks of the year. Interviewer assignment areas, consisting of several PSU's, were formed, and one interviewer interviewed all the PSU's in the area. As a result, travel to and from PSU's was extensive and consumed about 35 percent of the direct field costs for travel and interviewing. Travel to and from PSU's could be reduced by decreasing the number of PSU's or by hiring more interviewers. However, more interviewers would mean less work annually for each interviewer. Because adequate work loads must be preserved for NHIS interviewers, who work on no other Census Bureau surveys, the preferred way to reduce the number of PSU's with no resident interviewer was to reduce the total number of sample PSU's (Tadros, Moore, and Chakrabarty, 1982).

To determine the optimum number of PSU's for the NHIS it is necessary to estimate the components of variance (precision of NHIS statistics) and the components of cost. To estimate the precision of NHIS statistics under alternative numbers of PSU's, the following formula (see Moore, 1988, for derivation) was used to estimate the variance of potential designs for 1985–94 relative to that of the 1973-84 design:

$$\mathrm{VR}=Bf\frac{m_0}{m_1}+(1-B)\frac{u_0}{u_1}\;\frac{1+\delta_1(\bar{q}_1-1)}{1+\delta_0(\bar{q}_0-1)} \qquad (1)$$

where m = number of nonself-representing (NSR) PSU's in the sample,

u = number of housing units in the sample,

δ = intraclass correlation between housing units within clusters,

$\bar{q}$ = average number of housing units in the sample per cluster in the sample, and

B = between-PSU variance as a proportion of total variance.

The factor f reflects the effect of the changes in the NSR population relative to the 1973-84 design. It was expected to be greater than 1.0 for designs with larger NSR populations and was determined empirically to be about 1.03 for most design alternatives. The additional subscripts 0 and 1 refer to the 1973-84 and potential 1985-94 designs, respectively.

The values to be substituted into formula (1) were obtained from a variety of sources. Data presented in table A are from the National Center for Health Statistics (NCHS). These data showed that in the 1973-84 design the between-PSU variance proportion, B, for the major subgroups in the population was in the 0 to 20 percent range for many important health characteristics (NCHS, 1982). A value of B = 0.10 was used in the research to represent most important health items.

The intraclass correlation is a measure of homogeneity within clusters. Results of a study (NCHS, 1973) on δ showed that for compact clusters of six housing units, δ ranged from 0.01 to 0.17 for the majority of health characteristics, with the important items (for example, number of chronic conditions, number of bed days, number of restricted activity days) ranging from 0.02 to 0.06. A value of 0.04 was used for the intraclass correlation. Preliminary work showed that the optimal number of sample PSU's was not very sensitive to the value of δ.

The average number of sample housing units per cluster (secondary sampling units within PSU's) in the 1973-84 design was four. Most of the potential new designs were based on an average of six or eight, although other values were studied (see next section).

To allocate the sample optimally, it was also necessary to estimate variable and fixed costs for the survey. The variable costs include costs for data collection, recruiting staff, training, reinterview, observation, office work, pretests, sample maintenance, and associated overheads. Fixed costs primarily cover the Census Bureau's Washington, D.C., staff.

Data-collection costs are the direct payments to interviewers for interviewing and travel to PSU's. They depend on the sample size, the number of PSU's, and the cluster size.

In the 1973-84 design, an average of slightly more than four clusters of size four generally was assigned to one interviewer to be interviewed in 1 week. In 1 year approximately 3,175 weekly assignments were given to 120 interviewers, each working about 26 weeks a year.

If clusters of six housing units were used, the Census Bureau's field staff planned to form interviewer assignments of three clusters to yield expected assignment sizes of 18 housing units. For clusters of eight housing units, assignments would include two or three clusters, averaging 2.15 clusters per assignment. Thus, assignments would include 16 or 24 housing units. For a given sample size, the number of interviewers and weekly assignments can be estimated with this information.

An assumption was made that interviewers would be hired in PSU's based on the sample size of the PSU's. PSU's with about 26 weeks or more of work would all have at least one interviewer, and the remaining interviewers needed for that sample size would be hired from some of the smaller PSU's. This approach permitted estimation of the number of PSU's with no resident interviewer and the number of weekly assignments that would require travel. Data-collection costs were estimated as a function of the total number of weekly assignments and the number of assignments requiring travel to the PSU.

Table A. Estimated percent of between-PSU variance for the 1973–94 National Health Interview Survey design by selected health variables and demographic domains

		Sex		Race		Age				
Health variable	Total	Male	Female	Black	Other than black	Under 15 years	15-44 years	45-64 years	65 years or over	Average
Total restricted activity days	29	26	23	12	35	8	17	15	21	21
Total bed days	18	5	18	7	18	7	7	3	2	9
Total physician visits	27	13	30	9	25	15	5	(1)	5	14
Mean number of physician visits	31	10	28	16	19	13	6	6	(1)	14
Proportion seeing physician	3	1	(1)	8	3	4	5	6	(1)	3
Total dental visits	(1)	2	(1)	1	(1)	(1)	(1)	4	(1)	<1
Mean number of dental visits	4	1	4	1	4	(1)	(1)	2	(1)	2
Proportion seeing dentist	(1)	1	(1)	1	(1)	(1)	(1)	3	(1)	<1
Mean number of chronic conditions	21	17	15	6	20	3	8	18	7	13
Proportion with chronic condition	14	16	9	7	16	(1)	8	11	4	9
Average	15	9	13	7	14	5	6	7	4	9

[1]Estimated between-PSU variance is negative.

NOTE: National estimates are projected from Balanced Repeated Replication method calculations for the South Region. Program was developed by Bean and estimates were developed by Casady, of the National Center for Health Statistics.

Other variable costs were dependent on the sample size or the number of interviewers and could be estimated from recent historical data.

The cost-variance data showed that costs could be reduced by decreasing the number of sample PSU's and that the savings were more than sufficient to pay for the increase in the number of housing units needed to maintain the precision of most estimates at the level of the 1973–84 design. Designs that entirely eliminated travel between PSU's proved to be the least expensive of all the designs that maintained the reliability of the 1973–84 design.

Secondary sampling units

Secondary sampling units are clusters of housing units selected within sample PSU's. Since the beginning of the survey, clusters have been four, six, or nine housing units. They have been compact, that is, adjacent units for the most part, and noncompact, that is, spread over a larger geographical area such as a block. In the 1973-84 design, clusters in the address frame were compact with an expected size of four, and clusters in the area and new-construction frames were a noncompact four. Increasing the cluster size reduces travel costs but increases variances if the intraclass correlation is positive.

The change to clusters of size four came early in the 1970's after studies of unit costs and intraclass correlations for NHIS showed that, for constant costs, this would result in lower variances for health statistics than the larger cluster sizes considered. Actually, this was optimal only when taking into account the savings that would result from using the same cluster size in both CPS and NHIS. Considering only NHIS, the optimum cluster size would have been larger (Shapiro, 1971).

For the post-1980 census redesign, the Census Bureau updated the earlier cost study (Rogers, 1981). Costs were updated in three different ways. Regression techniques were used to estimate, first, total interviewer travel mileage for various cluster sizes, and second, unit costs for different cluster sizes. In the third method, a cost function was defined in terms of interviewing costs, costs of travel between clusters, and costs of travel between home and cluster.

The final step was to compare costs and within-PSU variances of items within a range of intraclass correlations and for different cluster sizes. All comparisons led to the conclusion that a larger cluster size would be more efficient for the NHIS. Table B summarizes the results (Huang, 1982).

Because data for some of the components of method 3's cost function were available only for clusters of size four, data for the other cluster sizes had to be estimated from existing data and reasonable assumptions. Because method 3's estimates were of lower quality than the other methods, the choice was eventually limited to the results of methods 1 and 2. Using clusters of size seven would have made coordination of the within-PSU sampling procedure with other Census Bureau surveys more complex, so it was

Table B. Optimum cluster sizes using 3 methods of estimating interviewing costs for different levels of intraclass correlation

Intraclass correlation	Method of estimating Interviewing costs[1]			
	1	2	3a	3b
0.02	8	8	12	9
0.03.	8	8	9	9
0.04.	7	8	9	9
0.05.	7	8	9	5
0.06.	6	7	9	5

[1] Methods are as follows:
1 = Mileage per household for various segment sizes estimated by regression, and travel costs added to interviewing costs. 2 = Total field costs estimated by regression. 3 = Field costs estimated from cost function with parameters estimated from available data In 3a, interviewing costs included only the time for interviewing. In 3b, other costs proportional to sample size (for example, processing) are added to interviewing costs.

dropped from further consideration, Several potential designs based on both six and eight households per cluster were evaluated through a sensitivity analysis presented in the next section.

As in the 1973-84 design, clusters of size four were recommended for new-construction units sampled from building permits. There were two reasons for this. As a rule, new-construction units are not close together. A new-construction cluster may be scattered over a relatively large area, so there is little to gain from using a larger cluster size. Interviewer work loads were a second consideration. NHIS interviewers would be better able to handle incremental work-load increases of four housing units as opposed to some larger number.

Sensitivity analysis of design parameters

Upon completion of the research on the number of PSU's and cluster size, the Census Bureau prepared 21 design alternatives to study the sensitivity and impact of changes in the design parameters on cost and variance, and to verify that the results from the models used in the research agreed with cost and variance estimates obtained from fully specified designs. The 21 designs are shown in table C.

The first six columns give the design number and the parameters of that design. Both the multiframe and area frame designs assumed that there would be a separate new-construction frame. The variance ratios shown in the next two columns were computed with values for the number of NSR PSU's and households and the cluster size substituted into formula (1). For the upper bound an intraclass correlation of 0.04 was used for all designs. This gave a conservative estimate of the change in variance. The lower bound accounted for changes in variances resulting from improvements in sampling methods. This was a more speculative estimate based on research for the NHIS and other Census Bureau surveys. The between-PSU portion was reduced by 20 percent to account for improvements in the stratification procedure. The intraclass correlation for the area designs was reduced to 0.035 because of the noncompact segments, and the within-PSU portion was reduced by an additional 1 percent to account for improved within-PSU sampling methods.

Table C. Estimated costs and variances for the 1985–94 National Health Interview Survey redesign alternatives

	Design parameter					Variance relative to 1973–84 design		Cost in thousands of dollars					
									Yearly operating costs				
Design alternative	Number of PSU'S	Primary sampling frame	Cluster size	House-holds	Inter-viewers	Upper bound	Lower bound	Total yearly costs	Variable cost	Fixed cost	Amortized phase-in cost	Amortized Implementa-tion cost	Additional travel costs due to use of non-compact **clusters**
1[1]	376	list	4	50,800	120	1.00	1.00	$4,449	$3,486	*$609*	$ —	$354	$. . .
2	376	area	4	50,800	120	1.00	0.98	4,647	3,563	*609*	–	475	42
3.	376	area	6	54,400	121	1.00	0.97	4,510	3,415	*609*	–	486	35
4.	376	area	8	58,100	130	1.00	0.97	4,635	3,531	*609*	10	485	32
5.	118	area	6	64,300	136	1.00	0.93	4,485	3,248	*609*	54	574	41
6.	129	area	8	67,000	152	1.00	0.93	4,509	3,271	609	69	560	37
7	200	area	6	57,600	131	1.00	0.94	4,570	3,398	609	49	514	37
8	200	area	6	57,600	131	1.00	0.97	4,552	3,398	609	31	514	37
9.	200	area	6	52,600	126	1.08	1.02	4,272	3,149	609	44	470	34
10.	200	area	6	51,500	125	1.10	1.04	4,210	3,098	609	43	460	33
11	200	area	6	50,000	124	1.13	1.07	4,150	3,053	609	42	446	32
12.	200	area	6	49,000	123	1.15	1.09	4,063	2,975	609	42	437	31
13.	200	area	8	61,400	137	1.00	0.94	4,653	3,476	609	55	513	34
14.	200	area	8	61,400	137	1.00	0.96	4,635	3,476	609	37	513	34
15.	200	area	8	53,100	131	1.13	1.07	4,180	3.078	609	49	444	29
16.	200	area	8	52,500	130	1.14	1.08	4,137	3,041	609	48	439	29
17.	200	area	8	51,500	129	1.16	1.10	4,100	3,014	609	47	430	28
18.	200	area	8	50,000	128	1.19	1.13	4,021	2,948	609	46	418	27
19.	222	area	6	56,700	112	1.00	0.95	4,585	3,439	609	31	506	36
20.	160	area	6	59,800	120	1.00	0.94	4,559	3,377	609	39	534	38
21	242	area	6	50,000	109	1.10	1.05	4,167	3.084	609	28	446	32

[1]Design used by the National Health Interview Survey from 1973 through 1984.

The remaining columns in table C show the costs for each design. The phase-in costs are the costs of recruiting, training, and initial observation of the additional interviewers who must be hired before the implementation of the new design. The implementation costs include research, the selection of up to 10 years of sample, which is a small portion of the costs, and all the preparatory work by field staff above that normally required to select the first year's sample, as well as the costs associated with the new-construction sample. The total phase-in and implementation costs have been spread over the lo-year life of the sample to facilitate comparisons between designs.

The last column in table C was included to give an estimate of the increase in travel due to using noncompact clusters of housing units. It was separated from yearly operating costs because of the very rough approximations used.

Designs 1 through 4, all using the PSU's of the 1973-84 NHIS, show the effect of moving to an area sample and then to clusters of six and eight housing units. Phase-in costs are low because only a few more new interviewers would be required. Implementation costs are also relatively low because use of the existing PSU's results in some savings.

Designs 5 and 6 are optimum designs under the constraint that the variance ratio upper bound equal 1.0. They are no-travel designs for cluster sizes six and eight. Every PSU would have a resident interviewer and there would be no need for travel to PSU's.

NCHS felt that to produce and analyze data for rural areas, a minimum of 200 PSU's was needed. Designs 7 through 12 are all 200-PSU designs with clusters of size six. Design 8 is the same as 7 except that the PSU's are a subset of the 376 PSU's of the 1973-84 NHIS. There is a savings in phase-in costs because fewer interviewers would be replaced, but the potential variance reductions (lower bound) are less. Designs 9 through 12 would save roughly $400,000 to $600,000; they do not maintain the reliability of the 1973-84 design. Designs 13 through 18 are analogous to 7 through 12 but have clusters of eight housing units and larger samples.

The last three designs were attempts to increase efficiency by considering how the annual interviewing work load is assigned. For example, in designs 7 through 12, interviewers in NSR PSU's would work in their home PSU's and one or two additional ones. Interviewers working in only one other PSU have less work than those working in two others and, therefore, may not be efficiently employed. If the sample size in design 7 could be reduced and the number of interviewers reduced by a proportionately greater amount so that each interviewer in NSR PSU's worked in exactly two PSU's requiring travel, and the number of PSU's were increased so that the variance ratio remained 1.0, then perhaps the total costs would also decrease. This strategy failed to work in design 19. Increased travel costs more than offset the savings in phase-in costs. The opposite approach, increasing the sample size and reducing the number of PSU's, worked better. Interviewers in NSR PSU's would work in their home PSU's and one other. The total costs for design 20 were less than those of design 7. Design 21 used the approach of design 19 but allowed a lo-percent variance increase. It compared favorably with design 10.

Selection of design parameters

Design 13 was ultimately selected by NCHS to be used during the 1985-94 decade. This design maintained the reliability of the 1973-84 design (upper bound); it had potentially greater improvements in reliability than designs 8 and 14, which used a subset of the 376 PSU's in the 1973-84 design; and it included a sufficient number of PSU's for rural estimates. Furthermore, design 13, with clusters of size eight, supported a larger sample size than design 7, which used clusters of size six. Larger sample cluster sizes would enhance the utility of the NHIS sample for other NCHS surveys of the same households to be conducted later in the decade.

Stratification variables

The purpose of stratification is to group PSU's into strata that are homogeneous with respect to characteristics of interest. This is expected to improve the precision of those and related characteristics. The 1973-84 NHIS shared the design of the CPS, including the PSU stratification, which was based primarily on demographic and labor force variables.

For the 1985–94 design, NHIS PSU's were stratified independently of other surveys. The best stratifiers would have been health variables, but health statistics were available only for sample PSU's and could not be used as stratifiers. Instead, variables that were highly correlated with health variables were sought for stratifiers. Relationships between health variables and the 1970 Census of Population and 1972 Economic Census variables that were available for all PSU's were examined. A set of 80 key health variables from the 1978 NHIS was identified by NCHS; 29 variables were selected from the 1970 Census of Population and 3 variables from the 1972 Economic Census. Data from the 1980 census were not available in time to be used during the research phase but were used in the actual stratification process.

Potential stratifiers had to satisfy four conditions:

1. Stratifying variables must be available for all PSU's.
2. They should be relatively highly correlated with health variables.
3. The correlations should be stable over a period of many years.
4. The stratifiers should be sufficiently variable between PSU'S.

The correlation analysis (Hsen, 1983b; Ponikowski, 1982, 1983c) produced a set of census variables that were most correlated with the NCHS key health variables. Because of the 6- to 8-year difference between the two data sets, any high correlations that were observed were considered to be sufficiently stable over time. Variables had to show sufficient fluctuation over the population PSU's to be considered useful. When potential stratifiers were highly correlated with each other, only one of them was kept. Percent Hispanic was added as a stratifier because of special interest in improving estimates for this subpopulation.

Table D. Correlations between 1978 health variables and stratification variables

Health variable	Stratification variable[1]					
	1	2	3	4	5	6
Bed days	-0.07	0.15	0.21	-0.05	0.12	-0.01
Persons with doctor visits	-0.32	-0.39	-0.21	0.20	-0.23	-0.01
Dental visits	-0.04	-0.25	-0.11	0.21	-0.10	-0.03
Persons with dental visits	-0.21	-053	-0.42	0.32	-0.34	-0.15
Persons with limitations of activity	-0.30	0.41	0.48	-0.34	0.41	0.27
Persons repotting poor or fair health	-0.26	0.70	0.52	-0.64	0.51	0.47
Work-loss days	-0.05	-0.03	-0.04	0.02	0.04	0.13
Currently employed persons	-0.06	-0.61	-0.48	0.38	-0.37	-0.06
Acute conditions	0.04	-0.12	-0.14	0.07	-6.07	-0.18
Restricted activity days	-0.09	0.18	0.06	-0.17	0.18	-0.02
Bed days due to acute conditions	-0.07	0.12	0.00	-0.03	0.09	-0.08
Hospital discharges	-0.12	0.31	0.40	-0.23	0.30	0.27
Doctor visits	-0.09	0.07	0.11	0.00	0.20	0.00
Persons with Medicaid	-0.07	0.49	0.24	-0.29	0.20	0.14
Persons under 65 years of age with no private insurance	0.23	0.45	0.25	-0.14	0.14	-6.24

'Stratification variables are the following: 1 — Hispanic; 2 — persons below poverty level; 3 — households with income less than $15,000 ($7,000 used for research); 4 — persons in urban areas; 5 — unemployed persons; 6 — persons employed in manufacturing.

Table D shows results of the correlation analysis for selected health variables and the final stratification variables.

Number of sample PSU's per stratum

In the 1973-84 design, one PSU was selected with probability proportionate to size from each NSR stratum. The between-PSU variance was estimated by collapsing "similar" strata so there would be two observations per superstratum. This biased the variance estimates by adding a between-stratum component. A two-PSU-per-stratum design would allow unbiased estimates of between-PSU variance.

The drawback to a two-PSU-per-stratum design was that only half as many strata would be formed as under a one-PSU-per-stratum design. There might be an increase in between-PSU variance because of a poorer stratification. For the 1985-94 design two PSU's were selected per NSR stratum based on research that indicated that any increase in the true between-PSU variance from a two-PSU-per-stratum design would probably be small (Moore, 1984; Ponikowski, 1983b).

PSU panels

One of the major requests from NCHS was that the Census Bureau study the feasibility of forming panels of PSU's for the 1985-94 NHIS (Sirken, Bryant, and Feldman, 1982). Panels of PSU's are subsets of sample PSU's that can be used alone or in combination to provide a national sample. Forming panels would facilitate linking other NCHS surveys to some or all of the NCHS sample design. Also, the panels could be used for sample reductions, as the remaining panels would form a national design. NCHS wanted to be able to select a subset of the full set of sample PSU's with the following properties:

- Each subset, or panel, is a representative sample of the U.S. population.
- Each NSR PSU is in only one panel.
- The larger self-representing (SR) PSU's are in all panels.
- The panels are equally distributed within census regions.

Because the 1985-94 design contained two PSU's per NSR stratum, a two-panel design would have been a natural division of the sample. However, a 50-percent sample reduction or subsample would be required to derive any benefits from the panel construction. Four panels would provide added flexibility. They could be constructed by combining pairs of NSR strata and assigning each PSU to a different panel. Then subsamples or reductions of approximately 25, 50, or 75 percent could be accommodated. Therefore, a decision was made to construct four panels in the 1985-94 design.

Overlap with other designs

Two possibilities for maximizing the overlap of NHIS PSU's with those of another design were investigated. One possibility was to maximize the overlap with the 1973-84 NHIS. This would give the greatest chance of retaining experienced interviewers. However, examination of this possibility showed that the number of interviewers retained would not have been much greater for maximum overlap than for independent selection.

The second possibility was to maximize the overlap with other Census Bureau surveys in order to share some of the sampling costs. However, because NHIS and the other surveys used different sampling techniques (area frame versus census address lists), NHIS would not be able to share very much of these costs. Furthermore, NHIS interviewers work exclusively on NHIS, so interviewing staffs could not be shared.

The advantages of maximizing overlap were small at best. One major disadvantage was that conditional probabilities for PSU selection would have to be computed in order to make unbiased between-PSU variance estimates. These probabilities would have been quite difficult, if not impossible, to compute. A decision to maximize overlap would have been inconsistent with the decision to choose two PSU's per stratum. Therefore, NHIS PSU's were chosen independently of other survey designs.

Selection of NHIS sample

PSU definitions

Primary sampling units in NHIS are generally single counties or, more frequently, groups of contiguous counties. In some States in New England, minor civil divisions (towns or townships) are used instead of counties. In some other States, county equivalents, such as parishes in Louisiana and independent cities in Virginia, are used.

The starting point for NHIS was the set of PSU's defined for the CPS redesign. Each metropolitan statistical area (MSA) or consolidated metropolitan statistical area (CMSA) in existence in 1983 was defined as a separate PSU. Metropolitan areas that crossed State borders had been split for the CPS because it was designed for State estimates. They were reunited for NHIS because of the greater interest in national estimates than in State estimates and to keep large MSA's intact for data analysis. Some other PSU's were redefined along natural boundaries when this allowed more efficient travel by the interviewer (Hsen, 1981b, 1982a; Kostanich, 1982).

PSU's are defined within a census region, with the exception of a few MSA's that cross regional boundaries. Approximately 1,900 PSU's were defined from the more than 3,000 counties and equivalents; after stratification, the sample PSU's were selected from these.

Stratification and selection of PSU's

Stratification was done within the four geographical regions by MSA or non-MSA status, using 1980 Census of Population data (Hsen, 1981a, 1983a, 1983b; Huang, 1982; Ponikowski, 1982, 1983a, 1983c). The computer program used for stratification was based on a clustering algorithm developed by Friedman and Rubin (1967). The procedure began with a random stratification within a specified number of strata. Pairwise exchanges were made between PSU's in different strata, and the resulting between-PSU variance was calculated. The criterion for comparing possible stratifications was the sum of between-PSU variances for the stratification variables-the smaller the sum, the better the stratification. For additional information on the stratification program, see Kostanich et al. (1981).

The sample design adopted for the NHIS requires that PSU's with the largest populations be designated as self-representing; each of the SR PSU's is treated as a separate stratum and is included in the sample with certainty. It also requires combination of the remaining PSU's into strata, referred to as NSR strata, with the selection of two PSU's from each stratum. The 52 largest MSA's were designated as self-representing; the rest of the Nation was divided into 73 strata. A summary of SR and NSR strata by region is given in table E.

Two PSU's per NSR stratum were selected. The selection methodology allowed sampling without replacement with probability proportionate to the projected 1985 population within the stratum (Ponikowski, 1983b).

Formation of PSU panels

Four national subdesigns, or panels, were constructed from the 1985-94 design. The panels were made approximately equal in terms of the number of sample housing units. NSR strata were paired within regions by combining

Table E. Number and population of strata for the 1985–94 National Health Interview Survey by census region

Type of PSU, stratum, and population	All regions	Northeast	Midwest	South	West
SR PSU'S					
Strata					
All strata	52	10	11	20	11
MSA strata	52	10	11	20	11
Strata outside MSA					
Population					
Total	124,949,213	34,229,347	27,869,858	32,364,581	30,485,427
Minimum	628,959	723,886	930,179	691,722	628,959
MSA population	124,949,213	34,229,347	27,869,858	32,364,581	30,485,427
Population outside MSA					
NSR PSU's					
Strata					
All strata	73	10	20	32	11
MSA strata	33	5	8	15	5
Strata outside MSA	40	5	12	17	6
Population					
Total	108,631,714	14,595,388	30,751,544	47,380,179	15,904,603
Average	1,466,106	1,459,539	1,537,577	1,480,630	1,455,873
MSA population	52,779,849	8,716,181	13,958,993	22,206,138	7,898,537
Population outside MSA	55,351,865	5,879,207	16,792,551	25,174,041	8,006,066

NOTE: PSU is primary sampling unit; SR is self-representing; NSR is nonself-representing; MSA is Metropolitan Statistical Area as defined by U.S. Office of Management and Budget.

strata that were similar with respect to the original stratification variables. Then the four PSU's in each combined stratum were randomly assigned to panels one through four.

To divide the SR area into panels, the PSU's were classified as small, medium, or large. Generally, small SR PSU's had a total population of less than 1.4 million persons; medium, between 1.4 and 2.4 million; and large, more than 2.4 million. Some exceptions were made so that the number of SR PSU's to be stratified within a region for panel assignment would be divisible by four.

In the 14 medium-size SR PSU's the weekly interviewer assignments were randomly divided into halves. The halves were treated as separate PSU's for panel assignment.

Small SR PSU's and the medium SR PSU halves were stratified within regions into groups of four using the same stratification variables used for NSR PSU's and a total population variable. The small SR PSU's were assigned to panels within the new strata, and each half-sample in a medium-size PSU was assigned to a different panel (Hsen, 1982b; Parsons, 1984a).

The weekly interviewer assignments in the 12 largest PSU's were randomly divided into four equal parts, each part being assigned to a different panel. Table F shows there were 12 large SR PSU's contained in all 4 panels; 14 medium SR PSU's, each in 2 panels; and 26 small SR and 146 NSR PSU's in 1 panel each.

Selection of SSU's

In the 1973-84 design, housing units in the Census Bureau's population surveys were selected in two steps. First, census enumeration districts (ED's), geographic areas defined for use in the 1970 Census of Population containing on average 350 housing units, were sorted according to geographical variables. ED's were sorted within CBUR categories (C, the central city of an MSA; B, the urbanized area not in category C; U, urban place, not an urbanized area and not in category C; and R, all other ED's) by identification number, a geographical code that tended to place contiguous ED's together in the sort. ED's were then sampled systematically with a random start and probability proportionate to size. The within-PSU sampling interval identified sample clusters within the ED.

The redesign provided an opportunity to evaluate new methods that would take into account specific needs of individual surveys and still avoid duplication of sample housing units in different Census Bureau surveys.

Census tracts, blocks, block groups, block faces, and housing units were considered as alternatives to ED's for the sort unit. Blocks are usually well-defined rectangular pieces of land bounded by streets, roads, streams, railroad tracks, or other features. Data limitations at the block level prevented evaluation of blocks as a sort unit. This also applied to block faces, the housing units on one side of a block, and to individual housing units. Use of block faces or housing units would also have increased the cost and complexity of sampling operations.

Table F. Distribution of self-representing (SR) and nonself-representing (NSR) primary sampling units (PSU's) by region and type

Region	Total number of PSU's	SR PSU'S			NSR PSU'S
		Large	Medium	Small	
All regions	198	12	14	26	146
Northeast	30	3	1	6	20
Midwest	51	3	4	4	40
South	84	4	4	12	64
West	33	2	5	4	22

Tracts were eliminated because research showed that most of the within-PSU variance was due to variability within ED's. Census tracts, geographical subdivisions of counties, are generally larger than ED's. They would be less effective in reducing within-PSU variance. Block groups, groups of blocks defined within tracts for the census, appeared to be similar in nature to ED's; however, ED's were more convenient because they provided coverage of the entire Nation, whereas blocks did not. Thus, ED's were retained as the within-PSU sort unit.

The computer program that was used to stratify PSU's was modified to cluster ED's using the variables given below. In order to increase stability of the procedure over time, the clustering was done within CBUR categories. This had the effect of imposing a CBUR sort on the ED's. The clusters were sorted first by CBUR categories, then by median value of owner-occupied housing if the population of the PSU was greater than 125,000 persons, or by proportion of owner-occupied housing if the PSU was smaller. Within a cluster, ED's were sorted by ED number, which is based on geography.

The variables used to cluster ED's were as follows:

1. CBUR.
2. Median value of owner-occupied housing.
3. Number of children under 6 years of age.
4. Total population aged 65 years and over.
5. Number of owner-occupied housing units.
6. Number of mobile homes or trailers.
7. Number of units lacking some or all plumbing.
8. Number of owner-occupied units of value less than $40,000.
9. Number of renter-occupied units with rent less than $200 per month.
10. Number of one-room housing units.
11. Total black population.
12. Total Hispanic population.

The last two variables were used only if the population was greater than 5 percent of the PSU population.

Research had shown this procedure to be superior to the sorting procedure used in the 1973-84 design. As most of the within-PSU variance was due to the variance within ED's, and because the stratification of ED's reduced only the variance between ED's (Kobilarcik, Statt, and Moore, 1982; Moore, 1982), the reductions in total within-PSU variances due to the new sort are small, but nontrivial.

Selection of housing units

Within a sample PSU, sample ED's are subdivided into blocks or other small land areas with well-defined boundaries. One such block or "chunk" is designated for the sample with probability proportionate to its estimated number of housing units. A Census Bureau interviewer visits the area well in advance of the scheduled interview date and creates a list of all the housing units in the sample block. Several subsamples are systematically drawn so as to form noncompact clusters of four housing units each. Clusters of four are formed in order to be compatible with other Census Bureau surveys. Then, for NHIS, clusters are paired to form the sample clusters of size eight. Samples for additional years can easily be identified at the same time.

In areas where a separate new-construction frame is used, an interviewer visits the permit-issuing office and lists all the building permits for residential housing issued during a specified time period. New-construction units are clustered in groups of four and systematically sampled. To avoid duplication between the area and new-construction frames, the interviewer finds out when the units in the area frame were built. Units built after a specified date are not interviewed unless selected from the new-construction frame. See U.S. Bureau of the Census (1978) for more information on the sampling of new construction.

The total number of housing units in the SR PSU's ranged from about 165 to about 4,500. In the NSR PSU's, the average sample size was nearly 200 housing units.

Oversampling the black population

The sample size of a subdomain can be increased by differential sampling within PSU's. This means that part of a PSU would be sampled at a higher rate (oversampled) than the rest of the PSU. For NHIS, PSU's with at least 5- but with less than 50-percent black population were divided into two unequal strata in such a way that the smaller stratum contained the higher concentration of the black population of the PSU. The sampling rate was increased in the smaller stratum and decreased in the larger stratum so that the total sample was the same as it would have been without differential sampling.

This was accomplished iteratively. In the first trial stratification, the smaller stratum included only one ED, the one with the largest percent black population. In successive iterations, the ED with the next largest percent was added to the smaller stratum. The process continued until either the smaller stratum would become the larger stratum if another ED were added, or until it contained all of the black population in the PSU.

For each pair of strata the following quantities were calculated. The index $i = 1$ represents the stratum with the high density black population; $i = 2$ represents the complement.

N_i = the population of stratum i ($i = 1, 2$)
t_i = the proportion of stratum i that is the black population
$\sigma_i^2(y)$ = the population variance for the total number of black persons in stratum i
$\sigma_i^2(z)$ = the population variance for a characteristic total of interest in stratum i (total heads of families was used in the research)

Without differential sampling, the sample size would be $n = r(N_1 + N_2)$, where r is the sampling rate within the PSU. To keep the same sample size with differential sampling, the following must hold:

$$r\,(N_1+N_2) = r_1 N_1 + r_2 N_2 \qquad (2)$$

where r_i is the rate for stratum i.

The following terms were also defined:

$$v = N_2/N_1,\ u = t_1/t_2,\ k = r_1/r_2$$

$$w = \sigma_1^2(y)/\sigma_2^2(y),\ c = \sigma_1^2(z) / \sigma_2^2(z)$$

Note that v, u, and k are all greater than or equal to 1.

Assuming simple random sampling within strata and ignoring the finite population correction factor, it can be shown that the ratio of the variance from the stratified sample to that of the unstratified sample (the design effect for differential sampling) for the estimate of the black population is

$$DE(y) = \frac{\sigma_1^2(y)\ (wu+kv)\ (k+v)}{\sigma^2(y)\ kw\ (u+v)\ (1+v)} \tag{3}$$

and that the design effect for the total population characteristic is

$$DE\ (z) = \frac{\sigma_1^2(z)\ (c+kv)\ (k+v)}{\sigma^2(z)\ kc\ (u+v)^2} \tag{4}$$

where $\sigma^2(y)$ and $\sigma^2(z)$ are population variances without regard to strata.

For a given stratification, all variables are determined except *k*. For a predetermined value of *DE(z)*, *k*, and then *DE(y)*, can be calculated.

Research showed that most of the potential reduction in *DE(y)* could be obtained with a value of 1.05 for *DE(z)*. Allowing increases to 1.10, 1.15, or more achieved only small additional reductions in *DE(y)*. For each trial stratification, *DE(z)* was set equal to 1.05. Then, the stratification giving the lowest value for *DE(y)* was used for sample selection.

In all, oversampling was used in 102 PSU's. According to preliminary estimates, the proportion of black persons in the redesigned sample is 16 percent, whereas without oversampling it would have been 11 percent (Chen, 1981; Mazur, 1983, 1984; Olsen, 1982; Tadros, 1982a, 1982b; Waksberg, 1973).

Randomization of assignments

In general, efficient work loads in a calendar quarter take the following into account. First, they give about 7 to 8 weeks of work for an interviewer. Second, they should not exceed 24 housing units per week. In the 1973-84 design, where clusters were an expected four housing units, the typical interviewer assignment comprised four clusters in a week. For the 1985-94 design, the clusters in a PSU scheduled for interview in a calendar quarter were combined into weekly assignments of two or sometimes three clusters each. Old-construction clusters were combined clerically so that, ordinarily, a week's interviews in a PSU would be in the same county or in two adjacent counties.

The scheduling of interviewer assignments for specific weeks is done so that each regional office will have approximately constant work loads from week to week, so that interviewers will work regularly without too many weeks between assignments, and so that weekly samples will be approximately equal with respect to the following characteristics (Parsons, 1984b; Tadros, 1984):

- The sample in each census region (Northeast, Midwest, South, and West) and in the Nation.
- The sample in SR and NSR areas in each region and in the Nation.
- The sample in CBUR categories in each region and in the Nation.

In addition to the operational benefits, this weekly balancing produced nationally representative weekly samples that could be accumulated over weeks, months, quarters, or years for estimation purposes.

Clusters of new-construction units are added to the weekly assignments as they occur. Sample balance is maintained to the extent possible, but restrictions on how these clusters are scheduled may lead to a slightly poorer balance than what existed at the beginning of the design. The main difficulty is that a single new-construction cluster in a PSU must be added to a week that already contains an assignment. If two new-construction clusters are to be added, they may be placed in a week in which the interviewer has no work in that or any other PSU.

References

Chen, I. 1981. HIS Redesign: Sampling Schemes to Improve Small Domain Estimates. Internal memorandum from Inez Chen to Rameswar P. Chakrabarty. May 28. U.S. Bureau of the Census. Washington, D.C.

Friedman, H. P., and J. Rubin. 1967. On some invariant criteria for grouping data. *J. Am. Stat. Assoc. 62: 1159-1178.*

Hsen, P. 1981a. HIS Redesign: Research on Order of PSU and Strata Formations. Internal memorandum from Robert T. O'Reagan to Gary M. Shapiro. June 29. U.S. Bureau of the Census. Washington, D.C.

Hsen, P. 1981b. HIS Redesign: Definition of PSU's—Number of SR PSU's Required. Internal memorandum from Rameswar P. Chakrabarty to Gary M. Shapiro. Nov. 4. U.S. Bureau of the Census. Washington, DC.

Hsen, P. 1982a. HIS Redesign: Definition of PSU's—Potential Problem Counties if these Counties Are Defined as PSUs. Internal memorandum for documentation. Oct. 25. U.S. Bureau of the Census. Washington, D.C.

Hsen, P. 1982b. HIS Redesign: Methods for Forming Panels of PSU's. Internal memorandum from Rameswar P. Chakrabarty to Gary M. Shapiro. Nov. 24. U.S. Bureau of the Census. Washington, DC.

Hsen, P. 1983a. HIS Redesign: Stability of County Level Data for Minority Stratifiers Within Four Urban Categories-South Region. Internal memorandum from Rameswar P. Chakrabarty to Gary M. Shapiro. Mar. 25. U.S. Bureau of the Census. Washington, D.C.

Hsen, P. 1983b. HIS Redesign: Followup of Correlation Analysis Between the Census Variables and the Health Variables. Internal memorandum for documentation. July 1. U.S. Bureau of the Census. Washington, D.C.

Huang, E. T. 1982. Report on Optimum Segment Size for Health Interview Survey. Internal memorandum. June 4. U.S. Bureau of the Census. Washington, DC.

Kobilarcik, E., R. Statt, and T. Moore. 1982. Within-PSU sampling for the Census Bureau's demographic surveys. *Proceedings of the Survey Research Methods Section,* American Statistical Association, pp. 205–210.

Kostanich, D. 1982. Decision on Single County vs. Multiple County PSU's—Meeting of October 22, 1982. Internal memorandum from Gary M. Shapiro to Charles D. Jones. Nov. 15. U.S. Bureau of the Census. Washington, D.C.

Kostanich, D., D. Judkins, R. Singh, and M. Schautz. 1981. Modification of Friedman-Rubin's clustering algorithm for use in stratified PPS sampling. *Proceedings of the Survey Research Methods Section,* American Statistical Association, pp. 285–290.

Mazur, C. 1983. HIS Redesign: Differential Sampling To Achieve a Reduction in Black Subgroup Variances. Internal memorandum from Rameswar P. Chakrabarty to Gary M. Shapiro. June 15. U.S. Bureau of the Census. Washington, DC.

Mazur, C. 1984. Documentation of Oversampling Within the Redesigned 1980 HIS PSU. Internal memorandum for documentation. April 25. U.S. Bureau of the Census. Washington, DC.

Moore, T. F. 1982. Modification to Recommended Procedure for ED Sort. Internal memorandum to Larry Cahoon. Oct. 27. U.S. Bureau of the Census. Washington, D.C.

Moore, T. F. 1984. HIS Redesign: Documentation of Decision To Select Two PSU's per Stratum. Internal memorandum from Gary M. Shapiro to Charles D. Jones. May 18. U.S. Bureau of the Census. Washington, D.C.

Moore, T. F. 1988. HIS Redesign: Derivation of Variance Formula for Comparing Alternate Designs. Draft based on internal memoranda. U.S. Bureau of the Census. Washington, DC.

National Center for Health Statistics, J. Waksberg, R. H. Hanson, and C. A. Jacobs. 1973. Reliability of estimates with alternative cluster sizes in the Health Interview Survey. *Vital and Health Statistics*. Series 2, No. 52. DHEW Pub. No. (HSM) 73-1326. Health Services and Mental Health Administration. Washington: U.S. Government Printing Office.

National Center for Health Statistics. 1982. Components of Variance From NHIS. Memorandum from Robert Casady to Rameswar P. Chakrabarty at U.S. Bureau of the Census. Apr. 1. Hyattsville, Md.

Olsen, C. 1982. HIS Redesign: Differential Sampling to Achieve a Reduction in Demographic Subgroup Variances. Internal memorandum from Rameswar P. Chakrabarty to Gary M. Shapiro. May 3. U.S. Bureau of the Census. Washington, D.C.

Parsons, V. 1984a. HIS Redesign: A Method for Forming Panels of PSU's Internal memorandum from Rameswar P. Chakrabarty to Gary M. Shapiro. Feb. 1. U.S. Bureau of the Census. Washington, D.C.

Parsons, V. 1984b. HIS Redesign: Final Scheduling and Balance of the Weekly HIS Assignments. Internal memorandum from Rameswar P. Chakrabarty to Gary M. Shapiro. July 2. U.S. Bureau of the Census. Washington, D.C.

Ponikowski, C. 1982. Results of Research on Selection of Variables for the Restratification of Nonself-representing PSU's for HIS. Internal memorandum from Rameswar P. Chakrabarty to Gary M. Shapiro. Dec. 18. U.S. Bureau of the Census. Washington, D.C.

Ponikowski, C. 1983a. HIS Redesign: Results of Correlation Study of Additional Census Variables With HIS Variables. Internal memorandum from Rameswar P. Chakrabarty to Gary M. Shapiro. Feb. 3. U.S. Bureau of the Census. Washington, D.C.

Ponikowski, C. 1983b. HIS Redesign: Request for Program for Selection of PSU's for HIS. Internal memorandum from Gary M. Shapiro to Robert T. O'Reagan. Mar. 23. U.S. Bureau of the Census. Washington, D.C.

Ponikowski, C. 1983c. HIS Redesign: Basic Decisions for Restratification of HIS. Internal memorandum from Rameswar P. Chakrabarty to Gary M. Shapiro. Aug. 12. U.S. Bureau of the Census. Washington, DC.

Rogers, P. 1981. HIS Redesign: Description of Cost Data From Field Division. Internal memorandum for documentation. June 30. U.S. Bureau of the Census. Washington, D.C.

Shapiro, G. 1971. Optimal Segment Size for Health Interview Survey. Internal memorandum from Joseph Waksberg to Monroe Sirken. March 2. U.S. Bureau of the Census. Washington, D.C.

Sirken, M., E. Bryant, and J. Feldman. 1982. Recommendations for the NHIS Redesign. Unpublished memorandum to Robert A. Israel. Dec. 2. National Center for Health Statistics. Hyattsville, Md.

Tadros, W. 1982a. HIS Redesign: Differential Sampling to Achieve a Reduction in Demographic Subgroup Variances, II. Internal memorandum from Rameswar P. Chakrabarty to Gary M. Shapiro. Sept. 22. U.S. Bureau of the Census. Washington, DC.

Tadros, W. 1982b. HIS Redesign: Estimating National Variance Changes if Black or Spanish Persons Are Oversampled. Internal memorandum from Rameswar P. Chakrabarty to Gary M. Shapiro. Oct. 8. U.S. Bureau of the Census. Washington, D.C.

Tadros, W., T. F. Moore, and R. P. Chakrabarty. 1982. Determining the optimum number of primary sampling units to be selected for the Health Interview Survey. *Proceedings of the Survey Research Methods Section,* American Statistical Association, pp. 217-222.

Tadros, W. 1984. HIS Redesign: Combining Sample Measures Into Weekly Assignments. Internal memorandum from Rameswar P. Chakrabarty to Gary M. Shapiro. June 4. U.S. Bureau of the Census. Washington, D.C.

U.S. Bureau of the Census. 1978. The Current Population Survey, design and methodology. *Technical Paper No. 40.* Washington: U.S. Government Printing Office.

Waksberg, J. 1973. The effect of stratification with differential sampling rates on attributes of subsets of the population. *Proceedings of the Social Statistics* Section, American Statistical Association, pp. 429-434.

Chapter IV. Estimation procedures for the 1985–94 NHIS

by Van L. Parsons, Ph.D., Office of Research and Methodology, National Center for Health Statistics

Introduction

The National Health Interview Survey (NHIS) is designed for the making of inferences about the civilian noninstitutionalized population of the United States. The NHIS program at the National Center for Health Statistics (NCHS) focuses its attention upon making inferences about the health of persons and households in the target population. This is accomplished by inflating the responses of each surveyed person or household in the NHIS, referred to as sample elementary units, to a national level. In either case, for each unit in the sample, a weight is assigned that permits estimation of population totals. An estimator, $\hat{X}$, for any given population characteristic total, X, can be expressed as a weighted sum over all sample elementary units, more precisely,

$$\hat{X} = \sum_{u} x(u)\ W_f(u) \qquad (1)$$

where u represents a sample elementary unit of the NHIS, $x(u)$ is the characteristic or response for unit u, and $W_f(u)$ is the NHIS national weight for unit u. This estimator is used to generate the NHIS estimates of totals, percents, and rates. In the sections that follow, the technical aspects of the procedures used in creating national weights, and the procedures used in estimating the variances of NHIS estimators, are discussed.

Conceptual design model for the NHIS

Complex estimation techniques are required for the NHIS because the survey is based upon a highly stratified multistage probability sample. The development of these techniques is facilitated by presenting the NHIS design, discussed in detail in the previous chapter, in a somewhat idealized framework. The conceptual NHIS design, presented below, will provide a structural model for the important design features.

The NHIS is a complex multistage probability sample with the following design structure: The primary sampling units (PSU's) of the universe are partitioned into 125 strata: 52 self-representing (SR) strata and 73 non-self-representing (NSR) strata. From this PSU universe, a sample of 198 PSU's is taken. All 52 PSU's from the SR strata are included in the sample with certainty; in each NSR stratum, two PSU's are chosen without replacement. The NSR sampling is done with probability proportional to relative size within the stratum using Durbin's (1967) procedure. In the sequel, for a given NSR stratum s, PSU selection probabilities will be denoted by π_{si}, the probability that sample PSU i, i = 1,2, is chosen, and $\pi_{s1,2}$, the joint probability that sample PSU's 1 and 2 are chosen.

For an SR stratum (or SR PSU) s, the certainty of selection will be expressed $\pi_{s1} = 1$.

Within each PSU, a sampling universe for second-stage sampling is created. As discussed in the section "Selection of NHIS sample" of chapter III, the eligible housing units of each sampled PSU are partitioned into as many as three substrata:

Substratum 1: All housing units from the building permit frame. For some PSU's this substratum is empty.

Substratum 2: All housing units in the area frame to be oversampled. It is anticipated that a large proportion of the population in this substratum will be black persons. For some PSU's this substratum will be empty.

Substratum 3: All housing units in the area frame not in substratum 2. As substrata 1 and 2 may both be empty, some PSU's have all their housing units in this stratum.

The housing units of each substratum are grouped into clusters to serve as secondary sampling units (SSU's). The expected cluster size is four housing units in substratum 1 and eight housing units in substrata 2 and 3.

A mean sampling interval (SI) for the U.S. target population of SSU's is determined. For a given sample PSU i within stratum s, the mean sampling interval is $SI \cdot \pi_{si}$, but different sampling rates are used within each of the three substrata. The sampling interval for substratum j can be expressed as $SI \cdot \pi_{si} \cdot c_j$ where $c_1 = 1$, $c_2 < 1$ (oversample), and $c_3 > 1$ (undersample). The values of c_2 and c_3 are chosen, in part, to satisfy the equation:

$$(M_2+M_3)/(SI\pi_{si}) = M_2/(SI\pi_{si}c_2)+M_3/(SI\pi_{si}c_3)$$

where M_2 and M_3 are the number of SSU's in substrata 2 and 3, respectively.

Using the appropriate sampling interval, the sample SSU's are selected by a systematic sample using a random start. Furthermore, the sampling of SSU's within each substratum is independent of the sampling of SSU's within any other substratum of any PSU.

For the NHIS all eligible households within a sampled cluster are targeted for sampling. If, however, a sampled SSU contains more households than the NHIS interviewer can survey within a week, then a probability sample of households will be taken. Although the NHIS can be considered as having three stages of sample selection, a fourth level of sample selection may be introduced by the administration of an NHIS supplemental survey. In general, these supplements select a probability sample of the household members previously chosen at the third stage of sampling.

National inflation weights

The NHIS estimator of a characteristic total as presented in equation (1) uses methodology based on the features of the complex multistage probability sample to define the national NHIS weight, W_f, for each sample elementary unit. This national weight is the product of up to four weighting factors:

1. Inverse of the probability of selection
2. Household nonresponse adjustment
3. First-stage ratio adjustment
4. Second-stage ratio adjustment (poststratification)

When the elementary unit is an individual, all four weighting factors define the person's final weight. Because the NHIS ratio adjustments are based on person characteristics, only the first two weighting factors are used to define the national household weight.

The first weight to be defined is the inverse of the probability of selection. The probability that a given person or household elementary unit is selected in the sample is determined by the product of the conditional probabilities of selection at each stage of sample selection: PSU within stratum, SSU within substratum of sampled PSU, household within SSU, and person within household (for supplemental surveys). The reciprocal of this product, say W_I, is the first component weight of W_f. Infrequently, this weight will be modified. In the rare case in which less than 25 percent of the households of an SSU are subsampled, the conditional probability of selection will be taken as 0.25. This truncation of a weighting factor usually occurs in fewer than 10 SSU's a year, and thus the biases introduced by such a modification should be small.

In an ideal hypothetical sampling situation having no nonsampling error components-for example, frame problems, nonresponse, or interviewer effects-equation (l), with W_I substituted for W_f,

$$\hat{X}_0 = \sum_u x(u)\ W_I(u) \qquad (2)$$

will provide an unbiased estimator for the population total, X. Such an estimator is referred to as a Horvitz-Thompson estimator.

Historically, usually less than 5 percent of all eligible NHIS sampled households do not respond. This household nonresponse will most likely bias the estimator of equation (2); consequently, the need for a weighting adjustment for household nonresponse is justified.

This need results in the second weight factor, the household nonresponse adjustment. Assuming that nonresponding households within an SSU have characteristics similar to those of the responding households in that same SSU, a natural way to define the household nonresponse adjustment weight is the ratio

HE/HR

where

HE = the number of eligible households within an SSU

HR = the number of responding households within an SSU

In the rare instances in which less than half of the eligible households in an SSU respond, this nonresponse adjustment is truncated to 2.0. Hence, the household nonresponse adjustment factor W_{nr} is defined

$$W_{nr} = \text{minimum (HE/HR , 2.0)}.$$

An illustration of this adjustment is given in table A.

The estimator produced by substituting the product of W_I and W_{nr} for W_f in equation (1),

$$\hat{X}' = \sum_u x(u)\ W_I(u)\ W_{nr}(u) \qquad (3)$$

should produce approximately unbiased estimators for the population total, X. The weight $W_I \cdot W_{nr}$ is used to define the final national weight, W_h, for households and is used to produce all NHIS estimates based upon household characteristics. For example, an estimate of the number of households below the poverty threshold that have a member with limited activity status due to a health problem would be computed using this weight. When person characteristics are to be estimated, further weighting adjustments lead to improved national estimates.

The third and fourth weighting factors to be defined are ratio adjustments. Statistical sampling theory (Cochran, 1977) has demonstrated that in many situations the estimators obtained by a ratio estimation procedure are often more reliable than the basic inflation weight estimators expressed by equation (2). More precisely, if X' and Y' are basic inflation weight estimators of two population characteristic totals, X and Y, respectively, and the "true" total Y is known, then the ratio estimator $X'' = (X'/Y')\,Y$ will often

Table A. Adjustment for nonresponse

SSU number[1]	Households eligible for infer-view	Eligible households not responding	SSU adjustment factor[1]
1	6		1.0000
2	6	1	1.2000
3	8		1.0000
4	4	3	2.0000

[1]SSU stands for secondary sampling unit.

Table B. First-stage factors for use with the full National Health Interview Survey sample, by region, residence, and whether other than black or black

Residence and whether other than black or black	Region			
	Northeast	Midwest	South	West
MSA[1]				
Other than black. . .	1.044548	1.038811	1.028855	1.018131
Black	0.820819	1.032812	1.031638	0.849872
Not-MSA[1]				
Other than black. . .	0.952168	0.968367	[2]0.983512	0.987550
Black	1.045485	1.300000	[3]0.942102	1.025181

[1]MSA = metropolitan statistical area as defined by U.S. Office of Management and Budget.
[2]0.983957 beginning January 1990 to account for replacement PSU's.
[3]0.940309 beginning January 1990 to account for replacement PSU's

Table C. 60 age-sex-race cells In the National Health Interview Survey

Age	Black		Other than black	
	Male	Female	Male	Female
Under 1 year	x	x	x	x
1–4 years	x	x	x	x
5–9 years	x	x	x	x
10–14 years	x	x	x	x
15–17 years	x	x	x	x
18–19 years	x	x	x	x
20–24 years	x	x	x	x
25–29 years	x	x	x	x
30–34 years.	x	x	x	x
35–44 years.	x	x	x	x
45–49 years.	x	x	x	x
50–54 years.	x	x	x	x
55–64 years.	x	x	x	x
65–74 years.	x	x	x	x
75 years and over . . .	x	x	X	X

be a more reliable estimator of X than is X' whenever there is a high positive correlation between X' and Y', and the sample size is large.

The ratio adjustment is also used to help correct survey bias due to undercoverage. Historically, there has been survey undercoverage of persons in the NHIS, and the estimator of equation (3) will, in general, provide underestimates for many target population characteristic totals. This bias due to undercoverage is often reduced by the use of the ratio adjustments.

The first-stage ratio adjustment is used in an attempt to reduce the between-PSU variance component of sampling variation among the NSR PSU's. For each of four race-residence classes within the NSR strata of each of the four geographical regions, the first-stage ratio adjustments are defined using the following methodology (table B).

The 16 race-residency-region classes defined within the NSR PSU's will be indexed by the letter c, $c = 1,2,. . . ,16$. Let

Z_c = the projected 1985 census total for class c over all NSR PSU's in the population.

Considering only the first stage of sample selection as presented in the conceptual design, the sample of NHIS NSR PSU's leads to an unbiased estimate of Z_c, say $\hat{Z}_c$,

$$\hat{Z}_c = \sum_{(s:\ \text{NSR stratum})} \sum_{i=1}^{2} Z_{sic} / \pi_{si}$$

where Z_{sic} is the projected 1985 census total for class c in sample NSR PSU i of stratum s.

The first-stage ratio adjustment factor associated with class c is defined as

$$F_c = Z_c / \hat{Z}_c$$

with the value being truncated if it falls outside the interval [0.7,1.3]. Now, the only sampled individuals who receive the first-stage adjustment as a weighting factor are those individuals who reside in housing units of substrata 2 or 3 in the NSR PSU's. Those individuals who reside in housing units of substratum 1, the building permit frame, of the NSR PSU's are not adjusted, as those housing units are not included in the projected 1985 census NSR class totals. A universal first-stage ratio adjustment, W_{r1}, can be defined for each sample person by defining a new class index, c = 0, to denote all persons not receiving the F_c ratio adjustment:

$$W_{r1} = W_{r1}(c) = F_c \qquad \text{if } c = 1,2,\ldots,16$$
$$= 1 \qquad \text{if } c = 0$$

A first-stage ratio-adjusted national estimator, $\hat{X}''$, of a population total, X, is defined by equation (1) when the weight W_f is replaced with the product of the first three component weights: $W_I \cdot W_{nr} \cdot W_{r1}$,

$$\hat{X}'' = \sum_u x(u)\ W_I(u)\ W_{nr}(u) W_{r1}(u) \tag{4}$$

About 55 percent of the NHIS sample is collected from residences that do not receive the first-stage ratio adjustment. Preliminary research has shown that the inclusion of the first-stage ratio adjustment factor will have very little impact upon NHIS estimates (Parsons and Casady, 1987).

The main advantages of the ratio-estimation process are exploited by the introduction of the second ratio factor, the poststratification adjustment weight. This weight assures that the NHIS estimates for 60 age-sex-race classes of the civilian noninstitutionalized population of the United States (shown in table C) agree with independently determined controls prepared by the U.S. Bureau of the Census. Furthermore, the independent controls are the same ones used for the Current Population Survey (CPS). Thus, the national population estimates for any combination of the age-sex-race groups from the two surveys are the same, which greatly enhances comparability of the two surveys.

Each month the U.S. Bureau of the Census produces national estimates for the age-sex-race classes. Although the NHIS is conducted weekly, the poststratification adjustment is computed only for NHIS quarterly estimates. The

NHIS quarters and the dates of the estimates are as follows:

NHIS quarter	Population estimates
January–March.	February 1
April–June	May 1
July–September	August 1
October–December.	November 1

Each quarter, 60 age-sex-race adjustment weights are computed. If a represents 1 of the 60 age-sex-race classes, $Y(a)$ represents the U.S. Bureau of the Census estimate for class a, and $Y''(a)$ represents the NHIS first-stage ratio-adjusted national total for class a,

$$\hat{Y}''(a) = \sum_u I_a(u)\, W_1(u)\, W_{nr}(u) W_{r1}(u)$$

where

$$I_a(u) = 1 \text{ if person } u \text{ is in class } a$$
$$= 0 \text{ otherwise}$$

then the second-stage ratio adjustment weight for class a, W_{r2}, is defined as

$$W_{r2} = Y(a) / \hat{Y}''(a)$$

The two-stage ratio-adjusted national estimator, $\hat{X}$, of a population total, X, is defined by equation (1) with the weight W defined by the product of the four component weights: $W_1 \cdot W_{nr} \cdot W_{r1} \cdot W_{r2}$

$$\hat{X} = \sum_u x(u)\, W_1(u)\, W_{nr}(u) W_{r1}(u) W_{r2}(u) \qquad (5)$$

Variance estimation

Most of the estimates produced from the NHIS are totals and ratios of totals, such as means, percents, and rates. All such totals are produced using the final national inflation weight as described in the preceding sections. These estimates are subject to both sampling and nonsampling errors. The nonsampling errors, such as response errors, defective sample frames, nonresponse, and undercoverage, are difficult quantities to measure, but every effort is made to minimize such errors at each step of the NHIS operation. The sampling error, however, can be measured by the variance of the estimator.

Equation (5), which generates the NHIS estimates, permits very simple computation of the estimates; however, the variances of such estimators are more difficult to compute. The functional form of a variance estimator depends upon the nature of the survey design and the first- and second-order estimation procedures. For the NHIS both the multistage, unequal probability design and the ratio adjustments complicate variance estimation methodology. Wolter (1985) and Rust (1985) present excellent reviews of variance estimation for complex surveys. Of the available methods, NCHS has chosen the Taylor-linearization methods for use with the current design. Prior to the 1985 NHIS, the Balanced Repeated Replication (BRR) method was used for variance estimation.

There are three primary reasons for this decision. First, the new design called for two PSU's to be selected without replacement from each NSR stratum. The NCHS BRR computer procedure (NCHS, 1977) assumes that the PSU's are chosen independently. This lack of independence would complicate the BRR procedure. Second, the linearization methodology is computationally simple when applied to NHIS estimators. Finally, the linearization method gives analysts greater flexibility in studying the operational and statistical aspects of the survey design and the estimation methodology.

Variance estimation for totals

The conceptual design for the NHIS permits the development of variance formulas for the NHIS estimators. The first step is to define a variance estimator for an NHIS estimator of a total where the sample weight is the inverse of the probability of selection. Such a total is that expressed by X_0 in equation (2). This defines a linear estimator, and its theoretical variance and an estimator of this variance can both be expressed in a mathematically tractable form. The theory for linear estimators will then lead to variance approximations for ratio-adjusted totals.

Consider the general form of the four estimators, $\hat{X}_0$, $\hat{X}'$, $\hat{X}''$, $\hat{X}$, given by equations (2) through (5):

$$\hat{X}_G = \sum_u x(u)\, W_G(u) \qquad (6)$$

where $\hat{X}_G$ is a generic total based upon the generic weight W_G.

In order to present variance formulas, some notation and definitions are needed.

The following indexes will be used to denote the nesting within the NHIS design:

s = stratum
i = PSU
j = substratum
k = SSU

For substratum j within sampled PSU i within stratum s let

n_{sij} = the number of SSU's sampled

$\hat{Y}_{sijk}$ = the aggregate of a characteristic over all sample elementary units within SSU k, $k = 1,2,\ldots,n_{sij}$, where each unit's response, x, is inflated by W_G; this total is computed by equation (6) when restricting units to one SSU

$$\bar{\hat{Y}}_{sij.} = \sum_{k=1}^{n_{sij}} \hat{Y}_{sijk} / n_{sij}, \text{ a substratum sample mean}$$

$$\hat{S}^2_{sij} = \sum_{k=1}^{n_{sij}} (\hat{Y}_{sijk} - \bar{\hat{Y}}_{sij.})^2 / (n_{sij} - 1), \text{ a substratum variance estimator}$$

$$\hat{Y}_{si..} = \sum_{j=1}^{3} \sum_{k=1}^{n_{sij}} \hat{Y}_{sijk}, \text{ a PSU estimator of total.}$$

An estimator for the variance of $\hat{X}$, discussed by Parsons and Casady (1987), is

$$\hat{VAR}(\hat{X}) = \sum_{(s:\ NSR\ stratum)} \frac{(\pi_{s1}\pi_{s2} - \pi_{s12})(\hat{Y}_{s1..} - \hat{Y}_{s2..})^2}{\pi_{s12}} + \sum_{(s:\ NSR\ stratum)} \sum_{i=1}^{2} \pi_{si} \sum_{j=1}^{3} n_{sij} \hat{S}^2_{sij} + \sum_{(s:\ SR\ stratum)} \sum_{j=1}^{3} n_{s1j} \hat{S}^2_{s1j} \qquad (7)$$

Based upon theoretical and empirical evidence (Parsons and Casady, 1987) and (Casady et al., 1987), the following conclusions can be drawn about using equation (7) as an estimator for the variance of a total. All conclusions are based upon the conceptual NHIS design discussed earlier.

For the estimator $\hat{X}_0$, equation (6) with $W_G = W_I$, the variance equation (7) will provide a slight overestimate (in terms of mathematical expectation) of the theoretical variance of X_0 whenever there is a 100-percent response rate.

For the estimator $\hat{X}''$, equation (6) with $W_G = W_I W_{nr}$, the variance equation (7) will provide a slight overestimate (in terms of mathematical expectation) of the theoretical variance $\hat{X}'$ if the household nonresponse adjustment can be treated as the inverse of a probability of selection. This artificial stage can be considered as the selection of responding households by means of a probability sample within the SSU. Furthermore, it must be assumed that person response within responding households is 100 percent.

The estimator $\hat{X}''$, equation (6) with $W_G = W_I \cdot W_{nr} \cdot W_{r1}$, provides NHIS estimates almost identical to those produced by X'. The introduction of the first-stage ratio adjustment appears to have little impact, and the variance equation (7) should be adequate.

The estimator $\ddot{X}$, equation (6) with $W_G = W_I \cdot W_{nr} \cdot W_{r1} \cdot W_{r2}$, provides NHIS person estimates. The use of the variance equation (7) will usually overinflate the sampling variance. For example, because estimates of poststratification age-race-sex classes are ratio-adjusted, the variances of such estimators are zero. The use of equation (7) may lead to nonnegligible variances.

The NHIS estimator X of equation (5) is not a linear estimator; therefore, a mathematically tractable expression for its variance does not exist. Woodruff (1971), however, has suggested a procedure in which a nonlinear estimator of total is "linearized" by a Taylor series approximation; and then equation (7) is applied to this linear form to produce a variance estimator for the original nonlinear total.

Variances for ratio-adjusted NHIS totals

Of the two ratio adjustments, only the second stage has any significant impact on the NHIS variance estimates (NCHS, 1977; Parsons and Casady, 1987). Consequently, for variance estimation purposes, the NHIS can be treated as having only one ratio adjustment, the second stage. With this in mind, the following methodology can be used to compute variances for NHIS person totals.

Let X be the NHIS ratio-adjusted estimator of a person population total as presented in equation (5). An estimator of the variance of X can be computed by the following procedure:

1. For each person unit define the W_G weight as

   $$W_G = W_I \cdot W_{nr} \cdot W_{r1}$$

2. For each poststratification cell, a, compute

   $$\hat{T}_x(a) = \sum_{u\ in\ a} x(u)\, W_G(u) \text{ and } \hat{T}_1(a) = \sum_{u\ in\ a} 1\, W_G(u)$$

These are first-stage ratio estimators for the characteristic count and the population count, respectively, in each poststratification cell.

3. Each person unit, u, has a second-stage ratio adjustment factor, $W_{r2}(a)$, corresponding to its class, a. For this unit define the transformation:

   $$x(u{:}L) = W_{r2}(a)\ [\,x(u) - \hat{T}_x(a)/\hat{T}_1(a)\,]$$

4. The estimator of variance for $\hat{X}$ is defined by equation (7), but now the Y and $\hat{S}^2$ terms are computed using the W_G weight specified in step 1 and replacing the elementary unit's response, x(u), with the transformed value $x(u{:}L)$ as specified in step 3.

Further discussion of this transformation has been documented (Casady et al., 1987; Parsons and Casady, 1987).

Variances for ratios of totals

The linearization procedures easily extend to the ratios of two NHIS totals. Suppose that X_n and X_d are two NHIS ratio-adjusted-totals, and the ratio of the totals is expressed as $R = X_n / X_d$. The variance estimator can be defined by an extension of the algorithm above.

1. Steps 1, 2, and 3, above, are followed to produce linear transformations for the numerator and denominator totals, $x_n(u{:}L)$ and $x_d(u{:}L)$, respectively.
2. The transformation for the ratio $\hat{R}$ is then defined:

   $$x(u{:}L) = [x_n(u{:}L) - \hat{R}\, x_d(u{:}L)\,] / \hat{X}_d$$

3. Step 4, above, is followed but uses $x(u{:}L)$ as just computed.

Equation (7) and the linearization transformations are easily programmed.

Presentation of variances in NHIS publications

Many of the NHIS publications contain a very large number of tabular estimates. Individual estimates of variance are usually not computed for all tabular estimates, but instead, special variance tables and charts, which allow the reader to approximate variances, are often included in the NHIS publications.

As mentioned earlier, most NHIS estimates are totals or ratios of totals. Hansen et al. (1953) present a procedure that permits the approximation of variances for a large number of estimated totals. This methodology uses two assumptions.

First, if $\hat{X}$ is an estimator of a population total, X, and $\hat{\sigma}^2(\hat{X})$ is an estimator of the theoretical variance of $\hat{X}$, then

$$\hat{\sigma}^2(\hat{X})/\hat{X}^2 = a + b/X + e$$

where a and b are unknown parameters, and

$$E(e) = 0 \text{ and variance}(e) = [a + b/X]^2$$

Second, estimates of total can be grouped by classes, each class having the same a and b parameters.

If a large number of estimates of totals can be grouped into classes satisfying the second assumption, then regression techniques can be used to estimate the parameters a and b for the class.

NCHS uses this procedure in its Current Estimates reports. First, the published totals are grouped by the main variable of interest, for example, doctor visits, bed days, and acute conditions. All subdomain totals are listed with the main variable. Next, a representative sample of totals is selected from the group. The methodology discussed in the previous sections is then used to compute variances for these sample totals. Finally, iterative regression techniques are used to fit the model of the first assumption to these computed values. The ultimate result is the estimation of the a and b parameters for the group from which the representative sample was taken.

As an illustration, in the 1985 Current Estimates report (NCHS, 1986) it is reported that for the health variable "total number of doctor visits" the a parameter is 0.000082 and the b parameter is 172,892.1. These parameters were estimated by using a representative sample of 64 doctor visit totals based on different subdomains. A partial listing of these doctor visit totals and their estimated coefficients of variation appear in table D.

These parameters are presented in the NHIS publications, and rules explaining their usage are given. The rules also explain how to estimate the variances of ratios of totals using the computed a and b parameters.

References

Casady, R. J., V. L. Parsons, and C. B. Snowden. 1987. Simplified National Health Interview Survey design and impact upon variance estimation. *Proceedings of the Section on Survey Research Methods,* American Statistical Association, pp. 412-417.

Cochran, W. G. 1977. *Sampling Techniques,* 3d ed. New York: John Wiley and Sons.

Durbin, J. 1967. Design of multi-stage surveys for the estimation of sampling errors. Appl. Stat. 16: 152-164.

Hansen, M. H., W. M. Hurwitz, and W. G. Madow. 1953. *Sample Survey Methods and Theory,* Vol. I, Methods and Application; Vol. II, Theory. New York: John Wiley and Sons.

National Center for Health Statistics, G. Jones. 1977. HES Variance and Cross Tabulation Program, Version 2. Unpublished documentation. Mar. 10. Hyattsville, Md.

National Center for Health Statistics, A. J. Moss and V. L. Parsons. 1986. Current estimates from the National Health Interview Survey: United States, 1985. *Vital and Health Statistics.* Series 10, No. 160. DHHS Pub. No. (PHS) 86-1588. Public Health Service. Washington: U.S. Government Printing Office.

Parsons, V. L., and R. J. Casady. 1987. Variance estimation and redesigned Health Interview Survey. *Proceedings of the Section on Survey Research Methods,* American Statistical Association, pp. 406-411.

Rust, K. 1985. Variance estimation for complex estimators in sample surveys. *J. 0ffic. Stat.* 1: 381-397.

Wolter, K. M. 1985. *Introduction to Variance Estimation.* New York: Springer-Verlag.

Woodruff, R. S. 1971. A simple method for approximating the variance of a complicated estimate. *J. Am. Stat. Assoc. 66:411–414.*

Table D. Coefficient of variation (CV) for doctor visit totals by selected subdomains

Subdomain	Estimate of total	Estimated cv [1]	Smoothed cv [2]
All ages	1,230.868	1.32	1.49
Currently employed	481,816	2.03	2.10
While, family income $9,000–$19,000	209,880	3.77	3.01
Living in MSA,[3] 55–64 years of age	118.396	4.16	3.93
Male, 65-74 years of age	53,246	4.80	5.77
West Region, 18-24 years of age	27,892	8.80	7.93
Family income $9,000–$13.000, 25–34 years of age	16,256	11.04	10.35
Unemployed, 18-24 years of age	7,175	13.34	15.55
Black, family income under $9,000, under 5 years of age	4,740	14.60	19.12

[1] Computed using linearization.
[2] Computed using formula $a + b/X$.
[3] MSA is metropolitan statistical area as defined by U.S. Office of Management and Budget.

Vital and Health Statistics series descriptions

SERIES 1. **Programs and Collection Procedures**-Reports describing the general programs of the National Center for Health Statistics and its offices and divisions and the data collection methods used. They also Include definitions and other material necessary, for understanding the data.

SERIES 2. **Data Evaluation and Methods Research**-Studies of new statistical methodology including experimental tests of new survey methods. studies of vital statistics collection methods, new analytical techniques, objective evaluations of reliability of collected data, and contributions to statistical theory. Studies also include comparison of U.S. methodology with those of other countries.

SERIES 3. **Analytical and Epidemiological Studies**-Reports presenting analytical or Interpretive studies based on vital and health statistics, carrying the analysis further than the expository types of reports in the other series.

SERIES 4. **Documents and Committee Reports**—Final reports of major committees concerned with vital and health statistics and documents such as recommended model vital registration laws and revised birth and death certificates.

SERIES 5. **Comparative International Vital and Health Statistics Reports**-Analytical and descriptive reports comparing U.S. vital and health statistics with those of other countries.

SERIES 6. **Cognition and Survey Measurement**-Reports from the National Laboratory for Collaborative Research in Cognition and Survey Measurement using methods of cognitive science to design, evaluate, and test survey Instruments.

SERIES 10. **Data From the National Health Interview Survey**—Statistics on illness, accidental injuries, disability. use of hospital, medical, dental. and other services, and other health-related topics, all based on data collected in the continuing national household interview survey.

SERIES 11. **Data From the National Health Examination Survey and the National Health and Nutrition Examination Survey**—Data from direct examination, testing, and measurement of national samples of the civilian noninstitutionalized population provide the basis for (1) estimates of the medically defined prevalence of specific diseases in the United States and the distributions of the population with respect to physical, physiological, and psychological characteristics and (2) analysis of relationships among the various measurements without reference to an explicit finite universe of persons.

SERIES 12. **Data From the Institutionalized Population Surveys**—Discontinued in 1975. Reports from these surveys are Included in Series 13.

SERIES 13. **Data on Health Resources Utilization**—Statistics on the utilization of health manpower and facilities providing long-term care, ambulatory care, hospital care, and family planning services.

SERIES 14. **Data on Health Resources: Manpower and Facilities**—Statistics on the numbers, geographic distribution, and characteristics of health resources Including physicians, dentists, nurses, other health occupations, hospitals, nursing homes, and outpatient facilities.

SERIES 15. **Data From Special Surveys**—Statistics on health and health-related topics collected in special surveys that are not a part of the continuing data systems of the National Center for Health Statistics.

SERIES 16. **Compilations of Advance Data From Vital and Health Statistics**-These reports provide early release of data from the National Center for Health Statistics' health and demographic surveys. Many of these releases are followed by detailed reports in the Vital and Health Statistics Series.

SERIES 20. **Data on Mortality**-Various statistics on mortality other than as included in regular annual or monthly reports. Special analyses by cause of death. age, and other demographic variables; geographic and time series analyses; and statistics on characteristics of deaths not available from the vital records based on sample surveys of those records.

SERIES 21. **Data on Natality, Marriage, and Divorce**—Various statistics on natality, marriage, and divorce other than as included in regular annual or monthly reports. Special analyses by demographic variables; geographic and time series analyses; studies of fertility; and statistics on characteristics of births not available from the vital records based on sample surveys of those records.

SERIES 22. **Data From the National Mortality and Natality Surveys**—Discontinued in 1975. Reports from these sample surveys based on vital records are Included in Series 20 and 21, respectively.

SERIES 23. **Data From the National Survey of Family Growth**—Statistics on fertility, family formation and dissolution, family planning, and related maternal and infant health topics derived from a periodic survey of a nationwide probability sample of women 15-44 years of age.

SERIES 24. **Compilations of Data on Natality, Mortality, Marriage, Divorce, and Induced Terminations of Pregnancy**-Advance reports of births. deaths, marriages, and divorces are based on final data from the National Vital Statistics System and are published annually as supplements to the Monthly Vital Statistics Report (MVSR). These reports are followed by the publication of detailed data in Vital Statistics of the United States annual volumes. Other reports including induced terminations of pregnancy issued periodically as supplements to the MVSR provide selected findings based on data from the National Vital Statistics System and may be followed by detailed reports in Vital and Health Statistics Series.

For answers to questions about this report or for a list of titles of reports published in these series, contact:

Scientlfic and Technical Information Branch
National Center for Health Statistics
Centers for Disease Control
Public Health Service
Hyattsville, Md. 20782

301-436-8500

U.S. DEPARTMENT OF HEALTH AND
HUMAN SERVICES
Pubic Health Service
Centers for Disease Control
National Center for Health Statistics
3700 East-West Highway
Hyattsville, Maryland 20782

OFFICIAL BUSINESS
PENALTY FOR PRIVATE USE, $300

BULK RATE
POSTAGE & FEES PAID
PHS/NCHS
PERMIT NO. G-281

DHHS Publication No. (PHS) 89-1364. Series 2. No. 110