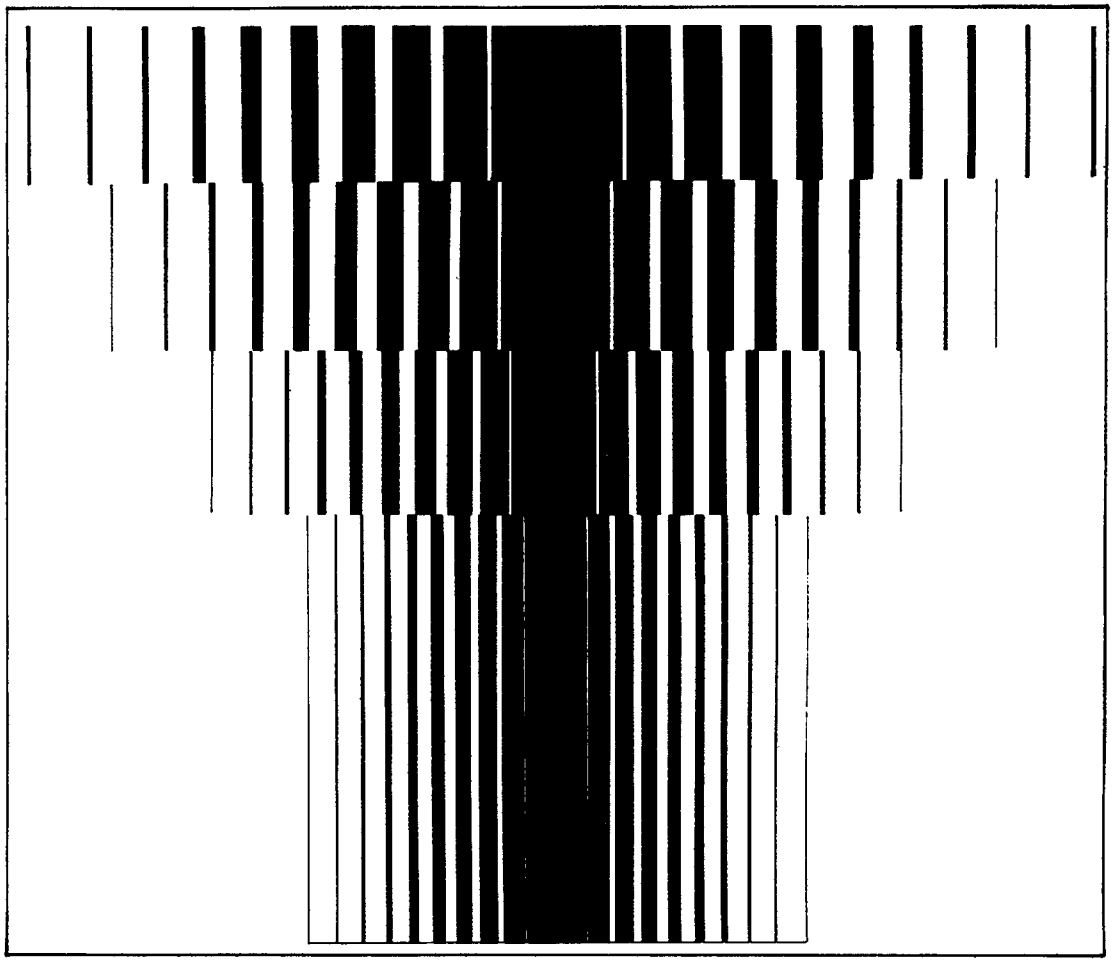
Statistical Methodologies for Analyzing a Complex Sample Survey

Series A, Methodological Report No. 4



U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES

Published by
Public Health Service
Centers for Disease Control
National Center for Health Statistics

March 1988

National Medical Care Utilization and Expenditure Survey

The National Medical Care Utilization and Expenditure Survey (NMCUES) is a unique source of detailed national estimates on the utilization of and expenditures for various types of medical care. NMCUES is designed to be directly responsive to the continuing need for statistical information on health care expenditures associated with health services utilization for the entire U.S. population.

NMCUES will produce comparable estimates over time for evaluation of the impact of legislation and programs on health status, costs, utilization, and illness-related behavior in the medical care delivery system. In addition to national estimates for the civilian noninstitutionalized population, it will also provide separate estimates for the Medicaid-eligible populations in four States.

The first cycle of NMCUES, which covers calendar year 1980, was designed and conducted as a collaborative effort between the National Center for Health Statistics, Public Health Service, and the Office of Research and Demonstrations, Health Care Financing Administration. Data were obtained from three survey components. The first was a national household survey and the second was a survey of Medicaid enrollees in four States (California, Michigan, Texas, and New York). Both of these components involved five interviews over a period of 15 months to obtain information on medical care utilization and expenditures and other health-related information. The third component was an administrative records survey that verified the eligibility status of respondents for the Medicare and Medicaid programs and supplemented the household data with claims data for the Medicare and Medicaid populations.

Data collection was accomplished by Research Triangle Institute, Research Triangle Park, N.C., and its subcontractors, the National Opinion Research Center of the University of Chicago, Ill., and SysteMetrics, Inc., Berkeley, Calif., under Contract No. 233–79–2032.

Co-Project Officers for the Survey were Robert R. Fuchsberg of the National Center for Health Statistics (NCHS) and Allen Dobson of the Health Care Financing Administration (HCFA). Robert A. Wright of NCHS and Larry Corder of HCFA also had major responsibilities. Daniel G. Horvitz of Research Triangle Institute was the Project Director primarily responsible for data collection, along with Associate Project Directors Esther Fleishman of the National Opinion Research Center, Robert H. Thornton of Research Triangle Institute, and James S. Lubalin of SysteMetrics, Inc. Barbara Moser of Research Triangle Institute was the Project Director primarily responsible for data processing.

For sale by the Superintendent of Documents, U.S. Government Printing Office
Washington, D.C. 20402

Copyright Information

All material in this report is in the public domain and may be reproduced or copied without permission; citation as to source, however, is appreciated.

Suggested Citation

National Center for Health Statistics, J. M. Lepkowski, J. R. Landis, P. E. Parsons, and S. A. Stehouwer: Statistical methodologies for analyzing a complex sample survey. *National Medical Care Utilization and Expenditure Survey.* Series A, Methodological Report No. 4. DHHS Pub. No. 88–20004. Public Health Service. Washington. U.S. Government Printing Office, Mar. 1988.

Library of Congress Cataloging-in-Publication Data

Statistical methodologies for analyzing a complex sample survey.

(National medical care utilization and expenditure survey. Series A, Methodological report ; no. 4) (DHHS publication ; no. 88–20004)

Written by James M. Lepkowski and others.

Bibliography: p.

Supt. of Docs. no.: HE22.26/4:7

1. National Medical Care Utilization and Expenditure Survey (U.S.) 2. Health surveys—Statistical methods. I. Lepkowski, James M. II. National Center for Health Statistics (U.S.) III. Series: National medical care utilization and expenditure survey (Series) Series A, Methodological report ; no. 4. IV. Series: DHHS publication ; no. 88–20004. [DNLM: 1. Expenditures, Health—United States—statistics. 2. Health Services—utilization—United States—statistics. 3. Sampling Studies. W 84.3 S797]

RA410.7.S724 1988 362.1'0973'021 86–600346

ISBN 0–8406–0355–X

Contents

Executive Summary . . . 1
Overview of the Report . . . 2
Survey Design . . . 3
- Survey Background . . . 3
- Sample Design . . . 4
- Data Collection . . . 6
- Survey Nonresponse . . . 7
- Weighting . . . 8
- Attrition and Item Imputation . . . 9
- Summary . . . 10

Public Use Data Files . . . 13
- Overview and Data Management Considerations . . . 13
- Data Modifications . . . 14
- Newborns . . . 14
- Hospital Stay Charges . . . 14
- Health Care Coverage . . . 15
- Bed-Disability Days, Restricted-Activity Days, and Hospital Nights . . . 15
- Hospital Admission or Discharge Dates . . . 15
- Poverty Status . . . 15
- Deliveries . . . 16
- Duplicate Hospital Stay Records . . . 16

Effects of Imputation on Survey Estimates . . . 17
- Empirical Findings . . . 17
- Strategies for Imputed Data . . . 20

Estimation Methods . . . 22
- Notion of an Average Population . . . 22
- Estimation Procedures . . . 23
- Variance Estimates and Confidence Intervals . . . 25
- Generalized Variance Formulas . . . 26

Analytic Strategies . . . 34
- Survey Analysis Strategies . . . 34
- Multiple Regression Methods for Continuous Dependent Variables . . . 35
- ANOVA Methods for Continuous and Categorical Dependent Variables . . . 36

Analysis Methods for Limited Dependent Variables . . . 37
- Logistic Regression Methods . . . 37
- Cumulative Logit Analysis Methodology . . . 40

References . . . 51
List of Detailed Tables . . . 53

List of Text Tables

A. Response rates of 16,902 eligible Round 1 sample persons during Rounds 2–5: National Medical Care Utilization and Expenditure Survey, 1980 . . . 7

B. Percent of data imputed for selected survey items: National Medical Care Utilization and Expenditure Survey, 1980 . . . 7

C. Characteristics of basic person sampling weights: National Medical Care Utilization and Expenditure Survey, 1980 . . . 9

D. Effect of person-year adjustment on counts and sampling weights for selected population groups: National Medical Care Utilization and Expenditure Survey, 1980 . . . 23

E. Coefficients for standard error formula for estimated aggregates or totals: National Medical Care Utilization and Expenditure Survey, 1980 . . . 29

F. Results from regression of predicted standard errors on observed standard errors for mean charge per person under 3 different conditions: National Medical Care Utilization and Expenditure Survey, 1980 . . . 30

G. Values of *roh* for standard error formula for estimated proportions: National Medical Care Utilization and Expenditure Survey, 1980 . . . 31

List of Figures

1. Overview of design of National Medical Care Utilization and Expenditure Survey, 1980 . . . 11

2. Mean charge for hospital outpatient department visits, by income group: National Medical Care Utilization and Expenditure Survey, 1980 . . . 19

3. Mean charge for hospital outpatient department visits using real data only, by imputation class: National Medical Care Utilization and Expenditure Survey, 1980 . . . 19

4. Dynamic population for 12 time period panel survey . . . 22

5. Cumulative frequency distribution of physician visit charges for adults 18–64 years of age by health care coverage: United States, 1980 . . . 41

6. Cumulative frequency distribution of physician visit charges for adults 18–64 years of age, with selected threshold values: United States, 1980 . . . 45

7. Lower triangle of the complex sample design covariance matrix V_p (x 10^{-4}) for the private and public health care coverage groups . . . 46

8. Lower triangle of the simple random sample covariance matrix (x 10^{-4}) for the private and public health care coverage groups . . . 46

9. Observed cumulative logits for adults 18–64 years of age, by health care coverage: United States, 1980 . . . 47

Symbols

- - -	Data not available
. . .	Category not applicable
-	Quantity zero
0.0	Quantity more than zero but less than 0.05

Statistical Methodologies for Analyzing a Complex Sample Survey

By James M. Lepkowski and J. Richard Landis, University of Michigan, P. Ellen Parsons, National Center for Health Statistics (formerly of the University of Michigan), and Sharon A. Stehouwer, The Upjohn Co., Kalamazoo, Michigan

Executive Summary

The purpose of this report is to provide researchers using the public use data tapes from the National Medical Care Utilization and Expenditure Survey (NMCUES) with guidelines for and illustrations of use of the data on those tapes. The effect of the design and execution of the survey on subsequent processing of, estimation from, and analysis of the data is especially addressed.

In order to provide potential users with an appreciation of the relationships between the design and execution of the survey and the subsequent analysis, the report begins with a historical description of how the data were obtained. Data management considerations are then reviewed, and a number of errors in the public use data tapes and corrections to those errors made by analysts at the University of Michigan are described in some detail. Because a number of important data elements received imputations to compensate for item missing data, illustrations are given of the impact that imputation can have on estimation and analysis. One illustration in particular indicates that imputed values can seriously attenuate the strength of relationships observed among nonmissing data.

Estimators of means, proportions, and totals, together with their corresponding standard errors, account for the numerous complex NMCUES design features and are presented here. The report concludes with a discussion of analytic methods suitable for investigating relationships among data items in NMCUES, particularly those appropriate for analyzing measures with limiting values, such as expenditures.

NOTE: The authors are grateful for the support received during all stages of the preparation of this document, both from colleagues at the University of Michigan and from the staff of the National Center for Health Statistics. At the University of Michigan, Kenneth Guire contributed greatly to the initial analyses of the NMCUES data, providing extensive data management and data processing support. High quality secretarial support in the preparation of this document was provided by Patrice Somerville. At the Institute for Social Research, University of Michigan, Nan Collier developed software for calculating sampling errors, and Judy Connor performed many of the analyses for generating sampling errors for national estimates.

Continual support was received from the National Center for Health Statistics. The project officer, Mary Grace Kovar, was instrumental in providing focus to the project and critical review of the report. The authors are indebted to Robert J. Casady, formerly Chief of the Statistical Methods Staff and now at the Bureau of Labor Statistics, for writing the major section in which the NMCUES survey design and estimation methodology are described. When potential errors in the data were identified during analyses, Robert Wright and Michele Chyba quickly solved the problems. Editors in the Publications Branch provided valuable assistance during all stages of the report, especially during the preparation of detailed tables.

Overview of the Report

In large-scale sample surveys, including the National Medical Care Utilization and Expenditure Survey (NMCUES), techniques such as stratified multistage sample selection, poststratified and nonresponse-adjusted estimation procedures, and longitudinal or panel data collection methods are often used. Because of these complexities in sample design, the direct application of standard statistical analysis methods to these data may yield results that are misleading. Considerable file management activity, such as the merging of data from two or more separate data files, is necessary before even the most basic types of analyses can be completed. Specialized estimation procedures are also required.

Many statistical methods have been modified so that sample design features can be incorporated into the analysis, and data processing methods and software are available for manipulation of multiple file data sets. Unfortunately, the methods for processing and analyzing such data have not been made widely available. These methods are necessary for people utilizing standard statistical software systems who want to make use of public use data tapes for large-scale surveys.

The methods and findings presented in this publication were developed in the process of producing a series of reports based on analyses of the NMCUES data (Berki et al., 1985; Parsons et al., 1986; Harlan et al., 1986; Harlan et al., to be published; Murt et al., 1986). The purpose of this report is to provide a set of analyses of NMCUES data, illustrating methods for processing and analyzing a complex sample survey. This publication is intended as a guide for users of NMCUES data who have acquired the public use data tapes and are now about to begin analyzing the data.

Experiences managing the NMCUES data files are presented together with reviews of issues, such as the impact of imputation on estimation, estimation procedures for longitudinal data, sampling error estimation, and analysis methods for sample survey data. The effect of NMCUES design features on estimates and analytic findings is emphasized throughout. Some analyses are conducted without considering the sample design, including only weights from the sample design, and including both weights and appropriate estimation procedures that account for the sample design to illustrate the effect of the design and weights on analytic findings.

The design of NMCUES from sample selection through the final preparation of sampling weights and imputation for missing data items is described in detail. These discussions are followed by data processing considerations for managing the NMCUES data files, including a discussion of modifications to the data files to correct problems that may affect some analyses. The effect of imputation for missing data on estimation is explored for several analyses, including subgroup means and regression relationships. Estimation procedures, including sampling error estimation procedures, appropriate to the NMCUES design are described. A detailed set of analyses using multiple logistic regression and cumulative logit techniques is illustrated using the NMCUES data.

Survey Design

The National Medical Care Utilization and Expenditure Survey was designed to collect data about the U.S. civilian noninstitutionalized population in 1980. Information was collected from a national probability sample of households, as well as samples of cases drawn from State Medicaid files. The questionnaire items concerned health, access to and use of medical services, charges and sources of payment for those services, and health care coverage. Cosponsored by the National Center for Health Statistics and the Health Care Financing Administration, NMCUES data collection was conducted by the Research Triangle Institute, Research Triangle Park, North Carolina, and its subcontractors, the National Opinion Research Center, Chicago, Illinois, and SysteMetrics, Inc., Santa Barbara, California.

NMCUES is a complex study designed to meet a variety of national and State policy information needs. The 1980 survey consisted of three integrated components—a national household survey, a State Medicaid household survey, and an administrative records survey. The national household component consisted of approximately 6,000 cooperating households that were interviewed four or five times during 1980 and 1981. The State Medicaid component consisted of a sample of approximately 4,000 households. Each household had one or more Medicaid-enrolled cases selected from the Medicaid eligibility files in California, Michigan, New York, and Texas. The State Medicaid survey households were interviewed at the same time as the national component households, and the same methods and questionnaires were used in both components. In the administrative records survey, Medicare and Medicaid eligibility and claims data were collected for persons in the national and State Medicaid household surveys who were enrolled in Medicare or Medicaid.

The complexity of NMCUES requires that an analyst examining available findings or seeking to investigate policy and other issues using NMCUES data be familiar with a range of design features. In particular, the data user must be able to select analytic methods that account for the survey design and level of inference.

The purpose of this report is to review the NMCUES design and a variety of methodological approaches to data analysis. Only one of the three NMCUES components, the national household survey, is considered. In this section, the overall survey design is addressed and several aspects of the design are considered, including survey background, sampling and data collection methods, survey nonresponse, and adjustments made to NMCUES data to compensate for nonresponse and other problems. This section of the report draws heavily from a paper by Casady (1983) presented to the 19th National Meeting of the Public Health Conference on Records and Statistics.

Survey Background

NMCUES can be considered one in a series of surveys concerned with health, health care, and expenditures for health care. The series began with a national survey of illness and medical care utilization and expenditures conducted by the Committee on the Costs of Medical Care during 1928–31 (Falk, Klem, and Sinai, 1933). It also includes the National Health Survey (Perrot, Tibbets, and Britton, 1939); studies conducted in 1953 and 1958 by the Health Information Foundation and the National Opinion Research Center (Anderson and Feldman, 1956; Anderson, Colette, and Feldman, 1963); and studies conducted by the Center for Health Administration Studies in 1963 and 1970 (Andersen and Anderson, 1967; Andersen, Lion, and Anderson, 1976). NMCUES is most closely related to two national surveys sponsored or cosponsored by the National Center for Health Statistics, the continuing National Health Interview Survey (NHIS) and the 1977 National Medical Care Expenditure Survey (NMCES).

NHIS is a multipurpose health survey that has been conducted continuously since 1957. Its primary purpose is to collect information on illness, disability, and the use of medical care. Although some information on medical charges and insurance payments has been collected in NHIS, the cross-sectional nature of the survey is not designed to provide data on annual charges and payments. A panel design with several interviews during the year was recognized as providing the potential for collecting more accurate and complete information on expenditures than could be obtained from a one-time interview with a yearlong recall period.

NMCES was a panel survey in which sample households were interviewed six times over an 18-month period in 1977 and 1978. NMCES was designed specifically to provide comprehensive data on how health services were used and paid for in the United States in 1977.

The NMCUES national household survey is similar

to NMCES in survey design, and it is similar to both NHIS and NMCES in the wording of questions in areas common to the three surveys. All three surveys provide information on illness and disability, but NMCES and NMCUES provide some information not available from NHIS, such as annual use of medical care, costs, sources of payment, and health insurance coverage. The similarities between NMCES and NMCUES, conducted 3 years apart, provide the opportunity for analysis of change during the 3 years between the surveys.

Sample Design

General plan—The sample design is a concatenation of two independently selected national samples, one provided by the Research Triangle Institute (RTI) and the other by the National Opinion Research Corporation (NORC). The sample designs used by RTI and NORC are quite similar with respect to principal design features; in both, extensive stratification and multistage area probability sample designs are used. Each can be characterized in terms of four stages of sample selection, with stratification of primary and secondary sampling units. The principal differences between the two designs are the type of stratification variables and the specific definitions of sampling units at each stage.

Primary sampling units (PSU's)—A PSU for a typical national household survey using area probability sampling methods in the United States is often defined as a county, a group of contiguous counties, or parts of counties. Both the RTI and NORC sample designs used similar types of PSU's, and both were based on counts from the 1970 Census of Population. RTI defined a PSU in terms of counties, groups of contiguous counties, or parts of counties with a minimum 1970 population of 20,000. A total of 1,686 nonoverlapping RTI PSU's cover the entire land area of the 50 States and Washington, D.C. For the NORC sample, a PSU consisted of a standard metropolitan statistical area (SMSA), part of an SMSA, a county, part of a county, or an independent city. NORC PSU's also covered the entire land area of the 50 States and Washington, D.C. Grouping of counties into a single PSU occurred when individual counties had a 1970 population of less than 10,000.

Stratification of PSU's—In both sample designs, PSU's were grouped into strata designed to have members relatively alike within strata and relatively unlike among strata. In the RTI design, the strata were explicitly created by placing entire PSU's into one and only one stratum. In the NORC design, a zoned selection procedure required an ordered list from which a systematic selection imposed an implicit type of stratification.

In the RTI sample design, the PSU's were classified as one of two types. The 16 largest SMSA's were designated as self-representing PSU's, and the remaining 1,670 PSU's in the primary sampling frame were designated as non-self-representing PSU's. The RTI self-representing PSU's, derived from the 16 largest SMSA's, had sufficient 1970 population size to be treated as 16 separate strata from which at least one subsequent secondary selection would be made with certainty. Of the 1,670 remaining non-self-representing RTI PSU's, a total of 1,659 were grouped into 42 strata, each of which had approximately the same population in 1970, about 3⅓ million. One additional stratum of 11 PSU's in Alaska and Hawaii, with a 1970 population size of about 1 million, was added to the RTI strata.

In the NORC sample, also, the PSU's were classified into two groups according to metropolitan status: SMSA and not SMSA. Within these two strata, PSU's were ordered by placing units with similar characteristics next to or near one another on the list. The ordered list was then partitioned into zones with a 1970 census population size of 1 million persons for the purposes of a zoned, or systematic, selection of units across zones. Zone boundaries could occur within a PSU, providing the opportunity for a single PSU to be selected more than once in the primary stage of selection.

First-stage selection of PSU's—The RTI primary-stage sample for NMCUES consisted of 59 PSU's: 16 self-representing PSU's and 43 non-self-representing PSU's, which were obtained by selecting one PSU from each of the 43 non-self-representing strata. Within non-self-representing strata, PSU's were selected with probability proportional to 1970 population size.

In the NORC primary-stage selection, a systematic selection procedure was used in which a single PSU was selected within each zone with a probability proportional to its 1970 population. Using this procedure, a PSU could be selected more than one time. For instance, an SMSA PSU with a population of 3 million would be selected at least twice and possibly as many as four times. The full NORC general-purpose sample contained 204 primary sample selections, which were systematically allocated to four subsamples of 51 each. A set of 76 primary sample selections was made for NMCUES by randomly selecting two complete subsamples of 51. One subsample was included in its entirety, and 25 of the primary selections in the other subsample were selected systematically for inclusion in NMCUES.

Second-stage units, stratification, and selection—For both the RTI and NORC sample designs, the primary selections were divided into nonoverlapping area units that covered the entire PSU. These secondary sampling units (SSU's) consisted of one or more enumeration districts defined by the 1970 census, block groups, or a combination of those units.

As in the first stage of selection, the RTI sample design grouped the SSU's into explicit strata of approximately equal size in each of the 59 NMCUES PSU's. Within each PSU, the SSU's were ordered and then partitioned to form secondary strata of approximately equal size. Two secondary strata were formed in the non-self-representing PSU drawn from Alaska and Hawaii, and four secondary strata were formed in each of the remaining 42 non-self-representing PSU's. Thus, the non-self-representing PSU's were partitioned into a total of 170 secondary strata. In a similar manner, the 16 self-representing PSU's were partitioned into 144 secondary strata. One SSU was selected from each of

the 144 secondary strata covering the self-representing PSU's, and two SSU's were selected from each of the remaining secondary strata. All second-stage sampling was with replacement and with probability proportional to SSU total noninstitutionalized population in 1970. The total number of sample SSU's was 2 × 170 + 144 = 484.

In the NORC sample design, the SSU's were ordered geographically to impose an implicit stratification through systematic selection from the ordered list. The cumulative number of households in the second-stage frame for each PSU was divided into 18 zones of equal width. An NORC SSU had the opportunity to be selected more than once, as was the case in the primary stage of selection. In addition, if a PSU had been selected more than once in the first stage, the second-stage selection process was repeated as many times as there were first-stage selections. A total of 405 SSU's were chosen in the NORC second-stage sample: Five SSU's were selected from each of the 51 primary selections in the subsample that was included in its entirety, and six SSU's were selected from each of the 25 primary selections in the group for which one-half of the primary selections were included.

Third-stage selection—In the third stage of selection in both designs, the selected SSU's were divided into nonoverlapping geographic areas for additional subsampling. For the RTI design, each SSU was divided into smaller areas, and one area within the SSU was selected with probability proportional to the total number of housing units in 1970. Then one or more nonoverlapping areas, called segments, were formed in the selected area. Each segment contained at least 60 housing units (HU's). One segment was selected from each SSU with probability proportional to the segment HU count. In response to the sponsoring agencies' request that the expected household sample size be reduced, a systematic sample of one-sixth of the segments was deleted from the sample. Thus, the total third-stage sample of 484 segments (one from each of the 484 SSU's) was reduced to 404 segments.

In the NORC sample, geographic areas were not created before a set of segments with a minimum number of housing units was defined. Instead, the selected SSU's were subdivided into area segments with a minimum size of 100 housing units. One segment was then selected with probability proportional to the estimated number of housing units.

Selection of housing units—NMCUES interviews were conducted at a sample of housing units and a sample of group quarters, hereafter jointly referred to as a sample of dwelling units. In both the RTI and NORC sample designs, once the segments were selected, all of the dwelling units within the segment (including group quarters) were listed. A systematic sample of dwelling units was selected from the listed dwelling units. The procedures used to determine the sampling rate for segments guaranteed that all dwelling units in the United States had an equal probability of selection.

Each selected dwelling unit was visited by an interviewer from the respective survey organization to determine whether any eligible sample persons resided there. All of the selected dwelling units with eligible persons were included in the sample. A control card was generated for each selected dwelling unit, and all household members (eligible and ineligible) were listed on it.

Target population—The collection of persons whose usual residence is a sample housing unit is typically defined as a household. In the case of NMCUES, the longitudinal survey design required that the usual definitions of household and sample person be modified to account for the unique dynamic nature of the population about which inferences were to be made. The concepts of key person and reporting unit, paralleling those of sample person and household in one-time cross-sectional survey design, were developed for NMCUES data collection and analysis purposes.

A *key person* was defined as a person whose usual residence at the time of the first interview was in a sample dwelling unit or a person who, although not a usual resident at the first interview, could be linked uniquely to a sample household. All key persons became part of the NMCUES national sample. Key persons included a number of persons who were not usual residents of sample households at the time of the first interview, and data concerning them were collected for the full 12 months of 1980 or for the portion of time that they were part of the U.S. civilian noninstitutionalized population. Unmarried students 17–22 years of age who lived away from home were considered to be usual residents of their parent or guardian's household. Hence, they were included in the sample as key persons when their parent or guardian's household was included in the sample. Persons who died or were institutionalized between January 1 and the date of first interview were included in the sample if they were related to persons living in the sampled households and were living in the household before their death. In addition, children born to key persons during 1980 were considered key persons, and data were collected for them from the time of birth. Relatives from outside the original population (i.e., institutionalized, in the Armed Forces, or outside the United States between January 1 and the first interview) who moved in with key persons after the first interview were also considered key persons. Data concerning them were collected from the time they joined the key person.

Relatives who moved in with key persons after the first interview but were part of the civilian noninstitutionalized population on January 1, 1980, were classified as *nonkey persons*. Data were collected for nonkey persons for the time that they lived with a key person. Because they had a chance of selection in the initial sample, their data are not used for general analysis of persons. However, data for nonkey persons can be used in an analysis of families because they contributed to the family's utilization of and charges for health care during the time that they were part of the family. Family analysis is not part of this investigation, though, and will not be discussed further.

Persons included in the sample were grouped into *reporting units* for data collection purposes. Reporting units were defined as all persons related to each other

by blood, marriage, adoption, or foster care status and living in the same dwelling unit. The combined NMCUES sample consisted of 7,244 reporting units, of which 6,600 agreed to participate in the survey. In total, complete data were obtained on 17,123 key persons.

Data Collection

The first step in the data collection process, enumeration of dwelling unit residents, has already been described. Once information about reporting unit members was recorded on the control card, it served as the primary source of information for following key persons during the course of data collection. The process of enumerating household members and verifying the status of each key and nonkey member was repeated each time a household was visited for interviewing.

The next step in data collection was administration of the household interview. In each round of data collection, a core questionnaire was administered to obtain information on illness, use of health care services, and health care expenditures since the previous interview and on health care coverage at the time of interview. During the first, third, and fifth rounds of data collection, a supplemental interview was also administered to collect data on topics that were expected to change minimally during the year or needed to be collected only once, such as employment status, 1980 income, and functional limitations. At the end of each interview except the first, a summary of health care and health care expenditures reported during previous interviews was reviewed. The computer-generated summary was mailed to the reporting units before the interview, and the interviewer carried a copy to the interview. The summary provided a means to verify previously reported events and expenditures and to update incomplete information.

Households were interviewed four or five times during 1980 and 1981 at approximately 3-month intervals. All households were interviewed in person in the first (February–April 1980), second (May–July 1980), and fifth (January–March 1981) rounds of data collection. In the third round of data collection (August–October 1980), households were interviewed by telephone whenever possible (83 percent of interviews in this round). Only about two-thirds of the households were interviewed in the fourth round (November–December 1980) because data collection for the fifth round began in January 1981, resulting in time constraints. Fourth round data collection was also conducted by telephone whenever possible (88 percent of the interviews).

Household respondents were required to be 17 years of age or over and a member of the household. Proxy respondents were used for households if all members were unable to respond because of health, language, or mental conditions.

The length of the recall period for which respondents were asked to report health care visits or expenditures varied by round. In the first round, the recall period was from January 1 up to the date of interview. With a 3-month interviewing period for this round, some respondents had to recall events during a 1-month period only, and others had to recall events over a 4-month period. The second and third rounds required recall since the last interview, a period of about 3 months for most households. Two-thirds of the households were interviewed during the fourth round. Each had a 3-month recall period for that round and less than a 2-month recall period for their round five interview. The one-third of the households not interviewed in the fourth round were interviewed at the beginning of the fifth. The average recall period for those households for their fifth round interview was approximately 3 months.

Several procedures were used to improve recall and assure high response rates. The computer-generated summary, mailed to each household prior to all interviews except the first, was designed to stimulate recall about health care events and expenditures and to update missing or incomplete information reported at an earlier interview. At the first interview, households were given a calendar and instructions to record all illnesses and health care utilization on the calendar. A pocket at the bottom of the calendar was provided for storage of receipts for review at the next interview.

A series of incentive payments was given to respondents to improve response rates and to encourage them to mail a change-of-address notification to the data collection organization if they moved between interviews. An incentive payment of $5.00 was made at the end of the first and second round interviews, and an additional incentive payment of $10.00 was made at the end of the fifth round interview. The respondent was also asked to sign an agreement to provide accurate information at each interview and to maintain the calendar.

The panel design for data collection, with approximately 13 weeks between interviews with each person, required a large data processing system in order to produce the documents for each round of interviewing. This system processed interviews and generated assignments for the next round in an average of 6 weeks from the time of receipt of an interview from the field. Processing included data receipt procedures, premachine editing, keying interviews, updating system control files, and production of the control card and summary document for the next round of interviewing.

At the end of data collection, data in the control system developed to process interviews and generate assignments had to be converted to a form suitable for analysis. Coding of conditions, geography, and other information was performed; a variety of machine edits were completed; and the control system files were restructured into analytic files. The control system files, originally organized in a format similar to the interview format, were reorganized into analytic files based on type of information, such as medical visits, hospital stays, and conditions. Numerous recoded variables were created and added to the analytic files. Sampling weights were developed for each case, and missing items in

otherwise complete interviews were filled in through a variety of imputation procedures.

Survey Nonresponse

Despite the best efforts of a data collection organization, information cannot be collected in a household survey from some of the designated survey respondents. NMCUES was no exception to this general rule. Three types of nonresponse occurred in NMCUES: Sample households or individuals refused to participate in the survey (total nonresponse); initially participating individuals dropped out of the survey at a later round (attrition nonresponse); and data for specific items on an otherwise complete interview were not collected (item nonresponse).

Response rates for reporting units and persons were high in NMCUES. Among the 7,244 reporting units eligible at the first round, 6,600 provided interviews (91.1 percent). The 644 first-round nonresponding reporting units (8.9 percent) failed to cooperate through refusals (7.2 percent), failure to find anyone at home during the survey period (1.0 percent), and other reasons (0.7 percent).

A total of 16,902 persons were enumerated in the 6,600 responding reporting units at the first round. Response rates for these persons over the course of the next four rounds of data collection were higher than 95 percent, as shown in Table A. For example, at the second round, 0.1 percent were ineligible and only 0.4 percent were nonresponding. By the fifth round, 96.5 percent of the original first-round enumerated persons were still responding to the survey. If the average number of persons per household for the first round was the same in responding and nonresponding eligible reporting units, the combination of reporting unit and person-level response rates indicates that 87.9 percent of persons eligible at the first round responded over all five rounds of data collection: (0.911)(0.965)(100) = 87.9.

Persons classified as initially responding to the survey may still fail to provide information for some or many items in the questionnaire. One instance of nonresponse among otherwise cooperating respondents is attrition nonresponse, which fortunately was a relatively small problem in NMCUES. On the other hand, item nonresponse was a problem, particularly for health care charges, income, and other sensitive topics. The extent of missing data varied by question. Table B illustrates the extent of the item nonresponse problem for selected items in the survey. The rates in Table B represent the amount of imputation, or substitution of nonmissing responses for missing data, that was required after as many missing entries as possible were completed through careful editing and checking. Although the rates in the table are not item nonresponse rates, they correspond closely to those rates.

Demographic items tended to have the lowest item nonresponse rates, some at insignificant levels, such as for age, sex, and education. Income items had higher levels of item nonresponse. Nearly one-third of the persons required imputation for at least one component of total personal income, which is a cumulation of earned income and 11 sources of unearned income. Bed-disability days, work-loss days, and cut-down days had levels of item nonresponse intermediate to the levels for demographic and income items.

Table A

Response rates of 16,902 eligible Round 1 sample persons during Rounds 2–5: National Medical Care Utilization and Expenditure Survey, 1980

Round	Responding	Nonresponding	Ineligible
		Percent	
2	99.5	0.4	0.1
3	97.9	1.5	0.6
4	97.1	2.0	0.9
5	96.5	2.3	1.2

Table B

Percent of data imputed for selected survey items: National Medical Care Utilization and Expenditure Survey, 1980

Description	Percent imputed
Person file (n = 17,123)	
Age	0.1
Race	[1]20.0
Sex	0.1
Highest grade attended	0.1
Perceived health status	0.8
Functional limitation score	3.2
Number of bed-disability days	7.9
Number of work-loss days	8.9
Number of cut-down days	8.2
Wages, salary, business income	9.7
Pension income	3.5
Interest income	21.6
Total personal income	[2]30.4
Medical visit file (n = 86,594)	
Total charge	25.9
First source of payment	1.8
First source of payment amount	11.6
Hospital stay file (n = 2,946)	
Nights hospitalized	3.1
Total charge	36.3
First source of payment	2.2
First source of payment amount	17.6
Prescribed medicines and other medical expenses file (n = 58,544)	
Total charge	19.4
First source of payment	2.8
First source of payment amount	10.0

[1]Race for children under 17 years of age imputed from race of head of reporting unit.
[2]Cumulative across 12 types of income.

The highest levels of item nonresponse occurred for the important charge items on the various visit, hospital stay, and medical expenses files. Total charges for medical visits, for hospital stays, and for prescribed medicines and other medical expenses were missing for 25.9, 36.3, and 19.4 percent of the events, respectively. The item nonresponse rates for the source of payment were small, but the nonresponse rates for the amount paid by the first source of payment were generally high. Nonresponse for nights hospitalized, located on the hospital stay file, was similar to nonresponse for the first source of payment.

Even though reporting-unit and person-level response rates are high, survey-based estimates of means and proportions using data from respondents alone may be biased if nonrespondents tend to have different health care experiences than respondents have or if a substantial response rate differential exists across subgroups of the target population. Furthermore, annual totals will tend to be underestimated unless allowance is made for the loss of data because of nonresponse. Similarly, data missing because of attrition or item nonresponse can introduce bias into survey estimates. When as many as one-third of the hospital stays are without charge information, total expenditures for hospital care or for all medical care will be severely underestimated.

Two methods commonly used to compensate for survey nonresponse are weighting procedures and imputation. Weighting procedures compensate for missing data by increasing the relative contribution of responding persons to survey estimates through the application of weights. Weights are also used to compensate for unequal probabilities of selection of sample units and to make other adjustments to survey estimates. Imputation is a process of replacing missing information for an item for one individual with data from the record of another individual who provided a response for that item. Imputation may also be made through a logical or a statistical relationship among nonmissing items within an individual's data. For NMCUES, weighting procedures were used to adjust estimates to account for reporting unit and person-level nonresponse. Imputation was used to compensate for attrition and item nonresponse. In the next sections, the methods used to develop sampling weights (including adjustments for total nonresponse) and imputation procedures used to complete attrition and item nonresponses are described.

Weighting

For the analysis of NMCUES data, sample weights are required to compensate for unequal probabilities of selection, to adjust for the potentially biasing effects of failure to obtain data from some persons or reporting units (nonresponse), and to adjust for failure to cover some portions of the population not included in the sampling frame (undercoverage). The NMCUES weighting procedure is composed of three steps: Development of base sample design weights for each reporting unit, adjustment for nonresponse and undercoverage at the level of the reporting unit, and adjustment for person-level nonresponse and undercoverage. A further adjustment was made for the number of days a person was an eligible member of the U.S. civilian noninstitutionalized population, but this adjustment affects only certain types of estimates from NMCUES and is discussed in a subsequent section, Analytic Strategies.

Basic sample design weights—Development of weights reflecting the sample design of NMCUES was the first step in the development of weights for each person in the survey. The basic sample design weight for a dwelling unit is the product of four components, which correspond to the four stages of sample selection. Each of the four components is the inverse of the probability of selection at that stage, when sampling was without replacement, or the inverse of the expected number of selections, when sampling was with replacement and multiple selections of the sample unit were possible.

As previously discussed, the NMCUES sample is comprised of two independently selected samples. Each sample, together with its basic sample design weights, yields independent unbiased estimates of population parameters. Because the two NMCUES samples were of approximately equal size, a simple average of the two independent estimators was used for the combined sample estimator. This procedure is equivalent to computing an adjusted basic sample design weight by dividing each basic sample design weight by 2. In the subsequent discussion, only the combined sample design weights are considered.

Total nonresponse and undercoverage adjustment—A weight adjustment factor was computed at the reporting unit level to compensate for nonresponse and undercoverage at this level. Every reporting unit within a dwelling unit is included in the sample, so the adjusted basic sample design weight assigned to a reporting unit is simply the adjusted basic sample design weight for the dwelling unit in which the reporting unit is located. A reporting unit was classified as responding if the reporting unit initially agreed to participate in NMCUES; otherwise, it was classified as nonresponding.

Initially 96 reporting unit weight-adjustment cells were formed by cross-classifying the race of reporting unit head (two levels), type of reporting unit head (three levels), age of reporting unit head (four levels), and size of reporting unit (four levels). These cells were then collapsed to 63 cells so that each cell contained at least 20 responding reporting units. Within each cell an adjustment factor was computed so that the sum of adjusted basic sample design weights would equal the March 1980 Current Population Survey estimate for the same population. Each reporting unit weight was adjusted for nonresponse and undercoverage by computing the product of the adjusted basic sample design weight and the nonresponse and undercoverage adjustment factor for the cell containing the reporting unit.

As a final step in the reporting unit weighting procedure, the distribution of the adjusted weights was examined to determine whether the range of weights was large. Several weights were identified as somewhat larger than the next larger weights in the sample. It was thought that the added contribution to variance from these large weights was greater than their potential for bias reduction in estimation. Therefore, a trimming procedure was applied in which the largest weights were changed to the values of the next largest weights in the data. The weights of all the other observations were then adjusted to restore the same sum of weights obtained prior to the trimming.

Poststratification adjustment—Once the reporting unit weights adjusted for nonresponse and undercoverage were computed, an adjusted weight was computed at the person level. Because each person within a reporting unit is included in the sample, the nonresponse-adjusted and undercoverage-adjusted weight for a sample person is the nonresponse-adjusted and undercoverage-adjusted weight for the reporting unit in which the person resides. Each person was classified as responding or nonresponding, as discussed subsequently in the section on attrition imputation.

Sixty poststrata were formed by cross-classifying age (15 levels), race (two levels), and sex (two levels). One poststratum (black males 75 years and over) had fewer than 20 respondents, so it was combined with an adjacent poststratum (black males 65–74 years), resulting in 59 poststrata.

Estimates based on population projections from the 1980 census were obtained from the Bureau of the Census. These estimates of the U.S. civilian noninstitutionalized population poststratified by age, race, and sex were for February 1, May 1, August 1, and November 1, 1980. The mean of these midquarter population estimates for each poststratum was computed and used as the 1980 average target population for calculating the poststrata adjustment factors. Table 1 presents the average target population in 60 sex by race by age subgroups, from which 59 poststrata were formed.

Survey-based estimates of the average poststratified population were developed using the weights adjusted for nonresponse and undercoverage. First, a survey-based estimate of the target population of each poststratum for each quarter was computed by summing the adjusted weights for respondents eligible for the survey on the midquarter date. Then the survey-based estimate of the 1980 average population was computed as the mean of the four midquarter estimates. Finally, the poststratification adjustment factor was computed in each poststratum as the ratio of the 1980 average target population (obtained from the Bureau of the Census data) to the NMCUES 1980 average population. The poststratified weight for each respondent was then computed as the product of the nonresponse-adjusted and undercoverage-adjusted weight and the poststratification adjustment factor for the poststrata containing the respondent.

Table C illustrates the nature of the final basic person weights that were produced by the weighting method outlined here. For the 17,123 key persons in the public use data files, the mean weight was 13,220. This means that, on average, each person record in the public use data file represents 13,220 persons in the 1980 U.S. civilian noninstitutionalized population. The distribution of the weights is likely to lead to slightly greater variance of estimates compared with a sample of the same size but without the need for weights. The ratio of the largest weight to the smallest is 8.3. Following the simple random sampling model of Kish (1965, section 11.7), the variance of an estimated mean will be increased by less than 5 percent as a result of the distribution of these weights.

Table C

Characteristics of basic person sampling weights: National Medical Care Utilization and Expenditure Survey, 1980

Item	Value
	Number
Sample size	17,123
Minimum value	5,508
Maximum value	45,689
Mean weight	13,220
	Number in thousands
Sum of weights	226,368
	Percent
Approximate increase in variance due to weighting	4.8

Attrition and Item Imputation

A sequential hot-deck imputation method was used for attrition imputation. First, each sample person with incomplete annual data (referred to as a *recipient*) was linked to a sample person with similar demographic and socioeconomic characteristics who had complete annual data (referred to as a *donor*). Second, the time periods for which the recipient had missing data were divided into two categories, imputed eligible days and imputed ineligible days. The imputed eligible days were those days for which the donor was eligible (in scope), and the imputed ineligible days were those days for which the donor was ineligible (out of scope). The donor's medical care experiences during the imputed eligible days—medical provider visits, dental visits, hospital stays, etc.—were imputed into the recipient's record for those days. Finally the results of the attrition imputation were used to make the final determination of a person's respondent status. If more than two-thirds of the person's total eligible days (both reported and imputed) were imputed eligible days, then the person was considered to be a total nonrespondent and the data for the person were removed from the data file. The

poststratification adjustment was made after these data were removed.

The methods used to impute data for missing values were diverse and tailored to the measure requiring imputation. Three types of imputation predominated: Deductive, sequential hot deck, and weighted sequential hot deck.

Deductive, or logical, imputations were used to fill in missing responses that could be determined readily from other data items that provided overlapping information. (These might even be referred to as edits.) For example, race was not recorded during the survey for children under 17 years of age. Instead, a logical imputation made during the data processing assigned the race of the head of the household (or the head's wife) to the child. Similarly, extensive editing was performed for the charge data before any imputations were made. For example, if first source of payment was available, only one source of payment was indicated; and if total charge was missing, the value of the first source of payment amount was assigned to the total charge item.

A sequential hot-deck procedure (Ford, 1983) was used primarily for small numbers of imputations for demographic items. In the sequential hot-deck procedure, the data are grouped into imputation classes and then sorted within those classes by measures that are correlated with the item for which imputations are to be made. An initial value, such as the mean of the nonmissing cases for the item within the imputation class, is assigned as a "cold-deck" value. The first record in the imputation class is then examined. If it is missing, the cold-deck value replaces the missing data code. If it is real, the real value replaces the cold-deck value and becomes a hot-deck value. Then the next record is examined. Again, if the value is missing, the hot-deck value replaces the missing data code; if it is real, the hot-deck value is replaced. The process continues sequentially through the imputation classes until all missing values have been replaced.

Finally, the weighted sequential hot deck (Cox, 1980) was used most extensively, providing imputed values for a variety of measures, many of which had substantial amounts of item missing data. This method is a modification of the sequential hot-deck method in which the sampling weights assigned to each record determine which real values are used to impute for a particular missing value. Records are classed and sorted by measures expected to be correlated with items requiring imputations. The procedure is applied to several items simultaneously to reduce the number of passes of the data that are required to complete imputations on many items. Because the selection of a record to serve as a donor for a particular missing value depends on the sampling weights, it is possible for a record to serve as a donor more than one time if it has a sampling weight that is much larger than that of other records.

Often a combination of methods was used to impute for a single item. Imputations for the charge items involved a combination of logical imputations, or edits, followed by the weighted hot-deck procedure. For example, an extensive edit was performed for medical visit total charges to eliminate as many inconsistencies between the source of payment data and total charge items as possible. Then the medical visit records were separated into three types: emergency room, hospital outpatient department, and doctor visits. Within each type, the records were classed and sorted by different variables prior to a weighted hot-deck imputation. For instance, records for doctor visits were classified by the reason for visit, the type of doctor seen, whether work was done by a physician, and the age of the individual. Within the groups formed by these classing variables, the records were sorted by type of insurance coverage and month of visit. The weighted hot-deck procedure was used with the classed and sorted data file to impute simultaneously for missing values of total charge, sources of payment, and sources of payment amounts.

Because imputations for missing items were made for a large number of the important items in NMCUES, they can be expected to influence the results of the survey in several ways. In general, the weighted hot deck is expected to preserve the means of the nonmissing observations when those means are for the total sample or classes within which imputations were made. However, means for other subgroups, particularly small subgroups, may be changed substantially by imputation. In addition, sampling variances can be substantially underestimated when imputed values are used in the estimation process. For a variable with one-quarter of its values imputed, for instance, sampling variances based on all cases will be based on one-third more values than were actually collected in the survey for the given item; that is, the variance will be too small by a factor of at least one-third. Finally, the strength of relationships between measures that received imputations can be substantially attenuated by the imputation.

A more complete discussion of these issues can be found later in this report.

Summary

Figure 1 is a summary of the steps in the NMCUES design from initial sample selection through the collection of data to the final weighting adjustments for nonresponse and undercoverage and imputations for item nonresponse. Each of these features of the survey design can have an impact on the choice of methods to use for analysis of NMCUES findings and on the interpretation of results.

For example, the sampling plan, which has stratified multistage selection, requires special variance estimation procedures that account for these design features. The data collection design over the course of a 1-year period leads to the consideration of average populations and estimators that resemble "risk rates" in epidemiology, with denominators that are average population estimates.

The imputation procedures may attenuate the strength of relationships observed from the data. The analyst (as well as the reader of reports based on NMCUES data) should be aware of the implications of the design for analysis and interpretation of findings. In the next section, the nature of the public use files available to the analyst is examined.

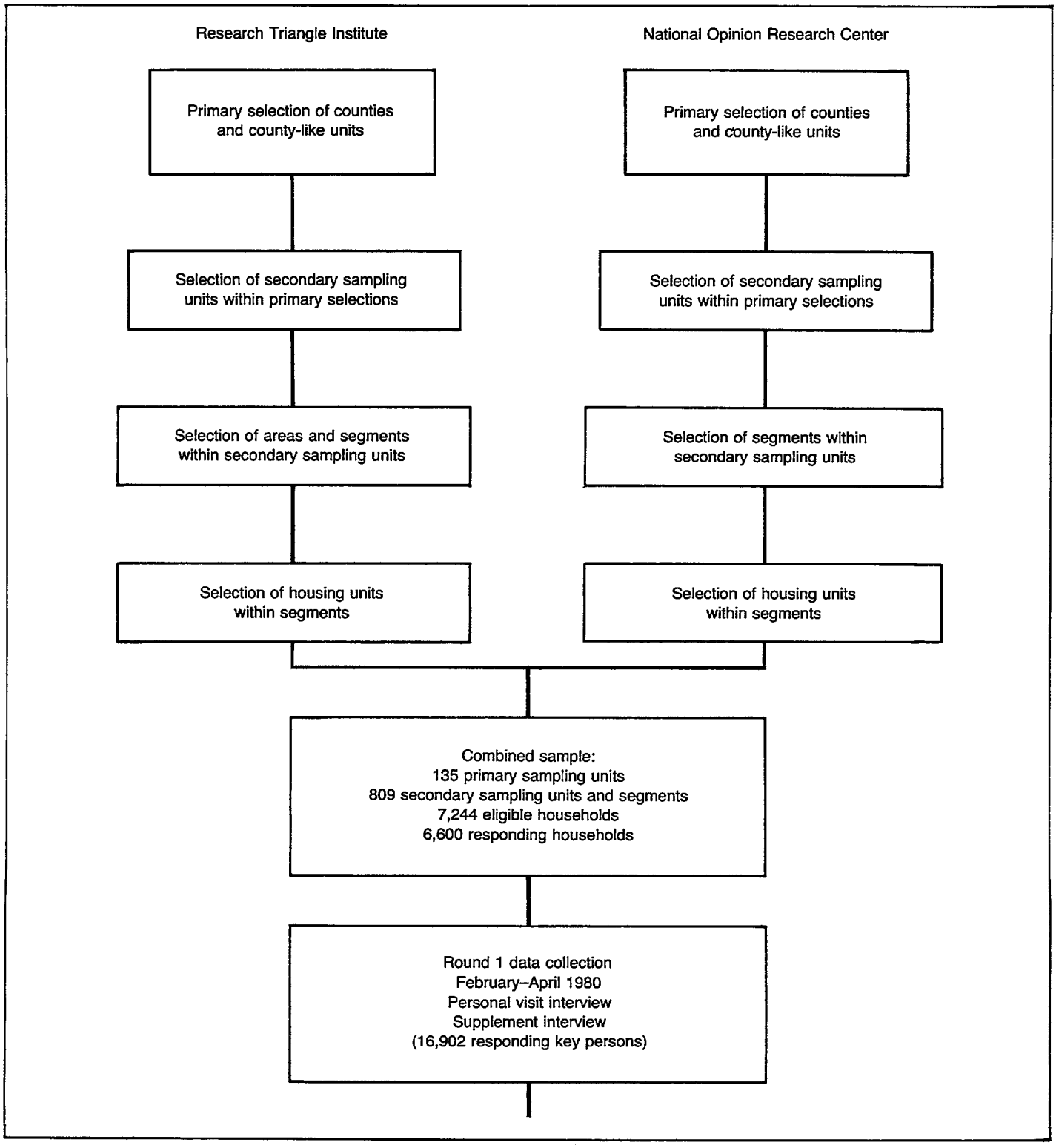


Figure 1

Overview of design of National Medical Care Utilization and Expenditure Survey: 1980

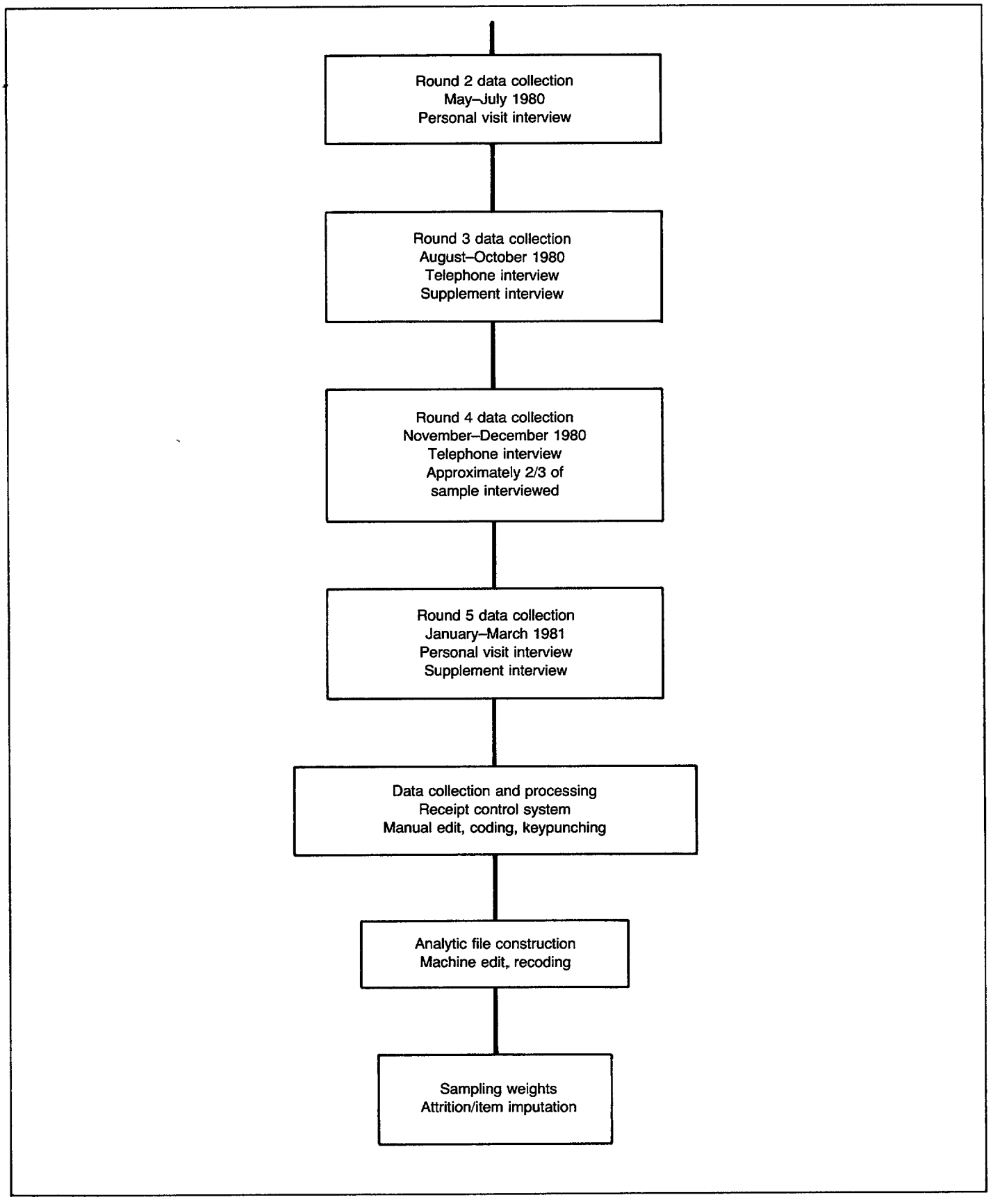


Figure 1

Overview of design of National Medical Care Utilization and Expenditure Survey: 1980—Con.

Public Use Data Files

Overview and Data Management Considerations

The NMCUES national household survey public use data files comprise six fixed-length rectangular files distributed on six magnetic tapes. The files are labeled as follows: Person, medical visit, dental visit, hospital stay, prescribed medicine and other medical expenses, and condition. Depending on the type of analysis required, the user may need to create various analytic files by combining information from these different files.

The pivotal data file is the person file, which contains a primary record of information for each of the 17,123 key individuals in NMCUES. Sampling weights for all of these individuals are recorded only in this file. All other data files contain medical event or condition records only for individuals reporting medical events or conditions; that is, the medical event files contain information, including sampling weights, only for users of the service covered in the file. Similarly, the condition file contains information only about persons reporting conditions that resulted in health care use and/or disability. Therefore, these files must be linked with the person file for any analyses on the individual level that include persons whose records are available only on the person file.

A study of different sources of payment for all health services, for example, requires manipulation of several data files. Data must first be aggregated for each person across the various medical event files, taking into account the specific source of payment for each event. The source-of-payment data can then be linked through the participant sequence number with information in the person file to produce source-of-payment estimates by various individual characteristics. This manipulation of files for source-of-payment data is necessary for three reasons. First, source-of-payment data are not available on the person file; only annual total charges for each type of service are available from the person file records. Second, data must be aggregated across different types of use for each individual. Third, the medical event files cannot be used to produce individual-level estimates for the total sample population (including nonusers of one or more types of service) independent of the person file because data for nonusers are not included on the event files.

Source-of-payment estimates for users of a specific type of service only, such as hospital care, can be produced using the relevant medical event file alone. Event-level estimates, such as charges by source of payment for hospital stays, can also be produced without linking to the person file.

Condition-specific analyses, such as those reported in Murt et al. (1986), Harlan et al. (1986), and Harlan et al. (to be published), require sophisticated manipulation of the data files because the various files are organized around different types of records and contain overlapping information.

The person file contains only summary information about conditions, medical events, and disability. A count of the number of conditions and some information about activity limitations and disabling conditions are included, but no information is given for linking conditions with specific events or disability days. Annual totals for use of and charges for the various types of health services and totals for disability days are also listed. The person file does not include source-of-payment data. However, it does include information about health care coverage and a variety of other health-related variables, such as usual source of care, as well as a full range of demographic and socioeconomic characteristics.

In the condition file, a specific condition is the unit around which all other data in the file are organized. Each reported condition is assigned a unique condition number, and the information concerning each condition is recorded on a separate record. The 51,465 records in the condition file include only conditions that resulted in disability and/or use of health services. Charges for health services are also associated with each condition, as are reasons for not seeking medical care and month and year of onset or accident.

It should be noted that respondents or informants were allowed to report more than one condition for each medical event or episode of disability days. No primary condition was indicated when multiple conditions were reported. All charges, use of service, and disability were associated with each reported condition. For example, all charges for a particular hospital stay are assigned to each condition reported for that hospital stay. Therefore, summation of use, charge, and disability data from the condition file results in a duplicated count of these measures on the person level of analysis; that is, one disability day would be counted twice for one person if two conditions were listed as causing that disability day.

In contrast, the various medical event files (hospital stay, medical visit, etc.) are organized around individual events. Each reported medical event is assigned a unique

number, and the information concerning each medical event is recorded on a separate record. In the medical event files, conditions, charges, and sources of payment, but not disability, are associated with each medical event. (Conditions are not associated with dental visits.) Individual-level analysis of medical event data requires linkage with the person file if nonusers are to be included in the analysis or if more than one type of service is being studied. There are 86,594 records in the medical visit file, 23,113 records in the dental visit file, 2,946 records in the hospital stay file, and 58,544 records in the prescribed medicines and other medical expenses file.

Imputation status is recorded on all six files for all variables included in NMCUES imputation procedures.

Analysis of condition-related data at the person level requires linkage at least between the condition and person files and may require linkage with one or more of the medical event files as well. Estimation of disability days for persons reporting neoplasms, for example, requires linkage between the condition and person files. An analysis of sources of payment for hospital and physician services for persons reporting neoplasms would require additional linkage with medical visit and hospital stay files.

In any person-level analysis of data from the condition file, the possibility of reporting multiple conditions must be taken into account, as noted earlier. Multiple counting of charges or disability days may be permissible for condition-level analysis. However, to avoid multiple counting, person-level analysis may require apportionment of charges or disability days among the various conditions reported for the same medical event or disability episode. Attribution of all charges or disability days to one of the reported conditions is another alternative. Because no primary condition is specified, assignment of all charges or disability days to the first condition listed may be the best option if the latter approach is selected.

Data Modifications

A number of data accuracy problems in the NMCUES public use files require modification before data analyses are conducted. Most of these modifications have some relationship to the hospital file; that is, changes must be made to variables in the hospital stay file or to hospital-related summary variables that appear in the person file. Other modifications involve newborn sampling weights, disability days, health care coverage, and categorical poverty status. Analyses that do not require the use of the hospital stay file and/or hospital-related variables in the person file or the other variables listed here can be performed without making these modifications.

The following problems were identified and addressed by modifying the data files.

(1) Sampling weights for 68 newborns were changed to reflect accurate survey eligibility status in accordance with instructions from the National Center for Health Statistics.

(2) Hospital charges were corrected for six respondents with extremely high values.

(3) Forty-seven respondents were reassigned to appropriate health care coverage categories based on source-of-payment data.

(4) Records for 175 persons had fewer bed-disability days than hospital nights. Records were edited to make the number of bed-disability days equal to the number of hospital nights.

(5) Coding errors were corrected for four respondents with incorrect hospital admission or discharge dates.

(6) Poverty status classification on the categorical variable was inconsistent with the continuous poverty status variable for four respondents. Categorical assignments were changed for those four individuals.

(7) A number of changes were necessary to correct information about nine respondents whose hospital records were incorrectly coded as deliveries in the hospital file.

(8) Hospital records were modified for one respondent who had duplicate records.

Newborns

Sixty-eight newborns were incorrectly considered eligible for the entire survey period. These errors were corrected by changing the eligible time-adjustment factor and the person time-adjusted weight for each of the 68 records. Table 2 presents the person identifying number (or participant sequence number) and revisions to time-adjusted weights and time-adjustment factors for these 68 newborns.

Hospital Stay Charges

Six hospital stay records with total charges of $90,000 or more contained the highest charges for individual medical events recorded in the NMCUES data files. Because of their importance to estimates of total charges and other expenditure measures, these records were carefully reviewed by University of Michigan and National Center for Health Statistics personnel. Several inconsistencies within these records suggested that the charge data were incorrect and required revision. These six records and related information in the person file were changed to conform with Medicare records maintained by the Health Care Financing Administration or with other information about each of the six hospital stays. In several cases, examination of Medicare records revealed that the error in the NMCUES hospital file resulted from a miskeyed decimal place. Table 3 presents the revised values in the hospital stay file, and Table 4 presents the revised values in the person file for these six records.

Health Care Coverage

Discrepancies between source of payment and health care coverage were noted in the course of analysis of the NMCUES data. All of the discrepancies involved Medicare coverage. Forty-seven respondents reporting Medicare as a source of payment in the medical visit, hospital stay, or prescribed medicine files were not properly coded as covered by Medicare. Health care coverage for these respondents was reclassified strictly according to source-of-payment data, except as follows.

- Respondents originally coded as covered by private insurance but not showing private insurance as a source of payment for any services were coded as having Medicare and private insurance coverage.
- For cases in which reassignment based on imputed data would conflict with reassignment based on real data, the real data were used.

The original and revised health care coverage categories for these 47 persons are presented in Table 5. Both categories are based on health care coverage definitions developed by the University of Michigan research staff and reviewed by the National Center for Health Statistics project staff. Although the University of Michigan health care coverage variables were derived from information on the public use files, these coverage categories are not directly available on the public use files. Therefore, if the University of Michigan categories are not used, other changes must be made to the health care coverage variables that are on the person file for respondents listed in Table 5. Information about those changes can be obtained directly from the Utilization and Expenditure Statistics Branch, Division of Health Interview Statistics, National Center for Health Statistics. The University of Michigan health care coverage categories are mutually exclusive and are specified as follows.

Under 65 years

Coverage all year:
  Private insurance only
  Medicaid only
  Other public program only
Mixed coverage:
  Private insurance and public program
  More than one public program
Part-year coverage
No coverage

65 years and over

Medicare coverage:
  Medicare only
  Medicare and private insurance
  Medicare and other public program
No Medicare coverage:
  Any other coverage
  No coverage

Specific codes for these categories are presented in Tables 6 and 7.

Bed-Disability Days, Restricted-Activity Days, and Hospital Nights

Examination of values for hospital nights and bed-disability days revealed that in 175 cases the value for hospital nights was greater than the value for bed-disability days (Table 8). According to interviewer instructions for the NMCUES questionnaire, hospital nights should be included in bed-disability days, except for newborns. Therefore, the value of bed-disability days was revised to equal hospital nights for these 175 cases. Restricted-activity days were also revised to reflect the added bed-disability days. This adjustment is a standard edit for this type of inconsistency in a data set. However, it does not fully compensate for errors in recording or computing bed-disability days. Bed-disability days are probably underestimated even after the revision. The edit was performed without regard to the imputation status of either bed-disability days or hospital nights.

Hospital Admission or Discharge Dates

Four cases identified in the examination of discrepancies between hospital nights and bed-disability days were discovered to have improperly coded hospital admission or discharge dates. Lengths of hospital stay were incorrectly coded for these cases. The admission or discharge dates and hospital nights were corrected for these four cases, but bed-disability days and restricted-activity days were not altered (Table 9).

Poverty Status

Cross-tabulation of poverty status classification derived from the continuous variable by poverty status categories coded on the public use files indicated that four respondents were miscoded on the categorical variable. The categorical variable for these four respondents was recoded to agree with poverty status on the continuous variable (Table 10).

Deliveries

Problems were discovered on nine records coded as deliveries in the hospital stay file. A variety of resolutions were identified as appropriate.

- Two deliveries were attributed to male respondents. Examination of individual records suggested that the sex variable was incorrectly coded in these two cases. The sex variable was therefore recoded to female.

- A third delivery attributed to a male was actually that of the respondent's spouse. In this case, the hospital record was reassigned and appropriate changes made in the person file for both persons.
- Four hospital stays for newborns were incorrectly coded as deliveries. These were recoded as newborn stays in the hospital stay file.
- A fifth newborn's hospital record was attributed to its mother. In this case, the hospital record was transferred to the newborn, and appropriate changes were made in the person file for both persons.
- One delivery was attributed to a woman 74 years of age. Following advice from the National Center for Health Statistics, the record was recoded to reflect "signs, symptoms, and ill-defined conditions" as the admitting condition.

These changes are summarized in Tables 11 and 12.

Duplicate Hospital Stay Records

Two sets of duplicate records (four records in total) in the hospital file were discovered for one respondent. The two duplicates were deleted in the hospital stay file, and necessary changes were made in the person file. Three of the four records had been imputed to another respondent as part of the attrition imputation process. No changes were made in the records for the person receiving the imputations. These modifications are presented in Table 13.

Effects of Imputation on Survey Estimates

The effects of imputation on survey analysis initially do not appear to be a concern because, when imputation occurs, the number of items and the amount of imputation for any given item are usually limited. Moreover, these effects can be misleading because it appears that the data are complete and that no special procedures are needed to compensate for missing data in the analysis, when in fact the imputed values can have substantial effects on survey estimates (Dempster and Rubin, 1983). For the analyst aware that imputation occurred for missing data for a particular item, it certainly is easier to assume that the effect of imputation on a particular analysis can be ignored safely or that the effect is completely beneficial (i.e., reduction of bias resulting from missing data). Unfortunately, for some types of items in a survey data set, the amount of item nonresponse and imputation can be substantial, producing large effects on the results of the data analysis.

In the section on survey design, the extent of missing data in NMCUES and the methods used to compensate for missing data were discussed. Imputation procedures that were used to compensate for item nonresponse were also reviewed. The purpose of this section is twofold: To examine the effects of imputed values on the analysis of NMCUES data and to review analytic strategies that might be used to handle imputed data. We begin by reviewing analyses for several utilization and expenditure measures with and without imputed values included in the computations. Based on these investigations, several strategies available for handling the imputed data in the statistical analysis are suggested. Finally, recommendations are made about suitable strategies for the routine analysis of NMCUES data with imputed values.

Empirical Findings

Extensive imputations were made for missing values for a large number of the key items in NMCUES. Imputation can be expected to influence estimates made from the survey in several ways.

Although the weighted hot deck is expected to preserve the means of nonmissing observations for the total sample or classes within which imputations were made (Cox, 1980), this will not be the case for sampling variances. Sampling variances can be underestimated substantially when imputed values are used in the estimation process (Kalton and Kasprzyk, 1982). For example, the estimated sampling variance for a variable with one-quarter of its values imputed will be based on one-third more values than were actually collected in the survey when the computations include all the data, real as well as imputed values. Thus the variance will be underestimated by a factor of at least one-third (Kalton, 1983).

Relationships observed among variables based on the real values can be altered significantly when imputed values are included in the analysis. In particular, the strength of a relationship may be attenuated by imputation. Santos (1981) demonstrates that the attenuation of correlations by imputed values can be substantial.

Table 14 presents estimated means and sampling errors for five survey measures from NMCUES that have small to modest levels of missing data (from 8 to 18 percent). Separate estimates are presented for computations using all data (both real and imputed) and using only the real data for each measure. The weighted mean and its standard error can be computed under two different assumptions. The simple random sampling (SRS) standard error of the mean is computed as though the observations were selected independently, but the sampling weights are included as part of the estimate. The weighted complex standard error accounts for the stratified multistage nature of the design, including the weights. The ratio of the complex to the SRS standard error is the square root of the design effect. (The design effect is the ratio of the variances.) This ratio is presented in the last column of Table 14. The design effect frequently is used to assess the effect of the sample design on survey estimates. It is generally found to be greater than 1 for complex sample surveys.

For each disability measure in Table 14, the means computed using all the data and using only the real data are quite similar. This similarity is expected because the weighted hot-deck imputation procedure is designed o preserve the weighted mean for overall sample estimates. However, the SRS standard errors are smaller when all the data are used. The SRS variance is inversely related to the sample size; therefore, imputed values increase the number of observations used in the variance calculation. Similarly, the complex standard errors for bed-disability and work-loss days are smaller when all the data are used. At the same time, the ratio of the

complex to SRS standard error essentially remains unchanged whether only real or all data are used. The decrease in the standard error when all data, instead of only real data, are used is proportionately the same for the SRS and the complex standard errors, except that the complex standard error for cut-down days and restricted-activity days is slightly larger when all the data are used.

One may conclude that imputation for the disability measures shown in Table 14 has little or no effect on estimated means or their standard errors. Although these results were computed for the total population, similar findings may be expected for most subgroups.

In contrast, imputation has larger effects for measures that have more extensive amounts of missing data. For example, more than one-half of the entries for total charge for a hospital outpatient department visit were missing. (See Table 15.) Of 9,529 hospital outpatient department visits (real visit records plus those generated from the attrition imputation process), 4,841 have a total charge that was imputed from one of the other hospital outpatient department visit records. Thus, 50.8 percent of the total charges were missing for this particular medical event. Despite the large amount of missing data, the weighted means computed using all the data and using only real values are similar, although the mean for all the data is slightly lower.

Table 15 also presents the means and other estimates for real data by the number of times a record was used as a donor in the weighted hot-deck imputation process. Donor records were chosen within classes in which the donor and imputed record were similar. Therefore, donors and recipients should have similar values for the measure being imputed. However, based on donor group means in Table 15, it appears that outpatient department visits with missing values received quite different mean total charges from those without missing values. On average across all donor classes, the visits with missing values received imputed values with lower mean total charges than the charges for real values.

It is also shown in Table 15 that sampling errors are changed substantially when imputed values are added to real values in the computation. The SRS standard error decreases 28.8 percent when the imputed values are added to the computation. Because the SRS variance (the square of the SRS standard error) is inversely proportional to the number of observations used in the calculation, it is expected to decrease by the proportionate reduction in sample size (50.8 percent). However, the SRS variance actually decreases by 49.3 percent, suggesting that sample size alone does not account for the magnitude of the observed reduction. This difference between expected and observed decrease in sampling variance is small and may be attributable to variability in the imputation process itself. Nevertheless, it is of interest to investigate the extent to which the amount of variability among elements may also be increasing when imputed values are added to the calculation.

An estimate of the element standard deviation, a measure of the variability among elements, was computed by multiplying the SRS standard error by the square root of the sample size. The estimated element standard deviations are presented in the last column of Table 15. The element standard deviation actually increases by 2.6 percent when imputed values are included. The expected decrease in standard error is reduced by this change in element variance. Furthermore, an examination of the element standard deviations for the various donor subgroups indicates that donors had greater variability with respect to outpatient department visits than did the total group with real values (which includes the donors).

The decrease in the SRS standard error when using all data is somewhat smaller than expected given the increased sample size, but the decrease in the complex standard error is actually larger than might be expected. The complex standard errors decreased 34.8 percent rather than 29.9 percent, the expected decrease based on sample size alone. As a result, the ratio of complex to SRS standard errors decreases somewhat when imputed values are included. This larger than expected decrease probably results from the imputation of values across strata and primary sampling units that are used to form computing units for complex variance estimation. Imputation across computing units has the effect of decreasing the variability among units and reducing the computed standard error.

Neither the complex standard error computed using only real data nor the one computed using all data is the actual standard error of the weighted mean computed using all the data. The mean computed using all data includes 4,841 values that were actually subsampled with replacement from the 4,688 real values. In addition, the imputations were made across the strata and primary sampling units used in the complex variance estimation procedure, a process which requires an assumption that the observations are selected independently between primary sampling units and strata. Hence, the complex standard error based on computations using all the data, shown in Table 15, fails to account for two sources of variability: The double sampling used to select values for imputation and the correlation between primary sampling units and strata induced by imputation.

At the same time, the complex standard error for the weighted mean computed using only the real data is an incorrect estimate of the standard error of the mean based on all the data. The actual standard error of the weighted mean computed using all the data is probably larger than that shown in Table 15. It may even be larger than the standard error computed using only the real data.

Unfortunately, none of the estimators used in Table 15 is totally appropriate for the actual standard error for the weighted mean computed using all the data. An appropriate estimator might be developed if information about the correspondence between a specific donor and recipient were available, but this information is not usually provided on public use tapes of survey data.

As an alternative, an estimation strategy was developed based on an adjustment to the sampling weights to compensate for item missing data rather than making use of the imputation process results.

The first step in the adjustment of sampling weights to compensate for item nonresponse for outpatient department visits is the creation of the original imputation classes. Sixteen classes were formed by cross-classifying age (under 17 years, 17–44 years, 45–64 years, and 65 years and older), sex, and whether or not a doctor was seen during the visit. Within each class, weights for recipients of an imputation were summed, and the total number of donations made in the class was counted. The sum of the weights for imputed records was then divided by the number of donations within each class to form an average weight that each donor contributed to an imputation within the class. In the final step in the process, the person weight for each donor was increased by the average donor weight multiplied by the number of donations in the imputation process for outpatient department visits made by that particular donor. These adjusted weights for donors have the property that they sum, within an imputation class as well as over all classes, to the total sum of weights for donors and recipients combined.

The adjusted weights permit the estimation of a mean or other statistic using only the real data as well as incorporation of an adjustment for item nonresponse directly into the estimate. Moreover, correct complex standard errors can be computed for these weighted means using the real data and the adjusted weights because imputation across strata and primary sampling units does not occur.

The estimated mean and its standard error under this adjusted weighting procedure, shown in the last row of Table 15, are similar to those obtained using all the data. However, they are greater than those obtained using only the real data with unadjusted weights and those obtained using all the data. The difference between the complex standard error for the weighted mean computed using only the real data and for the mean computed using adjusted weights is caused by the effects of increased variability of adjusted weights relative to the unadjusted weights.

As a final illustration of the effects that imputation can have on survey data analysis, the relationship between total charges for outpatient department visits and family income of the person making the visit is examined. Figure 2 presents estimated mean total charges per outpatient department visit computed using all the data and using only the real data for four family income groups. Using only the real data, the mean total charge per visit increases linearly as the family income increases. However, when all the data are used to estimate the mean total charge per visit, the mean charge does not increase as rapidly with increasing family income. This linear relationship between family income level and mean total charge per outpatient department visit in the real data has been attenuated by the imputed values. Table 16 presents the mean total charges for each of the four income groups for all data and for only the real data.

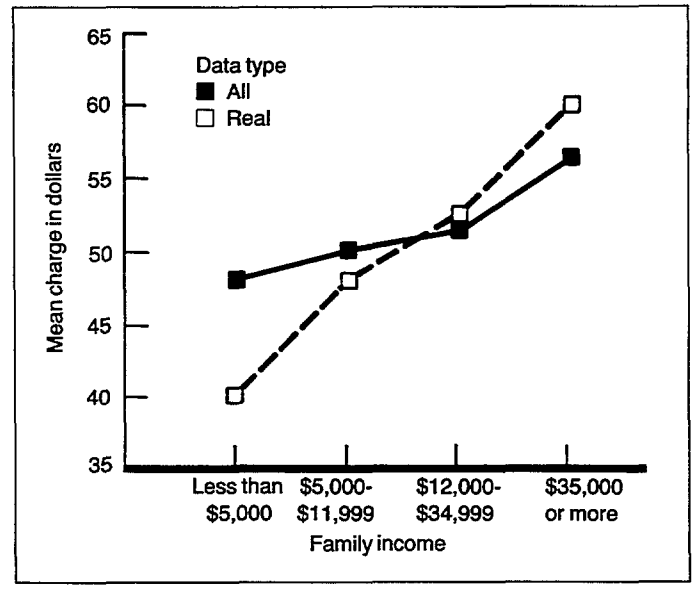


Figure 2

Mean charge for hospital outpatient department visits, by income group: National Medical Care Utilization and Expenditure Survey, 1980

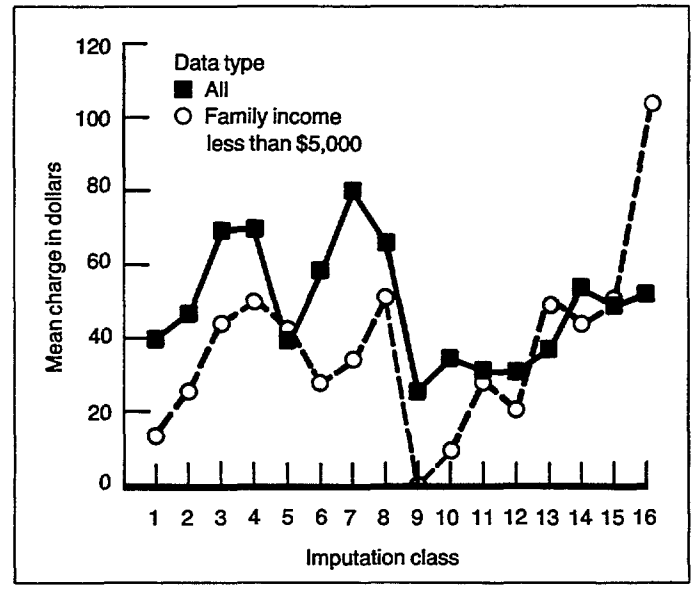


Figure 3

Mean charge for hospital outpatient department visits, using real data only, by imputation class: National Medical Care Utilization and Expenditure Survey, 1980

The reason for this attenuation can be deduced from Figure 3. Sixteen imputation classes—based on measures that did not include family income—were formed for the imputation of total charges for outpatient department visits. Figure 3 shows mean total charges for real data for the total sample and the subgroup with family income less than $5,000 in 1980.

The low-income group generally has lower mean charges than the total sample has. However, family income was not one of the variables used to form imputation classes. Therefore, low-family-income persons within an imputation class with missing outpatient department visit total charges were imputed a charge that was, on average, higher than the mean charge for low-income persons with real data. This occurred in almost every imputation class. When the real and imputed data for persons with family incomes less than $5,000 are combined, the effect of imputation is to increase the mean charge for this subgroup. Conversely, for persons with family incomes of $35,000 or more, total outpatient department visit charges for persons with real data tend to be larger than values imputed to persons with missing charges. The overall impact of the imputation process on the relationship between charges for outpatient department visits and family income is a regression toward the mean charge for real data for low- and high-income subgroups.

This attenuation occurs when the subgroups are formed from a measure that was not controlled for in the imputation process for the item receiving imputations. In addition, the relationship will not be attenuated as seriously if only real data and adjusted weights are used in the analysis, as shown in Table 16.

Strategies for Imputed Data

The results presented here demonstrate the effect that imputation can have on estimated means, estimated standard errors, and relationships among measures. The analyst of the NMCUES data must select a strategy for handling imputation in estimation. The results in this section allow comparison among four different strategies for handling imputations in the NMCUES data:

(1) Use all the data, real as well as imputed, in all analyses.

(2) For each item with missing data, create weights that adjust for the item missing data and use only the real data with adjusted weights in all analyses.

(3) Use only the real data with unadjusted sampling weights (i.e., ignore the effects of item missing data) in all analyses.

(4) Use an adaptive strategy, selecting Strategy 1 or 3 depending on the type of analysis to be conducted.

These strategies were examined for a series of analyses of the NMCUES data (Berki et al., 1985; Parsons et al., 1986). We considered each strategy with respect to three criteria:

- The practicality of implementing the strategy without developing special-purpose software for analysis.
- The accuracy of estimates of totals or aggregates, means, and standard errors of means.
- The degree of attenuation of relationships between two measures attributable to the strategy.

After reviewing findings similar to those presented in this report, we chose to follow an enhanced version of Strategy 1. It is useful to examine the practicality and accuracy criteria for each strategy in the specific context of analysis of the NMCUES public use data and the weighted sequential hot-deck imputation procedure used for those data.

Strategy 1 is practical to implement because no special adjustments to standard analytic methods are necessary. Estimates of aggregates or totals (such as total charges for outpatient department visits) are automatically adjusted for the failure to obtain responses from some persons, and weighted means for the total sample are similar to those obtained by using only the real data. However, standard errors estimated using all the data and standard variance estimation techniques will tend to underestimate the actual standard errors, the severity of the underestimate depending primarily on the amount of imputed data. Moreover, relationships between imputed measures and variables not used as control variables in the imputation process can be attenuated. Strategy 1 can be enhanced by presenting estimates of standard errors with an indication that some estimates may be subject to substantial underestimation because of imputation. In addition, multivariable analyses should be conducted both with and without imputed values to assess the impact of imputation.

In contrast to Strategy 1, Strategy 2 is not practical to implement. It would be a sizable task to create adjusted weights for each item used in a large-scale analysis. In addition, the choice of an estimation procedure for a multivariable analysis involving two or more measures, each with a different weight, is not obvious. The advantage of the adjusted weight strategy is that standard errors could be estimated for estimated means using standard estimation procedures applied to the real data with adjusted weights. In addition, weighted means computed using only the real data would essentially be identical to those computed using all the data, estimated totals would not be subject to underestimation because of item missing data, and the strength of relationships among survey measures would not be as seriously attenuated as under Strategy 1.

Strategy 3, like Strategy 1, is practical to implement, although some recoding of data items may be necessary on most statistical software systems to identify imputed values as missing rather than real. The weighted means for the total sample provided by Strategy 3 do not differ substantially from those estimated using all the data. Standard errors can be estimated for these real data means, and relationships among survey measures will not be attenuated by the imputation process. However, estimated totals for items with substantial amounts of missing data will be severe underestimates if only real data are used.

Analysts employing the final adaptive strategy might use only the real data to estimate means and their standard errors and to analyze relationships among survey measures. Estimates of totals would be computed using all

the data to avoid severe underestimation for survey measures with large rates of item missing data. Under Strategy 4, one still would be faced with the estimation of standard errors for totals for which imputed values were used; a suitable standard error estimation procedure is not readily available for sample survey data that have values imputed across strata and primary sampling units. For the most part, an adaptive strategy is practical to implement because it does not require any particular analyses that cannot be done using standard statistical software. However, extensive analysis of a large survey data set may be cumbersome when analytic methods are tailored to the amount of item missing data for measures used in the analysis.

Strategy 2 cannot be recommended because it is impractical to make item missing data adjustments to weights for the numerous estimates needing adjustment. Moreover, it is difficult to use the adjusted weights in multivariable analysis. Similarly, for routine analysis of survey data, Strategy 3 cannot be recommended because of the potential for highly inaccurate and misleading estimation in analysis.

For analysis of surveys that are smaller and less complex than NMCUES and that involve a limited number of analyses, Strategy 4 may be a feasible alternative. For smaller analytic tasks, greater attention can be given to refinements that may improve the accuracy of the reported results. Clearly Strategy 4 cannot be recommended when the size or complexity of the analytic task increases.

For a large complex survey such as NMCUES, we recommend that Strategy 1 be followed. However, significant attention must be given to investigating the effects of imputation on analysis. In particular, analysts should familiarize themselves with rates of item missing data for key survey items and the imputation methods used to compensate for missing data. For items with significant amounts of missing data, estimates of means, totals, and standard errors ought to be made both with and without imputed values to assess the effects of imputation on those estimates. Similarly, analyses of substantively important relationships should be conducted both with and without imputed values to assess whether imputation is seriously attenuating the strength of the observed relationship.

When this enhanced form of Strategy 1 was used in analyses of NMCUES data, the analytic task was increased by the need to conduct some analyses both with and without imputed values. However, the increased burden was reduced somewhat by careful selection of the types of measures for which both types of estimates were computed and by identifying the relationships that were substantively most important, for which an assessment of the effects of imputation was needed.

The results of investigations of the effects of imputation on analysis may cause the alteration of the presentation and interpretation of findings. Standard errors for estimates that are based on measures with substantial amounts of imputation can be indicated in tabular or other types of presentations. This approach may not be feasible in many cases. If it is not feasible, reports on analytic findings must at least include warnings to the reader that estimates of standard errors for some measures may be substantial underestimates because of imputed values. Thus, the interpretation of analytic findings concerning relationships among survey measures can be informed and guided by knowledge about the impact of imputation on results.

Estimation Methods

Notion of an Average Population

NMCUES was a panel survey in which members of the population were followed during the panel period (calendar year 1980). The nature of a dynamic population over time influences the rules used to determine who should be followed and for how long. It also has significant implications for the form of estimators for characteristics of the population during the panel period. Before discussing estimation strategies for the NMCUES data, it is useful to review the nature of a dynamic population over time.

Figure 4 illustrates the nature of a longitudinal population as members move in and out of eligibility. Stable members of the population appear at the beginning and at every time point during the life of the longitudinal time period. Even though these persons are termed "stable," they may of course change residence during the panel period and may be quite difficult to trace. They may even be lost to followup during the course of the panel, but as long as they remain in the population, they are referred to as stable population members. Leavers are persons who are eligible at the beginning of a time period but become ineligible at some later time. Leaving may occur through events such as death, institutionalization, or moving outside the geographic boundary of the population. Those lost to followup are not leavers. At the same time, new members (entrants) may enter the population through births or through returns from institutions or from outside the geographic boundary of the population. Finally, there also will be population elements that are both entrants and leavers from the population during different time periods. The majority of the population typically will be stable in nature, but it is the entrants and leavers, persons who may be experiencing major changes in their lives, who are often of particular interest to analysts of panel survey data. In order to assure adequate coverage of all elements in the dynamic population considered over the entire time period, NMCUES followup rules were carefully specified to properly include entrants, leavers, and mixed population elements.

As an illustration, consider a person who was in the Armed Forces on January 1, 1980, was discharged on June 1, 1980, and then became a key person (i.e., one to be followed for the rest of the year while eligible) in the NMCUES panel. Because NMCUES was designed

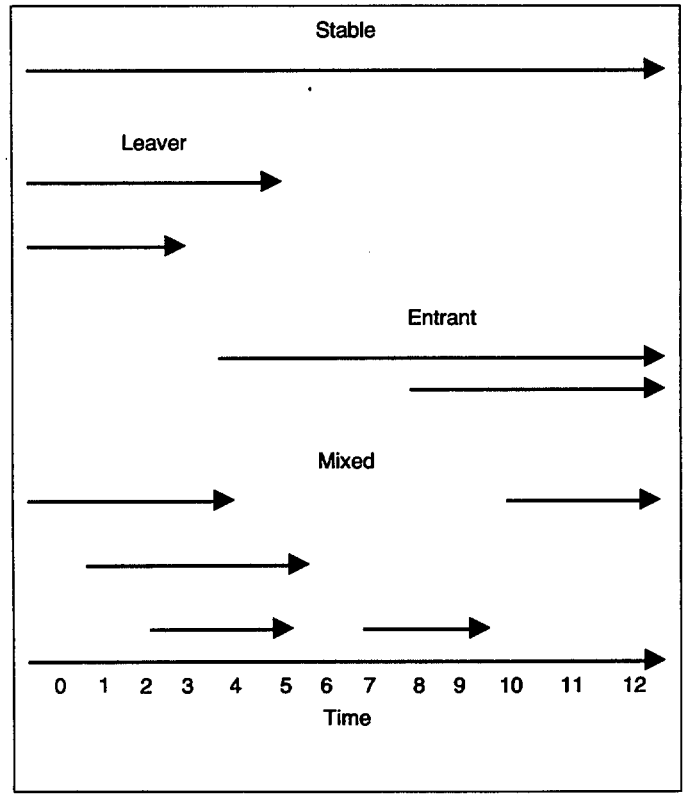


Figure 4
Dynamic population for 12 time period panel survey

to provide information about the civilian population, medical care use and charges for this person during the first 5 months of 1980 are outside the scope of the survey. Data about health care use and charges were not collected unless they occurred after June 1. At the same time, this person was eligible for only 7 months of the year, and he was also "at risk" of incurring health care use or charges for only 7 of the 12 months. This person thus contributes only 7/12, or 0.58, of a year of eligibility, or "person risk," to the study. This quantity is referred to as the "time-adjustment factor" in the NMCUES documentation and in this report.

For readers not familiar with the concept of "person years of risk," it may be useful to consider briefly the rules that were used to determine eligibility for a given person at a given moment during 1980. There were essentially two ways of becoming eligible for or entering the NMCUES eligible population. The obvious way was

to be a member of the U.S. civilian noninstitutionalized population on January 1, 1980, and hence to be a member of the original, or base, cohort about which inferences were to be made. The second way to become a member of the eligible population was to enter after January 1 through birth or through rejoining the civilian noninstitutionalized population during the year by returning from an institution, from the Armed Forces, or from outside the United States. There were also several ways by which persons who were eligible members of the population could become ineligible. Death obviously removes a person from further followup, as does institutionalization, joining the Armed Forces, or moving to a residence outside the United States. Information was collected to monitor the exact number of days that each person selected for NMCUES was eligible during the year. These eligibility periods are summarized by the time-adjustment factor on each record.

The use of "person years" to form sample estimates requires careful assessment of the characteristic to be estimated. Estimates that use only data collected from persons during periods of eligibility (e.g., total number of doctor visits or total charges for health care) do not need to account for time adjustments. Estimates for person characteristics (e.g., total population or proportion of the population in a given subgroup) must be based on person years to obtain estimates that correspond to the desired target population quantities. Some estimates require the use of the time-adjustment factor in the denominator but not in the numerator. For example, an estimate of the mean total charge per person for health care during 1980 must use the total charges for health care as a numerator without time adjustment. However, the denominator must be the number of person years that the U.S. population was exposed to the risk of such charges during 1980, a time-adjusted, or person-years, measure. The mean in this case is actually a rate of health care charges per person year of exposure for the eligible population in 1980.

When making estimates in which person years are important, the effect of the time-adjustment factor will vary depending on the subpopulation of interest, as can be noted in Table D. A cross-sectional cohort of N persons selected from the U.S. population on January 1, 1980, and followed for the entire year will contribute a total number of person years for 1980 that is smaller than N because of removals such as deaths or institutionalization. If entrants are added to the initial cohort during the year, the person years contributed by the initial cohort and the entrants may well exceed N, but it will still be less than the number of original cohort members plus the number of entrants.

The difference between persons and person years will vary by subgroups as well. Females 25–29 years of age on January 1 constitute a cohort for which few additions are expected because of entrants from institutions, the Armed Forces, or living abroad. Few removals are expected because of death, institutionalization, joining the Armed Forces, or moving abroad. On the other hand, males 80 years of age and over on January 1 will contribute a much smaller number of person years to the population than the total number of persons in the cohort at the beginning of the year, because a large number of the cohort will die or become institutionalized during the year.

Table D

Effect of person-year adjustment on counts and sampling weights for selected population groups: National Medical Care Utilization and Expenditure Survey, 1980

Population group	Sample size	Person years	Sampling weight	
			Basic	Adjusted
	Number		Number in thousands	
Total population	17,123	16,862.84	226,368	222,824
Female, 25–29 years	702	699.39	9,529	9,494
Male, 80 years and over . .	113	104.05	1,384	1,274
All persons born during 1980	251	121.02	3,560	1,713

Estimation Procedures

Sample estimators from the NMCUES data, regardless of whether they are totals, means, proportions, standard errors, or some other estimator, must account for the complexity of the sample survey design. Totals, means, and other estimates must include sampling weights to compensate for unequal probabilities of selection, nonresponse, and undercoverage. Stratification, clustering, and weighting must also be accounted for in the estimation of sampling errors. In addition, one must consider time-adjustment factors to account for persons not eligible for the entire year and imputations that were made to compensate for missing items. These factors all affect the precision and accuracy of the sample estimators to some extent, as discussed previously.

In NMCUES, a stratified multistage probability sample design was used to select the initial sample of reporting units followed in the survey. The stratification and clustered sample selection, as well as weighting adjustments, must be accounted for in the calculation of estimates. As described previously, the NMCUES sample design involved up to four stages of sample selection and stratification for each of those stages. It would be difficult to incorporate the full complexity of this sample design into estimation procedures.

Fortunately, it is not necessary to incorporate the full complexity of the design into the estimation procedures in order to obtain valid estimates of totals, means, standard errors, and other estimates. RTI has provided codes with the NMCUES data files that can be used to estimate various quantities of interest in a practical manner. Nonetheless, the estimation procedures for NMCUES are still complicated, and added notation is required to reflect special considerations needed in estimation.

For the NMCUES estimates, individuals are indexed by stratum, by primary sampling unit within stratum, and by individual within primary sampling unit. Codes are provided in the data set to identify the stratum and primary sampling unit for each person record. They are referred to in the public use data tape documentation as the pseudostratum and pseudoreplicate codes. The term "pseudo" is used because these codes do not correspond exactly with the strata and primary sampling units used in sample selection. In order to simplify subsequent variance estimation procedures, the NMCUES sampling statisticians grouped strata and primary sampling units used in the sample designs from the two survey organizations. For example, the actual sample selection may have involved such procedures as the selection of a single primary sampling unit within a stratum, but for purposes of variance estimation, strata have been collapsed such that exactly two primary sampling units are in a pseudo, or collapsed, stratum. There are 69 pseudostrata, each with exactly 2 pseudo-primary sampling units or pseudoreplicates, in NMCUES.

Let h denote the pseudostratum or stratum for each person in the NMCUES data set, where $h = 1,\ldots, 69$. Let a denote the pseudoreplicate within a particular stratum, where $a = 1$ or 2 only. Finally, let k denote the individual within a particular pseudoreplicate and pseudostratum, where $k = 1,\ldots, n_{ha}$. (The number of persons within pseudoreplicates varies across pseudoreplicates.) The (hak)th individual is uniquely identified by the pseudostratum and pseudoreplicate codes and a person-level code within those units.

A variety of estimators can be formulated from NMCUES data, depending on the focus of the analyses. To illustrate the role of time adjustments, weights, and other aspects of forming estimates for NMCUES, consider six estimates that appear in NMCUES reports:

(1) An estimated total charge for a selected subgroup, such as persons below the poverty level.

(2) An estimated total population, such as the number of females.

(3) The mean charge per visit.

(4) The mean charge per person.

(5) The proportion of persons whose charge is less than or equal to a fixed level.

(6) The proportion of all charges of a certain type that fall in a specified range of charges.

The following variables for the (hak)th individual will be used in the discussion of these six estimates.

y_{hak} = total charge for health care in 1980.

x_{hak} = total number of medical visits in 1980.

w_{hak} = nonresponse-adjusted and undercoverage-adjusted person weight.

t_{hak} = time-adjustment factor (the proportion of days in 1980 that the person was an eligible member of the population.

$$d_{hak} = \begin{cases} 1, & \text{if the total charge is less than or equal to a fixed value;} \\ 0, & \text{otherwise.} \end{cases}$$

$$e_{hak} = \begin{cases} 1, & \text{if the total charge is between two fixed values;} \\ 0, & \text{otherwise.} \end{cases}$$

$$\delta_{hak} = \begin{cases} 1, & \text{if the } k\text{th person is a member of a designated subgroup of the population;} \\ 0, & \text{otherwise.} \end{cases}$$

Estimating total charges from NMCUES, or any quantity that was recorded only during periods when the person was a noninstitutionalized civilian in the United States, is a relatively straightforward task requiring only a weighted sum of charge values. For example, the quantity

$$\hat{y} = \Sigma_h \Sigma_a \Sigma_k \, w_{hak} \, y_{hak} \, \delta_{hak}$$

is the estimated total charge for all health care received during 1980 for a selected subgroup. The total charge for health care for each individual in the subgroup is "inflated" by the person weight, and the weighted total charges are summed across strata, sampling error units, and persons.

In contrast, estimates of the population of persons require a time-adjusted estimator. For example, to estimate the size of a subgroup of the population, one would use

$$\hat{y}' = \Sigma_h \Sigma_a \Sigma_k \, w_{hak} \, t_{hak} \, \delta_{hak} \, ,$$

an estimate of the 1980 average subgroup population. The time-adjustment factor is inflated by the value of the person weight to obtain a person-year estimate of noninstitutionalized civilian members of the population for persons with characteristics similar to those of a given individual. The estimate is thus a weighted sum of time-adjustment factors.

Use of a time-adjustment factor is needed in the estimation of some types of means, but not all. For example, to estimate the mean charge *per visit* during 1980, no time adjustment is needed because the mean is based on a unit that was observed only during periods of person eligibility; that is, visits were recorded only during periods when the person was an eligible member of the population. Hence, the estimated mean per visit can be calculated as

$$\bar{y} = \Sigma_h \Sigma_a \Sigma_k \, w_{hak} \, y_{hak} \, / \Sigma_h \Sigma_a \Sigma_k \, w_{hak} \, x_{hak} \, .$$

In contrast, to estimate a mean *per person*, a time adjustment is required in the denominator, which is actually an estimate of the total average population in 1980. In particular, the estimated mean charge per person has the form

$$\bar{y}' = \Sigma_h \Sigma_a \Sigma_k \, w_{hak} \, y_{hak} \, / \Sigma_h \Sigma_a \Sigma_k \, w_{hak} \, t_{hak} \, .$$

Estimates of mean charges for subgroups have the same form as this estimator, with an indicator variable, δ_{hak}, added to the numerator and denominator of the mean for the appropriate subgroup of interest.

Estimated proportions can be formulated simply as means, with an indicator variable in the numerator indicating classification in a particular category and a count variable in the denominator for all units in the particular subgroup of interest. For proportions, time-adjustment factors may be used not only in the denominator, as for means, but also in the numerator. For example, to estimate the proportion of persons who had charges less than or equal to a fixed value, an estimate of the form

$$p' = \Sigma_h \Sigma_a \Sigma_k \, w_{hak} \, d_{hak} \, t_{hak} / \Sigma_h \Sigma_a \Sigma_k \, w_{hak} \, t_{hak}$$

is used. Appropriate indicator variables can be added to the numerator and denominator of the estimator to obtain estimates for a selected subgroup of the population.

On the other hand, an estimated proportion such as total charges within a range of charges does not require time adjustments in the numerator or the denominator. For example, the estimated proportion of all charges that were between two levels of charges is computed as

$$p = \Sigma_h \Sigma_a \Sigma_k \, w_{hak} \, y_{hak} \, e_{hak} / \Sigma_h \Sigma_a \Sigma_k \, w_{hak} \, y_{hak} \, .$$

Variance Estimates and Confidence Intervals

The NMCUES sample is one of a large number of samples that could have been selected, using identical sampling procedures, from the U.S. civilian noninstitutionalized population. Each of the possible samples could have provided estimates that would differ from sample to sample. The variability among the estimates from all the possible samples that could have been selected can be measured by the standard error of the estimates, or the sampling error. The standard error can be used to assess the precision of the estimate itself by creating a confidence interval. There is a specified probability that estimates over all possible samples selected from the population using the same sampling procedures will be within the confidence interval.

In variance estimation for stratified multistage sample surveys, the procedures used are different from the standard methods for simple random sampling. Often assumptions are made about the sample selection procedure that simplify variance estimation but tend to produce overestimates of the variance. It is beyond the scope of this report to examine the assumptions used to develop appropriate variance estimation procedures for NMCUES. Instead, one approach will be described for the types of estimators previously described.

Consider, for instance, estimating the sampling variance of the estimate of total charges for a subgroup, $\hat{y}$. Denote the weighted sum of charges for the (ha)th pseudoreplicate as

$$\hat{y}_{ha} = \Sigma_k \, w_{hak} \, y_{hak} \, \delta_{hak} \, .$$

The variance of $\hat{y}$ can then be estimated as

$$\mathrm{var}(\hat{y}) = \Sigma_h (\hat{y}_{h1} - \hat{y}_{h2})^2 \, ;$$

that is, the variance of $\hat{y}$ is computed as the squared differences of weighted pseudoreplicate totals summed across the 69 pseudostrata in the design. The computation requires that the pseudostratum and pseudoreplicate be available for each observation in the data file. These codes are supplied on the public use tapes for each type of data record.

Weighting is an implicit feature of this estimation procedure. The replicate totals are appropriately weighted, representing an estimate of the total value for the pseudo-primary sampling unit represented by the pseudoreplicate. The variance estimate is increased by the use of weights because of the larger variation in estimated totals introduced by the weighted values.

Two other features of the sample design are not included in this variance estimation procedure. Although the weights reflect a poststratification adjustment, the variance estimate does not account for poststrata in the calculation. Second, the variance arising from imputation is not included in this approach.

In estimation of means and proportions, these variance estimates for totals, as well as covariances of estimated totals in the estimation procedure, are used. However, because the denominators of the means and proportions are not fixed by the sample design but are actually random variables, means and proportions from NMCUES are ratios of random variables. The exact variances of such ratio means are not known, but in practice they can be approximated through a first-order Taylor series expansion. The Taylor series approximation can be programmed easily for routine estimation of the variances and covariances of ratio means.

Let $y = \Sigma_h \Sigma_a \Sigma_k \, w_{hak} \, y_{hak}$ and $x = \Sigma_h \Sigma_a \Sigma_k \, w_{hak} \, x_{hak}$ denote the numerator and denominator of the mean charge per visit estimator, $\bar{y}$, given previously. The Taylor series approximation to the variance of $\bar{y}$ is

$$\mathrm{var}(\bar{y}) \doteq \bar{y}^{\,2} \, [x^{-2}\mathrm{var}(x) + y^{-2}\mathrm{var}(y) - 2(xy)^{-1}\mathrm{cov}(x,y)],$$

where var(x) and var(y) are variances of estimated totals, described previously. The expression cov(x,y) is the covariance of the estimated totals x and y and is estimated as

$$\mathrm{cov}(x,y) = \Sigma_h (x_{h1} - x_{h2})(y_{h1} - y_{h2}),$$

where x_{ha} and y_{ha} are the sum of weighted charge and visit values, respectively, for the (ha)th pseudoreplicate. The variance estimates for the other mean, $\bar{y}'$, and for the proportions p' and p, considered previously, can

be constructed using the Taylor series approximation in a manner similar to that for $\bar{y}$.

Variances also are needed for differences between means and proportions, particularly differences between means or proportions from two subgroups. For example, a comparison of mean charges for males, denoted as $\bar{y}_1 = y_1 / x_1$, and for females, denoted as $\bar{y}_2 = y_2 / x_2$, requires an estimate of the variance of the difference. The stratified multistage sample design induces a nonzero and generally positive covariance between subgroup means even though the subgroups have no elements in common. Thus, the variance of the difference is computed as

$$\mathrm{var}(\bar{y}_1 - \bar{y}_2) = \mathrm{var}(\bar{y}_1) + \mathrm{var}(\bar{y}_2) - 2\,\mathrm{cov}(\bar{y}_1,\bar{y}_2).$$

The terms $\mathrm{var}(\bar{y}_1)$ and $\mathrm{var}(\bar{y}_2)$ are estimated by the Taylor series expansion approximation given previously, and the covariance term is estimated by the Taylor series approximation

$$\begin{aligned}\mathrm{cov}(\bar{y}_1,\bar{y}_2) = (\bar{y}_1\bar{y}_2)\{&(x_1x_2)^{-1}\,\mathrm{cov}(x_1,x_2) + \\ &(y_1y_2)^{-1}\,\mathrm{cov}(y_1,y_2) - (y_1x_2)^{-1}\,\mathrm{cov}(y_1,x_2) - \\ &(x_1y_2)^{-1}\,\mathrm{cov}(x_1,y_2)\}.\end{aligned}$$

Here, the covariances of totals are estimated in a manner similar to that described previously.

These procedures for estimating sampling errors from complex sample survey data are implemented in several statistical software packages. The sampling errors for the NMCUES companion reports were estimated using several programs available within the OSIRIS IV Statistical Software System (Computer Support Group, 1982). Sampling error programs that operate within the Statistical Analysis System (SAS) are available in the SESUDAAN package developed by the Research Triangle Institute (Shah, 1984). Other sampling error programs and packages are also available. Cohen, Burt, and Jones (1986) review the features of several of these.

The sampling variances calculated from these formulas can be used to form intervals for which confidence statements can be made regarding estimates from all possible samples drawn in exactly the same way as NMCUES. The confidence level is determined by multiplying the estimated standard error (the square root of the estimated variance) by a constant derived from the standardized normal probability distribution. In particular, for the estimate $\hat{\theta}$, with estimated standard error $ste(\hat{\theta})$, the upper limit for a confidence interval of $(1 - \alpha) \times 100$ percent can be formed by adding $z_{\alpha/2}$ times $\mathrm{ste}(\hat{\theta})$ to $\hat{\theta}$. The lower limit is formed by subtracting $z_{\alpha/2}$ times $\mathrm{ste}(\hat{\theta})$ from $\hat{\theta}$. The value of $z_{\alpha/2}$ is obtained from the standard normal probability distribution. For example, a 95-percent confidence interval corresponding to $\alpha = 0.05$ can be formed with $z_{0.025} = 1.96$; for a 99-percent confidence interval corresponding to $\alpha = 0.01$, $z_{0.005} = 2.346$ is used.

Confidence intervals for comparisons of estimates between two subgroups can be used to make inferences about whether a difference is statistically significant. If a confidence interval of $(1 - \alpha) \times 100$ percent does not include the value zero, one can conclude that the difference is significantly different from zero.

These confidence intervals depend on an assumption that the estimate $\hat{\theta}$ is a normally distributed random variable. For many types of estimators (e.g., medians, ranges), this assumption will not be appropriate. For others, normality of the sampling distribution of the estimate $\hat{\theta}$ depends on the Central Limit Theorem. If the sampling distribution is not normally distributed, the confidence interval based on the normality assumption will not include the actual population value the specified percentage of the time. However, for most sample survey estimates encountered in practice, the normality assumption is only one source of error among many in estimation. Typically in survey practice, the normality assumption is considered to be reasonable for most estimates.

Sampling variance estimates that are based on pseudoreplicate totals and incorporate the complexity of the sample design into estimation tend, on average, to be larger than estimates computed under the standard assumptions of independence between selections. Under SRS assumptions, the sampling variance of a sample mean would be estimated as

$$\mathrm{var}_{\mathrm{SRS}}(\bar{y}) = (1 - f)\,\frac{s^2}{n},$$

where $(1 - f)$ is the finite population correction, s^2 is the population element variance for the characteristic y, and n is the sample size. For weighted survey data, the population element variance can be estimated as

$$\hat{s}^2 = \frac{\Sigma_h\Sigma_a\Sigma_k\, w_{hak}(y_{hak} - \bar{y})^2}{\Sigma_h\Sigma_a\Sigma_k\, w_{hak} - 1}.$$

For means and proportions, the relative increase or decrease in sampling variance that can be attributed to the complex nature of the sample design is measured by the design effect,

$$\mathrm{deff}(\bar{y}) = \frac{\mathrm{var}(\bar{y})}{\mathrm{var}_{\mathrm{SRS}}(\bar{y})}.$$

The design effect is usually greater than one, indicating an increase in variance because of the complex sample design. The design effect is a summary measure of the effects of the design and includes the combined effects of stratification, clustered selection, and weighting.

Through empirical investigations of the design effect, Kish (1965) and others have observed that it varies by type of estimate and usually decreases with decreasing sample size for the same estimate as subgroups are examined. Based on these observations, the expression

$$\mathrm{deff} = (1 + [(n/A) - 1]\,roh)$$

has been proposed as a suitable summary model for design effects. Here, A denotes the total number of primary sampling units used in the calculation (A = 138 for NMCUES) and *roh* is a measure of similarity among observations from the same cluster. The value of *roh* is calculated from the design effect as $roh = (\text{deff} - 1)/[(n/A) - 1]$. As a consequence, *roh* is sometimes referred to as a "synthetic" measure of within-cluster homogeneity. In the instance of a complex sample design, *roh* also includes the effects of other design features such as stratification and weighting.

Sampling statisticians frequently use the design effect as a tool for examining the efficiency of existing and alternative survey designs. It may also be used to compute estimates of sampling errors. For example, if a value of *roh* and the sample size are available, the design effect can be computed. Using an estimate of the element variance, the sampling variance of an estimated mean or proportion can then be calculated as

$$\text{var}(\bar{y}) = \text{deff} \cdot \frac{\hat{s}^2}{n} .$$

A sample size should be shown in the table for the estimated mean or proportion. Users can be supplied with values for *roh* and $\hat{s}^2$ so that they can directly compute sampling error estimates using the design effect.

Generalized Variance Formulas

Computation of sampling errors for every estimate in the various NMCUES analysis reports would have been a sizable task. However, with currently available statistical software to implement the variance estimates, the cost and effort are not an unreasonable addition to much routine survey analysis. The presentation of sampling error estimates is a greater problem when the number of estimates presented in a survey analysis is considered. Presentation of a standard error for every estimate would increase the length of the report and detract from the clarity of the presentation of the estimates themselves.

A common practice in the analysis of survey data is to present a means for the reader of a report to derive or compute a standard error for a given estimate. Based on computations for a subset of estimates, empirical relationships are derived between basic information in a table (such as sample size, or the estimate itself) and the estimate of the standard error for a given estimate. The relationships that are derived can be presented in a variety of ways. For example, a variance curve in which sampling variance is related to sample size can be displayed graphically, allowing the user to obtain an estimated standard error for an estimate based on the number of cases used to calculate the estimate or the "base" estimated population size on which the estimate is computed. Alternatively, tables can be prepared that allow a reader who knows the sample size and value or type of an estimate (such as the type of charge or visit) to find an approximate standard error of the estimate. Formulas can also be developed that allow the reader of a report, with the aid of a calculator, to compute an estimate of a standard error directly using information such as sample size and value or type of estimate.

In each case, a large number of standard errors are first computed for a variety of types of estimates and sample sizes. Then analyses are conducted to determine the relationship between the standard errors and other quantities, such as sample size, the size of the estimate itself, or the type of estimate. For example, a set of functions may be fit through regression methods to determine a model for predicting a standard error from the sample size. Model fitting is conducted to obtain a generalized variance formula which produces a predicted standard error that is reasonably close to the actual estimated standard error.

The model fitting approach was used in the companion reports to derive generalized variance expressions for the estimates presented. Formulas were developed to allow computation of a predicted standard error using an electronic calculator with basic arithmetic operators and a square root function. The computed estimates are average, or smoothed, estimates of the estimated standard errors.

Formulas for standard error estimates were developed for three types of estimates:

- Totals or aggregates (e.g., total charges for all health services used in 1980, total person years for males).
- Means (e.g., per capita total charges, per capita charges for inpatient care for females).
- Proportions, percents, and prevalence rates (e.g., proportion of total charges paid for outpatient physician care; percent of the working-age population who were employed full time, full year in 1980).

Formulas for obtaining the variances of differences between estimates from two different subgroups of the population also were developed. Specifically, these formulas are used to compute standard errors for:

- Comparisons of two mutually exclusive subgroups (e.g., per capita total charges for males and females, male and female subgroups having no members in common).
- Comparisons between a subgroup and a larger group in which the subgroup is contained (e.g., total hospital stay charges for persons 65 years of age and over and for all persons in the NMCUES population).

The standard error of a difference is based on the standard error of the totals, means, proportions, percents, or prevalence rates of interest. Certain covariances between estimates, which typically are small relative to the standard errors of the estimates themselves, are ignored.

In the preparation of the companion reports, sampling error estimates were needed before a set of final tables to appear in the reports was specified. This requirement

posed a problem for the development of generalized variance formulas because the specific estimates that would appear in the report were not available for the model development. Table shells were prepared at an early stage of the analysis, and the subgroup definitions and variables used in those table shells served as a reference for the type of estimates that would appear in the final report tables. Generalized variance models were developed from the table shells, not from the final tables that appear in each report.

Similar types of estimates appear in many of the companion reports. For example, estimates of mean charge, proportion of persons, mean number of visits, and proportion of charges paid out of pocket are used in different ways in virtually all of the companion reports. At the same time, estimates that did not appear in any of the other reports were also used in individual reports. For example, in several reports about utilization and expenditures associated with specific conditions, such as cardiovascular disease or musculoskeletal conditions, condition-specific means and proportions were required. Often the nature of these estimates could not be specified during the preparation of table shells. As a result, generalized variance estimates were prepared in a two-stage process.

In the first stage, estimates and subgroups common to many or all of the reports were identified, and sampling errors were estimated for a large number of these. Generalized variance expressions were developed for these "core" estimates and are used in all of the reports. The second stage consisted of obtaining report-specific estimates and estimating sampling errors for them. Generalized variance expressions for these "special" estimates were then developed separately.

A large number of sampling errors were estimated in the preparation of generalized variance expressions. Numerous regression models were examined in order to develop suitable coefficients for a given model or to check the adequacy of prediction of the generalized variance expression for the estimated sampling error. The approach varied by the type of estimate being examined, whether a total, a mean, a proportion, or a difference. It is beyond the scope of this report to present the details of these methods. Instead, a summary of the methods and results is presented for five basic sampling error estimation problems: Totals, means, proportions, and two types of subgroup differences.

Totals—Estimates of several types of totals appear in the companion NMCUES reports: Persons or person years, charges, visits, acquisitions, disability days, and lost productivity. For each type of total, sampling errors were estimated for a large number of subgroups in the sample, as well as the total sample, using the sampling error program & PSTOTAL in the OSIRIS IV package. For some types of totals, sampling errors were estimated for several different types of estimates.

For example, sampling errors were estimated for 10 different types of total charges. Total as well as out-of-pocket charges were examined for ambulatory visits, emergency room visits, physician visits, hospital stays, and total for all types of services. For each type, subgroups were formed by cross-classifying several demographic and other variables. In particular, subgroups were formed by cross-classifying age (in nine categories) with type of health care coverage, age with sex, education with race, and age with perceived health status. More than 150 estimates were calculated for each type of charge examined.

In many surveys, sampling errors for totals have been observed to be directly related to the size of the estimate itself. Survey samplers have used this relationship to develop sampling error curves as well as generalized variance expressions so that they can derive a predicted sampling variance from the estimated total alone. Following these observations, the sampling errors computed for all types of charges were plotted against the estimate itself. A curvilinear relationship was observed, which suggested that a quadratic or perhaps logarithmic relationship existed between the sampling error and the estimated total. A variety of models were then fit to these data, and the relative adequacy of the fit was assessed using a simple coefficient of determination (R^2) criterion.

The best fitting models, both by type of charge and across types of charge, were based on a logarithmic relationship between the estimated total and the sampling error. However, the logarithmic model was rejected as unsuitable for the sake of simplicity and the desire to provide a generalized variance expression that could be used by a reader having a calculator with no more than a square root function.

A quadratic model,

$$\mathrm{var}(\hat{y}) = a\hat{y} + b\hat{y}^2,$$

also provided an adequate fit to the data both within and across different types of charges. For example, for ambulatory visit charges, the quadratic model accounted for 95.6 percent of the variation in estimated standard errors for 146 estimates examined. Although some predictive accuracy is lost by using a single model across types of charges, for simplicity a single quadratic model applied to all charge data was found to provide an adequate fit. An R^2 of 89.4 percent was achieved for 1,741 observations in an analysis of standard errors for totals.

The coefficients for the charges model, as well as for each of the other types of totals to appear in a given report, were then provided in an appendix to be used for approximating standard errors of totals. Letting $\hat{y}$ denote the estimated total or aggregate for which a standard error is desired, the standard error for the estimate can be calculated by the expression

$$\mathrm{ste}(\hat{y}) = [a\hat{y} + b\hat{y}^2]^{1/2},$$

where a and b are constants chosen from Table E for the particular estimate of interest.

As an illustration of the use of this formula, suppose

Table E

Coefficients for standard error formula for estimated aggregates or totals: National Medical Care Utilization and Expenditure Survey, 1980

Estimator	Coefficient	
	a	*b*
Person years	3.0476×10^4	4.7081×10^{-4}
Charges	1.0986×10^8	4.5524×10^{-4}
Lost productivity	1.1593×10^1	9.1757×10^{-4}
Visits, acquisitions, or disability days	4.6408×10^2	5.7634×10^{-1}

that the standard error of the estimated total charges for all health services for women 17–44 years of age is needed. Here, $\hat{y}$ = \$34,550,000,000, the estimated total health care charges accumulated in 1980 by women 17–44 years of age, as reported in Parsons et al. (1986). From Table E we obtain the coefficients $a = 1.0986 \times 10^8$ and $b = 4.5524 \times 10^{-4}$ to use in the formula to calculate the standard error of $\hat{y}$. The estimated standard error is then computed as

$$\begin{aligned} \text{ste}(\hat{y}) &= [(1.0986 \times 10^8)(34.55 \times 10^9) + \\ &\quad (4.5524 \times 10^{-4})(34.55 \times 10^9)^2]^{1/2} \\ &= [(3.7957 \times 10^{18}) + (5.4342 \times 10^{17})]^{1/2} \\ &= 2{,}083{,}100{,}000. \end{aligned}$$

This estimated standard error for the total $\hat{y}$ can be used to create confidence intervals for total charges for women 17–44 years of age. For example, a 68-percent confidence interval can be obtained by adding and subtracting the standard error from the estimate. In this case, in 68 out of 100 samples drawn in exactly the same way as NMCUES, the estimated total charges for women 17–44 years of age will range from \$32,467,000,000 to \$36,633,000,000. Similarly, a 95-percent confidence interval can be obtained by adding and subtracting from the estimate 1.96 times the standard error. Thus, for 95 out of 100 samples drawn in the same way as NMCUES, the estimated total charges for women 17–44 years of age will range from \$30,467,000,000 to \$38,633,000,000.

Means—Sampling errors for means are not, for the most part, directly related to the estimated mean itself. A more complicated relationship exists between the type of estimate and the sample size. Some regression analyses similar to those used for totals were examined, but the fit to the data was not satisfactory, so the approaches were not examined further.

The generalized variance expression for means follows the general form

$$\text{var}(\bar{y}) = \text{deff} \cdot \frac{\hat{s}^2}{\hat{n}} ,$$

where $\hat{n}$ is the estimated subgroup size in the population on which the estimated mean is based. A value of $\hat{n}$ is obtained from the table in which the mean is displayed, and values for *roh* used to compute deff and $\hat{s}^2$ are derived from the data.

As for totals, a large number of sampling errors were estimated for many different types of means appearing in the reports. The element variances varied considerably across different types of means. Many more sampling errors were computed for means than for totals. For example, for mean charges per person (or person year), separate estimates were made for all charges and for charges paid out of pocket. For each type of charge, sampling errors were examined for 12 different types of events, such as ambulatory visits, hospital stays, hospital outpatient department visits, and acquisition of a prescribed medication. For each type of charge and type of event, sampling errors were computed for more than 50 subgroups formed by cross-classifying age with health insurance coverage, age with sex, education with race, and family income with perceived health status.

Both the sampling error estimated using the appropriate stratified cluster variance formula and that estimated using the SRS variance formula were calculated for every estimated mean. The design effect, *roh*, and $\hat{s}^2$ were calculated from these two sampling error estimates. A variety of methods were then examined for obtaining a predicted sampling variance for the mean. An average *roh* and $\hat{s}^2$ were computed for each type of charge and type of event, and a predicted standard error was computed using these average values. Average *roh* values across types of events were also computed. Sometimes they were computed across all types of events and types of charges. Other times, *roh* values were averaged only across selected types of events for which they appeared to be of a similar order of magnitude.

Each set of standard error models involved the regression of the predicted standard error on the observed standard error, with regression forced to pass through the origin. The R^2 for the regression and the value of the slope coefficient were used to assess the adequacy of the prediction method for obtaining reasonably accurate estimates of standard errors using the estimated sample size, *roh*, and $\hat{s}^2$. (R^2 is not an entirely satisfactory measure of the goodness of fit of a linear regression model without an intercept, but it was used here as a convenient measure for selecting a suitable model.) For example, for mean charge per person based on all charges, average *roh* values were computed separately and across six different types of events for 75 different sampling error estimates. Three different predicted standard errors were computed: One based on an average *roh* and $\hat{s}^2$ for each type of event, a second based on an average *roh* and $\hat{s}^2$ across event types, and a third based on an average *roh* computed across event types but with an average $\hat{s}^2$ computed within event type. The predicted values were then compared with the observed standard errors using a simple linear regression model. The results are displayed in Table F.

The largest proportion of variance is explained when the average *roh* and $\hat{s}^2$ are computed within event type,

Table F

Results from regression of predicted standard errors on observed standard errors for mean charge per person under 3 different conditions: National Medical Care Utilization and Expenditure Survey, 1980

Computation of average		Sample		Slope
roh[1]	$\hat{s}^2$	size	R^2	coefficient
Within events	Within events	450	0.743	0.990
Across events	Across events	450	0.076	0.536
Across events	Within events	450	0.728	0.929

[1]Synthetic measure of intraclass homogeneity.

but the proportion explained when only *roh* is averaged across event types is quite similar. Clearly, averaging $\hat{s}^2$ across event types is not satisfactory. Whether or not *roh* is averaged across event types, the predicted value tends to underestimate the observed value slightly because the slope coefficient is less than 1.0. To simplify presentation and use, the generalized variance expression using an average *roh* across types of events but $\hat{s}^2$ averaged within type of event was chosen.

Similar analyses were conducted for many types of means: Mean charges per event, mean charges per user, mean visits per user, mean visits per person, mean percent paid out of pocket, mean length of hospital stay, and others. In each case, the appropriateness of computing predicted values using average *roh* and $\hat{s}^2$ values was examined, usually across different types of events. In the end, a decision was usually made to provide the reader with an average *roh* value across types of events but an average $\hat{s}^2$ computed within an event type. In nearly all cases, the proportion of variance among observed standard errors that was explained exceeded 70 percent. In one instance, it was discouragingly low (approximately 25 percent). However, few means of that particular type appeared in any given report. Hence it was decided that, despite the poor predictive ability of the generalized variance expression in that instance, the average *roh* and $\hat{s}^2$ values would be used.

The formula recommended for estimating the standard error of a mean can be calculated as

$$\text{ste}(\bar{y}) = \left[\text{deff} \cdot \frac{\hat{s}^2}{\hat{n}}\right]^{1/2}$$

$$= \left[\left\{1 + \left(\frac{\hat{n}}{1{,}795{,}637} - 1\right) roh\right\} \cdot \frac{\hat{s}^2}{\hat{n}}\right]^{1/2},$$

where $\hat{n}$ is the estimated sample size on which the estimated mean is based. The values of *roh* and $\hat{s}^2$ for a variety of means can be obtained from Table 17. Values of *roh* and $\hat{s}^2$ for mean charges and mean utilization measures of various types are included in the table.

As an illustration, suppose that the standard error of the per capita charges for all health care in 1980 for males 17–44 years of age is needed. Here, $\bar{y}$ = \$473 for males 17–44 years of age. The values *roh* = 0.029644 and $\hat{s}^2 = 7.2407 \times 10^{10}$ are obtained from Table 17 under the entry for "Mean charge per person, All charges, Total." There were an estimated $\hat{n}$ = 45,576,000 males 17–44 years of age. Substituting these values into the expression for ste($\bar{y}$),

$$\text{ste}(\bar{y}) = \left[\left[1 + \left(\frac{45{,}576{,}000}{1{,}795{,}637} - 1\right)(0.029644)\right] \cdot \frac{7.2407 \times 10^{10}}{45{,}576{,}000}\right]^{1/2}$$

$$= \left[[1 + (25.382 - 1)(0.029644)]\,(1588.7)\right]^{1/2}$$

$$= [(1.7228)(1588.7)]^{1/2}$$

$$= 52.316.$$

The standard error of the per capita total charges for males 17–44 years of age is \$52.32.

Approximate confidence intervals can be constructed for the population mean by adding to and subtracting from the estimated mean a constant times the estimated standard error. For example, to form a 95-percent confidence interval for estimated per capita charges for males 17–44 years of age, 1.96 times the estimated standard error is added to and subtracted from the estimated mean, $\bar{y}$ = \$473. In this case, the 95-percent interval ranges from \$370 to \$576.

When the estimated sample size is about the same size or smaller than the constant 1,795,637 in the standard error formula, the design effect estimate will be less than or equal to one. When $\hat{n} \leq$ 1,795,000, the design effect portion of the standard error formula is not used, and the estimated standard error can be calculated simply as

$$\text{ste}(\bar{y}) = [\hat{s}^2/\hat{n}]^{1/2},$$

where $\hat{s}^2$ is again chosen from Table 17.

For example, there are an estimated $\hat{n}$ = 1,468,000 unemployed males. To estimate the standard error of the per capita charges for all health care for these persons in 1980—$\bar{y}$ = \$802 from Parsons et al. (1986)—the value $\hat{s}^2 = 7.2407 \times 10^{10}$ is obtained from Table 17 as before and

$$\text{ste}(\bar{y}) = \left[\frac{7.2407 \times 10^{10}}{1{,}468{,}000}\right]^{1/2}$$

$$= 222.09.$$

An approximate 95-percent confidence interval for the per capita charges is obtained by adding to and subtracting from the mean (\$802) 1.96 times the standard error, or approximately \$435. Thus, the 95-percent interval ranges from \$367 to \$1,237.

Proportions, percents, and prevalence rates—Models for the standard error of a proportion were developed in a manner similar to that used for means. The formula recommended for computing a standard error estimate for a proportion is therefore similar to that recommended for the standard error of a mean. However, because the element variance for a proportion can be estimated directly from the value of the proportion itself, the approach can be simplified somewhat. Let $\hat{p}$ denote the estimated proportion for which a standard error is needed. The population variance can be estimated simply as

$$\hat{s}^2 = \hat{p}(1 - \hat{p}).$$

Hence, no value of $\hat{s}^2$ must be displayed in tables with *roh* values.

Following this simplification, the standard error for a proportion $\hat{p}$ can be estimated as

$$\text{ste}(\hat{p}) = \left[\{1 + (\frac{\hat{n}}{1{,}795{,}637} - 1)\, roh\}\frac{13{,}012 \cdot \hat{p}(1-\hat{p})}{\hat{n}}\right]^{1/2},$$

where $\hat{n}$ is the estimated sample size on which the proportion is based, *roh* is a value selected from Table G, and the constant 13,012 is the average time-adjusted weight for all persons in the sample. The design effect—the ratio of the actual sampling variance for the estimated proportion to the standard error that would be achieved for a simple random sample of the same size—is calculated for proportions in the same way as it was calculated for means.

Sampling errors were estimated for a large number of proportions of various types. After extensive analyses similar to those used for means, several basic types of proportions were identified for which separate *roh* values were needed. For example, proportion of persons (or person years), proportion of charges, proportion of charges paid out of pocket, and proportion of visits each had sufficiently different average *roh* values that the predictive accuracy of the generalized variance expression was improved by using different average *roh* values for them. In the regression of predicted on observed standard errors for the proportion of person years, for instance, the use of an average *roh* value across 285 different standard errors yielded an R^2 of 0.903 and a slope coefficient of 1.061. In this case, the predicted standard error was slightly smaller, on average, than the observed value, but this underestimation was not considered a serious problem.

As an illustration of the use of the formula for ste($\hat{p}$), consider obtaining the standard error for the proportion of total productive person years lost from morbidity ($\hat{p}$ = 0.667) attributable to the population unable to work. The formula used to calculate the standard error for proportions can also be used for percents that have been divided by 100. There are an estimated $\hat{n}$ = 7,885,000 persons in the category, and *roh* = 0.069992 (Table G). Substituting these values into the standard error formula for $\hat{p}$, we obtain

Table G

Values of *roh* for standard error formula for estimated proportions: National Medical Care Utilization and Expenditure Survey, 1980

Estimator	*roh*[1]
Person years	0.069992
Charges	0.041917
Charges paid out of pocket	0.019816
Visits	0.084014

[1]Synthetic measure of intraclass homogeneity.

$$\begin{aligned}\text{ste}(\hat{p}) &= \left[[1 + (\frac{7{,}885{,}000}{1{,}795{,}637} - 1)(0.069992)]\ \frac{13{,}012 \cdot (0.667)(1-0.667)}{7{,}885{,}000}\right]^{1/2} \\ &= \left[\left[1 + (3.3912)(0.069992)\right]\frac{2{,}890.1}{7{,}885{,}000}\right]^{1/2} \\ &= [(1.2374)(3.6653 \times 10^{-4})]^{1/2} \\ &= 0.021297.\end{aligned}$$

This estimated standard error of 0.021297 is for the proportion $\hat{p}$ = 0.667, derived from the percent of interest. To obtain the standard error of the percent, simply multiply ste($\hat{p}$) by 100 for a standard error of 2.1297 for the percent 66.7.

An approximate 95-percent confidence interval for the percent can now be calculated by adding to and subtracting from the estimated percent 1.96 times the estimated standard error. Thus, for the percent of total productive person years lost from morbidity that is attributable to the population unable to work, the 95-percent interval ranges from 62.5 to 70.9 percent.

When the estimated sample size is less than or equal to 1,795,637, the design effect is close to one. The formula can then be simplified to

$$\text{ste}(\hat{p}) = \left[\frac{13{,}012\,\hat{p}(1-\hat{p})}{\hat{n}}\right]^{1/2},$$

as previously described for the standard error of a mean. For example, 72.7 percent of the value of productivity lost because of morbidity in the black population is attributable to those who are unable to work. For the $\hat{n}$ = 1,153,000 estimated persons in this subcategory, the standard error of the proportion associated with this percent is estimated as

$$\left[\frac{13{,}012 \cdot (0.727)(1 - 0.727)}{1{,}153{,}000}\right]^{1/2} = 0.047327.$$

A 95-percent confidence interval for the estimated percent is calculated by multiplying this estimated standard error by 100 × 1.96, or 9.3, and adding the result to and subtracting the result from the percent. Thus, the 95-percent interval ranges from 63.4 to 82.0 percent.

Differences among mutually exclusive subgroups—Many comparisons between the same estimate for two different subgroups in the population are made in the companion reports. Let $\hat{d} = \hat{\theta}_1 - \hat{\theta}_2$ denote the difference between two subgroup estimates, $\hat{\theta}_1$ and $\hat{\theta}_2$. For example, suppose that the per capita charges for females 17–44 years of age are to be compared with the per capita charges for males in this age group. Then $\hat{\theta}_1 = \bar{y}_1 = \$711$ for females 17–44 years of age, $\hat{\theta}_2 = \bar{y}_2 = \$473$ for males 17–44 years of age, and $\hat{d} = \bar{y}_1 - \bar{y}_2 = \$238$.

The standard error of this difference can be computed using a Taylor series approximation similar to that described previously. However, a simpler method was needed for routine use by readers of companion reports. The standard error of the difference $\hat{d}$ is

$$\text{ste}(\hat{d}) = [\text{var}(\hat{\theta}_1) + \text{var}(\hat{\theta}_2) - 2 \cdot \text{cov}(\hat{\theta}_1, \hat{\theta}_2)]^{1/2},$$

where $\text{var}(\theta_1)$ and $\text{var}(\theta_2)$ are the estimated sampling variances for $\hat{\theta}_1$ and $\hat{\theta}_2$, and $\text{cov}(\hat{\theta}_1, \hat{\theta}_2)$ is the estimated covariance between $\hat{\theta}_1$ and $\hat{\theta}_2$. The covariance between the two subpopulation estimators $\hat{\theta}_1$ and $\hat{\theta}_2$ arises in clustered sampling because elements from the two different subpopulations are typically present within each cluster and, hence, were not selected independently.

In examining the difference between per capita charges for males and females ages 17–44 years, it should be noted that both males and females appear in each of the 138 clusters formed for sampling error estimation purposes. If the per capita charges for males and females are positively correlated within clusters, the covariance term $\text{cov}(\hat{\theta}_1, \hat{\theta}_2)$ will be nonzero and positive in value. The size of the covariance will depend on the strength of the correlation between values of male and female per capita income within clusters. Alternatively, subpopulations may be formed for which the subpopulation covariance is negligible or zero by definition. For instance, comparing per capita income for two nonoverlapping geographic regions will eliminate any overlap between cluster elements, and the subpopulation covariance will be zero. However, most subpopulation comparisons of mutually exclusive subgroups will involve subpopulations that have elements in the same clusters.

From a large number of empirical results, Kish (1965) has noted that, for subpopulations that tend to be distributed across all or most clusters, the subpopulation covariance $\text{cov}(\hat{\theta}_1, \hat{\theta}_2)$ is, on average, positive in value and small relative to the variances of the two subpopulation estimates. To the extent that this observation is correct for a given subpopulation difference, ignoring the subpopulation covariance when computing the variance of the subpopulation difference will overestimate the variance. Across a large number of subpopulation differences for which, on average, the subpopulation covariance is small and positive in value, the variance of the subpopulation difference computed ignoring the covariance will also tend to be an overestimate. For some instances, the overestimation may be severe, and in others when the covariance is negative, the variance of the difference without the covariance will be an underestimate. For the sake of simplicity, we decided to accept the error inherent in ignoring the subpopulation covariance, anticipating that, on average, the estimated variance would tend to be an overestimate.

In the companion reports, comparisons were made among subgroups for so many different types of estimates that it was difficult to present standard errors for differences. Therefore, it was decided to accept an overestimation of the standard error of a subgroup difference. In particular, the standard error of the difference between estimates from two mutually exclusive subgroups is given as

$$\text{ste}(\hat{d}) \simeq [\text{var}(\hat{\theta}_1) + \text{var}(\hat{\theta}_2)]^{1/2}.$$

Returning to the example of computing the standard error of a difference in per capita charges between males and females 17–44 years of age, $\hat{n}_1 = 48{,}626{,}000$ and $\hat{n}_2 = 45{,}576{,}000$. It can be seen from Table 17 that $roh = 0.029644$ and $\hat{s}^2 = 7.2407 \times 10^{10}$. Thus,

$$\text{ste}(\bar{y}_1) = \left[\left(1 + \left(\frac{48{,}626{,}000}{1{,}795{,}637} - 1\right)(0.029644)\right) \frac{7.2407 \times 10^{10}}{48{,}626{,}000}\right]^{1/2}$$

$$= 51.384$$

and

$$\text{ste}(\bar{y}_2) = \left[\left(1 + \left(\frac{45{,}576{,}000}{1{,}795{,}637} - 1\right)(0.029644)\right) \frac{7.2407 \times 10^{10}}{45{,}576{,}000}\right]^{1/2}$$

$$= 52.316.$$

Hence, the standard error of the difference is computed as

$$\text{ste}(\hat{d}) = \left[(51.384)^2 + (52.316)^2\right]^{1/2} = 73.330.$$

This standard error can be used to form an approximate confidence interval for the difference, as described previously for estimates of totals, means, proportions, percents, and prevalence rates. In this instance, the 95-percent confidence interval ranges from \$94.27 to \$381.73. Because this interval does not include the value zero, one could conclude with 95-percent confidence

that per capita charges differ for the two categories. In other words, the chances are only 5 in 100 that the difference over a large number of identical surveys will be equal to zero.

Differences between subgroups and total group— Another type of comparison made in the companion reports is between an estimate for a subgroup and the same estimate for a group that contains the subgroup. Because one group is entirely contained within another, some simplification is possible in deriving a generalized variance expression.

Let $\hat{d} = \hat{\theta}_1 - \hat{\theta}_T$ denote the difference between a subgroup estimate and the estimate for a group in which the subgroup is contained, where $\hat{\theta}_1$ is the subgroup estimate and $\hat{\theta}_T$ is the estimate for the larger group. Let s_1^2 denote the element variance for the subgroup and s_T^2 the element variance for the total population, and let n_1 and n_T denote the sample sizes from the two respective groups. If an SRS had been selected, the variance of the difference $\hat{d}$ could be estimated as

$$\mathrm{var}_{\mathrm{SRS}}(\hat{d}) = \frac{s_1^2}{n_1} + \frac{s_T^2}{n_T} - 2\,\frac{r_{1T}\,s_1\,s_T}{\sqrt{n_1 n_T}}\ ,$$

where r_{1T} is the correlation between the subgroup and the total sample. Under the assumption that $s_1^2 = s_T^2$, and noting that for random subgroups $r_{1T} \doteq \sqrt{n_1/n_T}$,

$$\mathrm{var}_{\mathrm{SRS}}(\hat{d}) = \frac{s_1^2}{n_1}\,[1 - (n_1/n_T)].$$

To the extent that the element variance for the subgroup differs from that for the total population, this expression will be an underestimate of the actual SRS variance for most reasonable values for s_1^2 relative to s_T^2.

Following this argument for simple random sampling, the standard error of the difference $\hat{d}$ for a complex sample design such as NMCUES may be approximated as

$$\mathrm{ste}(\hat{d}) = \mathrm{ste}(\hat{\theta}_1)[1 - (\hat{n}_1 \,/\, \hat{n}_T)]^{1/2},$$

where $\mathrm{ste}(\hat{\theta}_1)$ denotes the standard error of the estimator θ_1, and $\hat{n}_1$ and $\hat{n}_T$ denote the estimated sample sizes for the subgroup and for the larger group, respectively. This formula is based on an assumption that the covariance between $\hat{\theta}_1$ and $\hat{\theta}_T$ is the same as the variance of $\hat{\theta}_1$ (i.e., $\mathrm{var}(\hat{\theta}_1)$). The assumption results in an estimated standard error for the difference that is, on average, somewhat larger than the actual standard error.

The adequacy of this formula was assessed for the NMCUES companion reports by computing approximately 50 differences between subgroups and the total sample, as well as the standard errors of those differences, using sampling error software in the OSIRIS IV system. These observed standard errors were compared with those obtained from the simplified expression. The predicted values of the standard errors were then regressed on those observed from the simplified expression, with the regression forced to pass through the origin. The predicted values explained 99 percent of the variation in the observed values, with a slope coefficient of 0.921. This fit was considered adequate for the purpose of providing a practical generalized variance expression for comparisons between subgroups and the total population.

As an illustration of the use of this expression, suppose that the standard error of the difference between per capita total charges for black persons and per capita total charges for all persons is needed. Here, $\hat{\theta}_1 = \bar{y}_1 = \$573$, $\hat{\theta}_T = \bar{y}_T = \$690$, $\hat{n}_1 = 26{,}046{,}000$, and $\hat{n}_T = 222{,}824{,}000$. Using the formula for estimating the standard error of the mean and values from Table 17, ($\hat{s}^2 = 7.2407 \times 10^{10}$ and $roh = 0.029644$),

$$\mathrm{ste}(\bar{y}_1) = \left[\left(1 + \left(\frac{26{,}046{,}000}{1{,}795{,}637} - 1\right)(0.029644)\right)\frac{7.2407 \times 10^{10}}{26{,}046{,}000}\right]^{1/2}$$

$$= 62.393.$$

Hence, the standard error of the difference, $\hat{d} = \$573 - \$690 = -\$117$, is computed as

$$\mathrm{ste}(\hat{d}) = 62.393\,[1 - (26{,}046{,}000 \,/\, 222{,}824{,}000)]^{1/2}$$

$$= 58.633.$$

A 95-percent confidence interval can be constructed for the difference by adding to and subtracting from the estimated difference 1.96 times the estimated standard error of the difference. In this instance, the 95-percent confidence interval is from $-\$231.92$ to $-\$2.08$. One can conclude with 95-percent confidence that black persons have lower per capita total charges than all persons because this confidence interval does not include zero.

Analytic Strategies

Data from NMCUES can be used to address a wide variety of policy and scientific questions. Answers to many of those questions can be sought through statistical analysis of relationships among measures available on the NMCUES public use files. In many situations, the statistical methods involve use of descriptive estimators, such as totals, means, or percentiles, compared across subgroups of the population. For example, the relationship between charges for hospital outpatient department visits and race can be investigated by comparing mean charge per hospital outpatient department visit for white and black subgroups. In the previous section, a confidence-interval method for making such comparisons is described.

Although subgroup comparisons can provide answers to simple questions, they are inadequate for many of the more complex questions investigators seek to answer using NMCUES data. For example, racial differences in mean charges for outpatient department visits may be attributable to differences in income or health care coverage by race. Means for race groups can be compared within income or health care coverage subgroups to control for these factors. However, a preferred statistical approach is to fit a regression or analysis-of-variance model to the survey data. From the model, the joint effect of several predictors on the response variable can be assessed. More sophisticated statistical methods are available for continuous as well as categorical dependent variables.

This section contains a brief review of the application of various regression and analysis-of-variance models to NMCUES data. Several strategies for survey analysis in which the sample design is incorporated in the analysis are described. Regression and analysis-of-variance methods for analyzing continuous as well as categorical types of dependent survey variables are then briefly reviewed. The discussion of strategies and regression and analysis-of-variance (ANOVA) methods is a summary of Landis et al. (1982). More detailed treatment of these topics is given in their report.

Many NMCUES measures have distributional properties that are awkward for standard methods of statistical analysis. For example, data on selected types of expenditures are available only for persons who utilized the health care system, and these data are quite skewed in distribution. In a subsequent section, two techniques are described that are appropriate for the analysis of these NMCUES measures.

Survey Analysis Strategies

The complexity of the NMCUES sample design can be incorporated into an analysis by using the sampling weights provided with the data files and computing variance and covariance estimates using procedures such as those described previously. However, some analysts may not have access to statistical software that handles weights properly or has programs for estimating sampling errors correctly. Even when software is available to compute sampling errors, the computation is more expensive than simple random sampling calculations. Further, empirical evidence suggests that, for some surveys, the effects of the sample design on analytic results becomes smaller as the analytic statistics become more "complex" (Kish and Frankel, 1974). It might appear that ignoring the sample design in an analysis of sample survey data has several attractive features: Standard, more familiar software can be used; computations are less expensive; and more complicated analyses are less likely to be affected by the sample design. However, other statistical issues must be considered in choosing an appropriate analytic strategy.

Landis et al. (1982) describe three ways that survey data can be analyzed based on a consideration of the use of sampling weights and the type of sampling error computations employed, as follows:

	Feature used:	
Option	*Weighting*	*Sampling error estimation*
1	Unweighted	Simple random
2	Weighted	Simple random
3	Weighted	Complex

NMCUES analyses conducted entirely under option 1 may produce biased estimates of totals, means, and other point estimates, and it can be expected that the inferences will be based on estimated levels of precision that are smaller than appropriate given the sample design. On the other hand, option 1 analyses are readily performed using standard statistical analysis software. Option 2 analyses require software that computes weighted estimates, which will be adjusted for unequal probabilities of selection and nonresponse and noncoverage errors. The estimated levels of precision under option 2 will still be smaller than appropriate. Option 3 requires the use of more specialized statistical analysis software, and it is the most appropriate form of analysis for

NMCUES data. As described previously, even option 3 analysis does not correctly account for all the complexities of the NMCUES design. Several approximations and assumptions have been used to simplify procedures.

As a practical matter, analysis options 1 and 2 may be useful in conducting preliminary analyses to identify findings that are quite likely to be the same as findings under option 3 analysis. For example, Landis et al. (1982) describe a sequential strategy in which standard statistical software is used to conduct option 1 analyses and find statistically significant relationships. More definitive analyses are then conducted for the option 1 statistically significant relationships using the sampling weights and complex sampling error estimation procedures of option 3. Some conclusions from the option 1 analysis may be reversed under option 3, but all final models and inferential conclusions appearing in a report are based on estimated covariance structures that account for the complexities of the sample design.

Of course, under a sequential strategy, option 1 and 2 relationships that are nonsignificant will not be reanalyzed using the option 3 methodology. Thus, there is no opportunity to detect relationships in the less likely situations in which the sample design actually reduces variances relative to option 1 or 2. Although estimated variances from complex samples may be smaller than estimates computed under simple random selection assumptions, reductions in true variance are generally considered to be unusual in highly clustered sample designs. To protect against failing to detect an important relationship under the sequential strategy, substantive relationships of particular interest can be investigated under option 3 even if the option 1 or 2 analysis is nonsignificant.

Although a sequential strategy may reduce the cost of analyzing survey data, the need to process some analyses more than once, using different statistical software each time, is an unattractive feature. In addition, statistical software that appropriately handles the sample design in the analysis has become more readily available recently. For example, Shah (1984) describes the SESUDAAN statistical package, which, although not part of the popular Statistical Analysis System (SAS Institute, Inc., 1985), operates within the SAS environment. The SESUDAAN package includes programs for generating tables of means, proportions, or totals and their standard errors as well as a general linear model program. A set of linked SAS procedures has also been developed that estimate variances for regression and logistic regression statistics incorporating the sample design in the calculations (Mohadjer et al., 1986). The OSIRIS Statistical Software System (Computer Support Group, 1982) includes sampling error estimation programs to generate tables of estimates and their standard errors, a general ratio estimation program, and a regression analysis program that incorporates the sample design in the estimation. All of the statistical programs within the OSIRIS system are able to handle weights in estimation. SUPERCARP (Hidiroglou, Fuller, and Hickman, 1974) and a version for microcomputers called PC CARP (Schnell et al., 1986) are other separate packages of sampling error and regression analysis programs that do not operate within a general statistical software system. Cohen, Burt, and Jones (1986) review the features and performance of some of the programs from several of these packages.

There are still many types of statistical analyses for which suitable software has not been developed, and the sequential strategy may be useful for those analyses. However, given the increasing availability of software designed specifically for the analysis of survey data, the sequential strategy suggested by Landis et al. (1982) can be replaced by a straightforward use of option 3 analyses alone when a user has access to appropriate statistical software.

Multiple Regression Methods for Continuous Dependent Variables

Consider the joint effects of several factors—such as age, sex, race, and family income—on physician visit charges for persons visiting a doctor's office in 1980. A general model for assessing these effects is the multiple regression model

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_p x_{pi} + e_i,$$

where y_i denotes the value of the dependent variable Y for the ith individual, x_{ki}, $k = 1, 2, \ldots, p$ denotes the value of the kth predictor or explanatory variable for the ith individual, e_i denotes random variation of y_i, and the β_k are constants. β_0 is the intercept term in the model. β_k are the partial regression coefficients and represent the change in the expected value of the dependent variable that corresponds to a unit change in the kth predictor.

The classical assumptions for this model are

- The model is correctly specified (e.g., no important factors have been omitted from the model).
- The x_{ki} are fixed constants (not stochastic).
- The x_{ki} are linearly independent.
- The e_i are independent and identically distributed $N(0, \sigma^2)$ random variables.

Even for nonsurvey data, it is unlikely that all of these assumptions will be met. However, departure from one or more of these assumptions may not seriously affect inferences made from a fitted model. For example, dependent variables such as charges are skewed in distribution, and hence the random error term e_i is not normally distributed. A transformation may be employed to provide a new dependent variable that is approximately normally distributed. On the other hand, for some dependent variables, transformations may be of little value, but the procedure may still be robust under such departures when the sample sizes are large.

There are several approaches to estimation of model parameters and their standard errors for survey data. Model parameter estimates are computed using the standard least-squares or weighted least-squares estimation formulas. The standard errors of the parameter estimates can be computed by a Taylor series expansion approximation. Alternatively, pseudoreplication methods can be used to compute standard error estimates. In pseudoreplication, replicates of the sample design are created by selecting sets of primary sampling units (or sampling error computing units) from across the strata in the design. For example, to form a replicate of the design in NMCUES, a single sampling error computing unit is selected at random from each of the 69 paired units used in sampling error calculations. Other replicates are then formed by choosing single units at random from each stratum. A large number of such replicates are formed, and the regression model terms are then estimated once for each replicate and once for the overall sample. The standard errors of the estimates are computed as functions of the variability of the replicate estimates relative to the overall sample estimates. Rust (1986) provides a more detailed description of these estimation procedures, and Landis et al. (1982) illustrate their use for a complex sample survey.

ANOVA Methods for Continuous and Categorical Dependent Variables

When the predictor variables are factors X_k, each of which has several levels, the familiar analysis-of-variance methods are appropriate for a continuous response variable, Y. Cross-classifications of the levels of the factors are used to divide the sample into a number of groups, and a measure of the total variation in Y, such as the sum of squared deviations about the sample mean, is partitioned into components, each of which corresponds to a factor or group of factors in a model. The usual hypothesis tests can be examined under assumptions of equality of variance across groups, independence among groups, and simple random selection.

ANOVA for data from a complex survey such as NMCUES requires modification of the standard procedures. One approach is the large sample weighted least-squares algorithms described by Grizzle, Starmer, and Koch (1969) for the analysis of multivariate categorical data. This methodology was adapted to ANOVA for complex survey data by Koch and Lemeshow (1972) and Koch, Freeman, and Freeman (1975). A vector of subclass means, say $\boldsymbol{F}$, is estimated from the survey data using appropriate weighted estimation procedures. In addition, the variances and covariances of the elements of $\boldsymbol{F}$ are estimated consistently using an approach such as the Taylor series approximations, described earlier. A general linear model framework is then used to develop variational models for $\boldsymbol{F}$ relative to its associated covariance matrix $V_{\boldsymbol{F}}$. The usual ANOVA hypotheses about which factors or combinations of factors contribute to variation among group means are investigated by fitting linear models to $\boldsymbol{F}$ by weighted least-squares methods. Landis et al. (1982) describe and illustrate this approach for a complex sample survey.

Landis et al. also describe the application of the same generalized weighted least-squares linear model-fitting methodology to the analysis of multidimensional contingency table data from complex sample surveys. Following the multifactor multiresponse framework described by Bhapkar and Koch (1968), independent variables or factors, as well as dependent variables or responses, are identified among the variables used to define the rows and columns of a multiway contingency table. The factors are cross-classified to define subpopulations or rows of a table, and the responses are cross-classified to define a response profile that forms the columns of the table. The distribution across the response profile is estimated for each subpopulation. Models are then fit to functions of the response profile distribution. The model-fitting methodology is identical to that described for ANOVA of continuous dependent variables.

As shown in this brief overview, a wide variety of statistical techniques are available for analyzing relationships among survey measures and incorporating the sample design into the analysis. The discussion is by no means exhaustive. For example, logistic regression methods have been adapted to handle the complexities of sample survey data both through Taylor series (LaVange, Iannacchione, and Garfinkel, 1986) and replicated variance estimation procedures (Mohadjer et al., 1986). Adaptation of other analytic procedures to handle complex sample design features is also likely to occur.

Despite the potentially wide range of applications of these methods to NMCUES data, a number of analytic problems in NMCUES cannot be addressed appropriately with the methods summarized here. We now turn to one of these problems, the analysis of relationships between predictor variables and a dependent variable that has a nontrivial proportion of the sample with a limiting value and a skewed distribution for the other nonlimiting values.

Analysis Methods for Limited Dependent Variables

In a number of statistical analyses, the dependent variable has the characteristic of having a lower limiting value (often zero) for a sizable portion of the population, with the remaining members of the population having a distribution of values that is often highly skewed. For example, the charges for medical care incurred by an individual during a 1-year period has this type of distribution. A substantial portion of the population has a zero value either because persons received no medical care during the year or because they received care without charge. The remainder of the population has charges, and the distribution is skewed by a small number of persons incurring very large charges.

A model characterizing the relationship between the limited dependent variable and other measures needs to account for the limited nature of the measure (e.g., limited to medical care users) and the skewed distribution of the nonlimiting values. One approach would be a multipart model with a transformation of the nonlimiting portion of the data (Duan et al. 1982). One part of the model characterizes the relationships among predictors and the proportion of persons with the limiting value; the second part relates the transformed nonlimiting values to either the same or another set of predictors. The multipart model is less satisfying than a single comprehensive model because it uses a mixture of proportions and transformed observations.

A second approach is the limited dependent variable methods proposed by Tobin (1958) and Heckman (1976), which are prominent in the econometric literature. The nonlimiting values of the dependent variable (charges) can be transformed to reduce or eliminate skewness.

A third approach is to consider the charge measure in terms of an ordinal variable with meaningfully selected cutoff points, one of which can be assigned to the limiting value. Modeling methods for ordinally scaled measures can then be applied to the categorized distribution. For example, McCullagh (1980) discusses the application of proportional odds and proportional hazard models to the analysis of ordinally scaled response variables. Williams and Grizzle (1972) and Agresti (1985) propose forming logits from cumulative proportions on an ordinally scaled measure and examining the fit of various models to the cumulative logits.

Unfortunately, existing applications of these approaches treat the observations as independent and identically distributed random variables. In applications of these methods to data obtained from sample surveys with complex sample designs, this assumption is invalid. In this section, two methods for handling limited dependent variables are presented. First, a two-step approach using logistic regression methods is reviewed. In the first step, a logistic regression model is fit to the data to predict the probability that an individual will use a particular type of health care during 1980. In the second step, logistic regression is also used to develop a model to predict the probability that a user of a particular type of health care will have a charge or use level above a fixed threshold. In each logistic regression analysis, the NMCUES sampling weights are used in the computation of estimated coefficients. The sample design is accounted for in standard error estimates through an ad hoc procedure. The second method for handling limited dependent variables is a modification of the ordinal scaling approach. This modification incorporates the sample design into an analysis of cumulative logits through the weighted least-squares (WLS) methodology of Grizzle, Starmer, and Koch (1969). Following Koch, Freeman, and Freeman (1975), Freeman et al. (1976), and Landis et al. (1982), variational models for the observed cumulative logits from complex sample survey data are fit using the WLS estimation and hypothesis testing framework.

Discussion of the logistic regression approach is limited to the basic features of the logistic regression model and interpretation of its results. The two-step procedure itself and its interpretation are described in some detail in Berki et al. (1985) and will not be repeated here. Application of the cumulative logit methodology to limited dependent variables is described in more detail here. The methodology is illustrated with an analysis of data from NMCUES.

Logistic Regression Methods

In Berki et al. (1985) logistic regression models are used to identify characteristics of individuals in the U.S. civilian noninstitutionalized population in 1980 that are predictive of use of hospital services and high use

of hospital services. In the standard regression situation, the dependent variable Y is intervally scaled (continuous in distribution), and, as described previously, the model predicts the mean of Y as a linear combination of the predictors X_k. The logistic regression model is applied when the dependent variable is not continuous but dichotomous (i.e., Y assumes only one of two possible values). Use of standard regression methods in this case is inappropriate for a variety of reasons. Most importantly, several assumptions of the regression model are violated, and predicted values for a dichotomous Y from the ordinary regression model can fall outside the range of allowable values. Logistic regression avoids these deficiencies by modeling the logit of the probability that the dependent variable will assume one of the two possible values.

Specifically, suppose that Y can assume only two values, 1 or 2, and that $Y = 1$ is the event of interest. The logit is defined as the logarithm of the ratio of the probability that $Y = 1$ to the probability that $Y = 2$, denoted by

$$L(Y) = log_e[Prob(Y = 1) / Prob(Y = 2)].$$

Here $log_e[\cdot]$ denotes the natural logarithm of the argument $[\cdot]$, and $Prob(\cdot)$ is the probability of event $(\cdot)$. Thus, the multiple logistic regression model,

$$L(y_i) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_p x_{pi} + e_i,$$

formulates the logit of Y as a linear combination of the predictors x_{ki}. The β_k's again are partial regression coefficients, but in this case β_k denotes the partial linear regression of the logit of Y on the kth predictor, given that the other predictors are also in the model. An iterative maximum likelihood procedure is most commonly used to estimate the coefficients.

An alternative interpretation of these logistic regression coefficients can be made by observing that $L(Y)$ is actually the logarithm of the odds that an individual will be classified as $Y = 1$ rather than $Y = 2$. For example, suppose that $Y = 1$ corresponds to the event that an individual was hospitalized in 1980, and $Y = 2$ corresponds to the event that the individual was not hospitalized. Then $Prob(Y = 1) / Prob(Y = 2)$ denotes the odds that an individual was hospitalized during the year, and the logit is simply the logarithm of this odds, referred to as the log odds.

The logistic regression coefficients can be interpreted in terms of the log odds as well as in terms of the ratio of odds for two different individuals with different values for the predictors. For example, consider the following three indicator variables:

$$X_1 = \begin{cases} 1, \text{ if the individual is under 35 years of age,} \\ 0, \text{ otherwise.} \end{cases}$$

$$X_2 = \begin{cases} 1, \text{ if the individual is 55–74 years of age,} \\ 0, \text{ otherwise.} \end{cases}$$

$$X_3 = \begin{cases} 1, \text{ if the individual is 75 years of age and over,} \\ 0, \text{ otherwise.} \end{cases}$$

Through the combined use of these three indicator variables, individuals aged 35–54 years are assigned to a comparison, or reference, group against which the odds of being hospitalized are contrasted. The logistic regression coefficients represent the unit change in the log odds that occurs for an individual in one of the three age groups defined by these indicators relative to the age group 35–54 years.

For example, consider the estimated logistic regression coefficients from a predictive model for use of hospital care, displayed in Table 18. Here the coefficient corresponding to X_1 is 0.1510. Thus, persons under age 35 have somewhat higher odds of being hospitalized than persons aged 35–54 (the reference group) when all the other variables in the model are controlled at an average value. Similarly, persons aged 55–74 and 75 and over have higher odds of being hospitalized than persons aged 35–54 years have.

These coefficients can be used to predict the probability of hospitalization for a hypothetical person. The estimation of predicted probabilities requires three steps:

(1) Identify characteristics of the hypothetical person in terms of characteristics in the model.

(2) Identify coefficients that apply to that individual and combine them to obtain a predicted logarithm of the odds of the event occurring.

(3) Use the predicted log odds in step (2) to obtain the predicted probability.

For example, suppose one wants to estimate the probability of hospitalization in 1980 for a hypothetical individual who is male, white, and 35–54 years of age; reported good health status and income that is 200–499 percent of the poverty level; has private health insurance and a usual source of health care; and resides in the Northeast Region. The predicted log odds that this hypothetical individual is hospitalized, relative to not being hospitalized, is estimated by summing the appropriate coefficients from Table 18.

For characteristics that have two levels, such as sex, either a coefficient will be included in the sum or nothing will be added. For instance, with males as the reference group, the hypothetical person would not have a coefficient for sex added to the log odds sum. On the other hand, because the reference group for race is black, this hypothetical white individual would have the coefficient for race added to the log odds sum.

For characteristics with several levels, the coefficient for persons in the reference group is zero, and the coefficients for the others are nonzero. Thus, the health status coefficient for the hypothetical individual reporting good health is zero. On the other hand, for type of health care coverage, for which the reference group is none or some mixture of part-year coverages, the private insurance coefficient would be added to the sum.

Coefficients for variables parametrized with the standard analysis-of-variance approach, in which the parameter values sum to zero, must be handled in a different way. For example, in the model reported in Table 18, the sum-to-zero parametrization is used for poverty level. As a result, those with income more than 700 percent of the poverty level have a coefficient determined by adding the negative value of the sum of the other poverty level coefficients to the logarithm of the odds. For the hypothetical person with income 200–499 percent of the poverty level, the contribution of poverty level to the log odds sum would simply be the value of the coefficient for that group, -0.0728. If the hypothetical person had an income that was more than 700 percent of the poverty level, the contribution of poverty level would be the negative of the sum of the coefficients for the other poverty level groups, or $-0.1483 + 0.0728 + 0.1191 = 0.0436$.

Once the coefficients of all separate characteristics are determined and combined, the estimated log odds is calculated by adding the constant term, which corresponds to the effect for a person who has all the reference group characteristics, to the sum of coefficients. Thus, the predicted log odds for the hypothetical individual is computed using coefficients from Table 18 as the sum of the constant term and the coefficients for the white and other race group and for the groups with income 200–499 percent of the poverty level, with private insurance coverage, and with a usual source of care:

$$(-2.8445) + (0.0984) + (-0.0728) + (0.3468) + (-0.0608) = -2.5329.$$

Next, the natural exponent of the predicted log odds is used to obtain the predicted odds of being hospitalized relative to not being hospitalized,

$$[\hat{p}/(1 - \hat{p})] = \exp(-2.5329) = 0.0794.$$

Finally, the predicted odds are converted to a predicted probability by computing

$$\hat{p} = \frac{0.0794}{1 + 0.0794} = 0.0736.$$

Thus, the hypothetical individual has a predicted probability of hospitalization in 1980 of 7.4 percent.

Returning to the three indicator variables for age, the logistic regression coefficients can also be interpreted in terms of a ratio of odds. The function exp(0.1510) is the ratio of the odds of hospitalization for an individual under age 35 to the corresponding odds for an individual aged 35–54, given that the other predictors in the model are fixed at some average value. As noted in Table 18, this odds ratio is 1.1630, indicating that persons under age 35 have approximately 16-percent higher odds of hospitalization than persons aged 35–54 have.

There are several logistic regression programs that can be used to estimate the coefficients of the model and to incorporate the sampling weights in the computation. For example, the estimates in Table 18 were computed using the DREG (Dichotomous REGression) program for logistic regression analysis in the OSIRIS system (Computer Support Group, 1982). The program uses an iterative maximum likelihood procedure to estimate the logistic regression coefficients and incorporates sampling weights directly into the estimates.

The program also estimates standard errors for the logistic regression coefficients, but these estimated standard errors are computed under the assumptions of simple random selection. Standard error estimates for the coefficients that incorporate the sample design into the estimation procedure are available in more recently developed programs (LaVange, Iannacchione, and Garfinkel, 1986; Mohadjer et al., 1986). Alternatively, an ad hoc procedure can be used to adjust the estimated standard errors derived from the DREG program to account for the NMCUES sample design. A model identical to the logistic regression model was estimated using standard regression methods for complex sample designs in which estimated standard errors are computed by balanced repeated replication methods (Frankel, 1974), the REPERR program in the OSIRIS IV system. For each standard regression coefficient, the program calculated the ratio of the actual standard error to the corresponding standard error computed under the assumptions of independent sample selection. Assuming that this ratio of standard errors applies to the corresponding logistic regression coefficients, the estimated standard errors from the DREG program were multiplied by the ratio from the standard regression method to obtain an adjusted standard error for the logistic regression coefficient.

These adjusted standard errors were used to create confidence intervals for the estimated odds ratios computed from the logistic regression coefficients. For each coefficient, a 95-percent confidence interval was computed by adding to and subtracting from the estimated coefficient 1.96 times the adjusted standard error. The upper and lower confidence limits for the logistic regression coefficients, β_U and β_L, were then converted to upper and lower confidence limits for the odds ratio by the transformations e^{β_U} and e^{β_L}. For example, in Table 18, the estimated odds ratio for persons under age 35 (relative to persons aged 35–54) is 1.1630, with 95-percent confidence limits ranging from 1.0321 to 1.3105. Thus, with 95-percent confidence, one can conclude that the odds of being hospitalized for an individual under age 35 are greater than those for an individual aged 35–54.

In standard regression analysis, a useful measure for assessing the goodness of fit of the model to the data is the proportion of variance explained by the model, referred to as the multiple correlation coefficient, or R^2. A parallel measure can be developed for the fit of a multiple logistic regression model. For each individual in the sample, the probability of hospitalization, for example, can be estimated from the logistic regression coefficients by substituting the individual's values for

each predictor into the model and computing the predicted log odds of hospitalization for that individual. The predicted probability of hospitalization under the logistic regression model for each individual who actually was hospitalized can be computed from the predicted log odds. Similarly, the probability of nonhospitalization for individuals in the sample who were not hospitalized can be computed as the complement of their predicted probability of hospitalization. If these predicted probabilities are close to one for both the hospitalized and nonhospitalized persons, the predictive power of the model is good. Probabilities close to zero suggest that the predictive power of the model is poor.

As an overall assessment of the predictive power of the model, a mean of the predicted probabilities for all sample individuals can be computed. In this case, a simple average is not appropriate because the logistic regression model operates on a logarithmic scale rather than the linear scale used in the standard regression model. A more appropriate averaging procedure is a geometric mean of predicted probabilities. Specifically, let y_i denote the observed value (1 or 0, depending on whether the person was hospitalized), and let $\hat{p}_i = Prob(Y_i = y_i)$ denote the predicted probability that the observed event will occur for the ith sample individual based on the logistic regression model coefficients. The predictive power of the model over all individuals can be computed as the geometric mean of these probabilities,

$$\hat{\pi} = \left[\prod_i \hat{p}_i\right]^{1/n},$$

where n denotes the total sample size. If the observed values for all individuals are predicted well from the model, the predicted probabilities $\hat{p}_i$ will each be close to one, and the geometric mean $\hat{\pi}$ will be close to one. Because $\hat{\pi}$ is a measure of how well the model fits the data, $\hat{\pi}_e = 1 - \hat{\pi}$ can be used as a measure of the error associated with the model.

Without any of the predictors $X_1, X_2, \ldots, X_p$, the logit of Y would be predicted using only the intercept. The importance of the X_k values in prediction of the logit of Y can be assessed by examining how much of the predictive error for the "mean only" model is accounted for when the predictors are added to the model. The proportion of predictive error $\hat{\pi}_e$ accounted for by the addition of these predictors is a measure corresponding to the multiple correlation coefficient R^2 in standard regression analysis. Berki et al. (1985) use the proportion of predictive error to examine alternative models and assess whether the predictors add predictive power to a logistic regression model.

Cumulative Logit Analysis Methodology

Consider again a response variable for a characteristic with a lower limiting value, quite possibly zero, and a skewed distribution to the right for the nonlimiting values. As previously, assume that a sizable portion of the population has the limiting value and that development of a model that summarizes the relationship between the distribution of this dependent variable and various predictor variables or factors is desired.

For example, suppose that the dependent variable of interest is physician visit charges, and the relationship between the distribution of physician visit charges and health care coverage is to be examined. For purposes of analysis, health care coverage is divided into four categories that define subgroups of the population:

- None—No health coverage at any time during the reference year (1980).
- Part-year coverage—Public or private insurance during a portion of the reference year or a mixture of government and private insurance throughout the year.
- Public coverage—Health care coverage by any of several U.S. Government health care coverage programs during the entire reference year.
- Private insurance—Health care coverage by a private company during the entire reference year.

Suppose also that the analysis is restricted to adults aged 18–64 years.

Figure 5 presents the cumulative distributions of physician visit charges for the four health care coverage groups for persons aged 18–64. A sizable proportion of each health care coverage group had no physician visit charge in 1980. The proportion with no charge appears lowest for public and private coverage groups and appears highest for the group with no health care coverage. Once the differences in the proportion with no charge are accounted for, it is not clear whether the distributions of charges are similar across the four groups.

Several features of these cumulative frequency distributions need to be examined further:

- Are the proportions with no charge significantly different across the four groups?
- If the proportions with no charge are different, how do they differ?
- Given initial differences in the proportions with no charge, do the groups have the same distributions among those with charges?
- If not, how do the distributions differ? Is one group experiencing higher charges than the rest throughout the distribution or for only a portion of it?

Methods for the analysis of ordinally scaled categorical data can be adapted to answer questions such as these. Using these methods, it is more natural to partition the charge distribution into distinct levels defined by a small number of cutoff points, or thresholds, than to examine the cumulative frequency distributions directly. The estimated cumulative proportions of the popu-

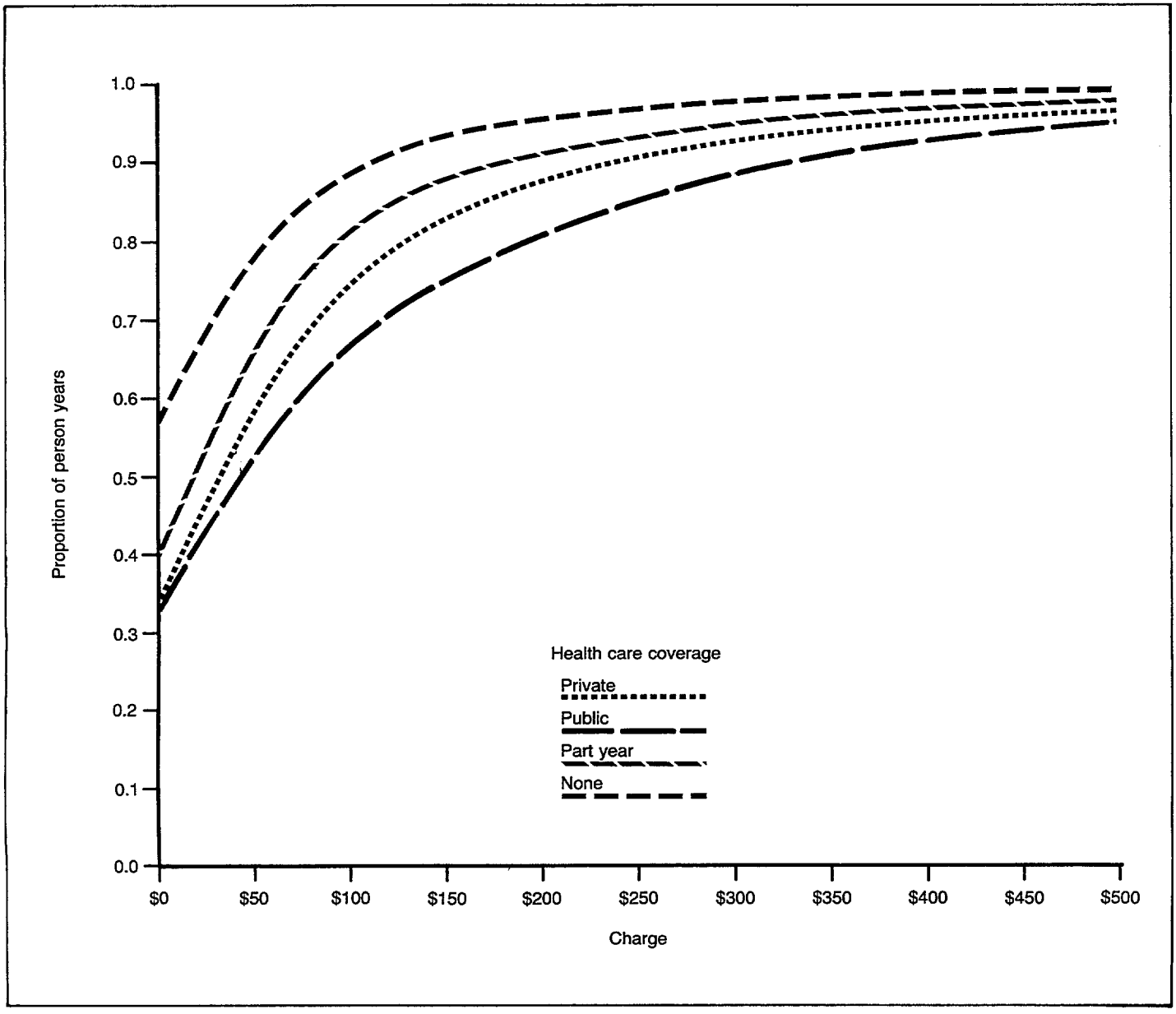


Figure 5

Cumulative frequency distribution of physician visit charges for adults 18–64 years of age by health care coverage: United States, 1980

lation at each threshold value can then be compared across the subgroups. The cumulative proportions themselves could be compared, but a more useful type of comparison will be in terms of whether the cumulative distributions change proportionately or disproportionately from one threshold to the next. Comparisons of proportionate change are more suitably made on a logarithmic scale than on an additive one, so a logarithmic function of cumulative proportions at each threshold is compared across the subgroups of interest.

In the following pages the methodology is described for data obtained from a simple random sample of observations. The description begins with estimation procedures for the cumulative proportions and logits of the cumulative proportions (referred to as cumulative logits), continues with linear models that can be fit to the cumulative logits, and concludes with hypothesis testing procedures that provide a framework for examining the importance of the model parameters. The methodology is then adapted to data arising from complex sample designs such as NMCUES. This is accomplished by estimating the cumulative proportions and their standard errors, taking the sample design into account in the estimation process. Finally, the relationship between physician visit charges and health care coverage is investigated by applying the cumulative logit methodology to NMCUES data.

Cumulative Logit Analysis Under Simple Random Sampling

Suppose that t threshold values are chosen to divide the distribution of the response variable into $t + 1$ levels. Let the cumulative frequency distribution be deter-

mined relative to this set of thresholds to form an ordinally scaled variable. The first threshold is fixed at the limiting value of the distribution and thus defines a subgroup of sample units, each of which has the limiting value of the dependent variable. Other thresholds are chosen to represent points of substantive interest on the distribution of the response variable (or some other criteria may be used). In addition, each element of the population is further classified into s subpopulations based on categories of a single predictor variable or on the cells of the cross-classification of a set of predictor variables.

Suppose that a simple random sample of size n is selected from the population of interest and consider the cross-classification of observations by subpopulations and thresholds shown in Table 19. Here, $n_{i\cdot}$ denotes the number of observations in the ith subpopulation; n_{ij} denotes the number of sample elements classified into the ith subpopulation and having values of the dependent variable that are greater than the $(j - 1)$th and less than or equal to the jth threshold values for $j = 1, \ldots, t$; and $n_{i,t+1}$ denotes the number of sample elements in the ith subpopulation with values of the dependent variable greater than the tth threshold. Let $p_{ij} = n_{ij} / n_{i\cdot}$ be the observed proportion of the ith subpopulation with values of the dependent variable between the $(j - 1)$th and jth threshold values, with estimated variance $\mathrm{var}(p_{ij}) = p_{ij}(1 - p_{ij}) / n_{i\cdot}$ and estimated covariance with $p_{ij'}$ of $-p_{ij}p_{ij'} / n_{i\cdot}$ and with $p_{i'j'}$ of 0 for $i \neq i'$. The observed cumulative proportion of sample elements in the ith subpopulation with a value of the dependent variable that is less than or equal to the jth threshold is thus $F_{ij} = p_{i1} + \ldots + p_{ij}$, for $j = 1, \ldots, t$, and the logit of the (ij)th cumulative probability is $L_{ij} = \log_e[F_{ij} / (1 - F_{ij})]$. The variances and covariances among the L_{ij} can be obtained directly through appropriate transformations of the p_{ij} and use of Taylor series approximations, as described subsequently.

Now consider a saturated linear model that characterizes the variation in the cumulative logits as

$$L_{ij} = \mu + \tau_i + \alpha_j + \gamma_{j(i)},$$

where μ denotes a mean of subpopulation cumulative logits at the first threshold. The τ_i denote the effect of the ith subpopulation on the cumulative logit. The τ_i can be parametrized in a variety of ways, including the familiar sum-to-zero ANOVA parametrization, $\Sigma_i \tau_i = 0$. Alternatively, a reference group parametrization can be used in which $\tau_{i'} = 1$, for $i' = i$, and $= 0$, otherwise, for $i = 1, 2, \ldots, s - 1$, where the reference group is the sth subpopulation. The α_j values may be termed threshold effects, and they also can be parametrized in a variety of ways.

For this model, it is instructive to parametrize the threshold effects relative to the level of the first cumulative logit, L_{i1}. For example, if $\alpha_1 = 0$ and $\alpha_j = \sum_{j'<j} \beta_{j'}$, it follows that β_j corresponds to an incremental slope parameter for the jth cumulative logit relative to the $(j - 1)$th cumulative logit. In this way, the parameters are directly analogous to those in standard profile analysis for repeated measures, with the levels of the repeated measure being the cumulative logits corresponding to the successive thresholds in the model. Finally, the $\gamma_{j(i)}$ denote a departure from the mean slope of the $(j - 1)$th to the jth cumulative logit line segment for the ith subpopulation. The $\gamma_{j(i)}$ can also be parametrized in several ways, but usually the parametrization is made in a manner consistent with that for the τ_i. Thus, if a sum-to-zero parametrization is used for the τ_i, a parallel parametrization is used for the $\gamma_{j(i)}$.

If the $\gamma_{j(i)}$ are equal to zero for $i = 1, 2, \ldots, s$, the line segments connecting the $(j - 1)$th to the jth cumulative logit are parallel across the s subpopulations. Moreover, if the $\gamma_{j(i)}$ are equal to zero for $j = 1, 2, \ldots, t$ and $i = 1, 2, \ldots, s$, then the resulting additive model $L_{ij} = \mu + \tau_i + \alpha_j$ corresponds to the proportional odds model described by McCullagh (1980). Agresti (1985) refers to the parallel profile model in which the predictors are simply nominally scaled as a logit row effects model; when the predictors are ordinally scaled and assigned integer scores, Agresti terms it the logit uniform association model. Finally, if in the profile analysis parametrization of the slope terms, the β_j are equal to one another (i.e., $\beta_j = \beta$ for $j = 2, \ldots, t$) and $\gamma_{j(i)} = 0$, then the odds ratios across successive thresholds are identical, and the successive line segments for a given subpopulation have identical slopes.

The cumulative frequency distributions can be compared across the subpopulations in terms of hypotheses corresponding to those in profile analysis:

- Are the line segments between two adjacent thresholds parallel across subpopulations? (Are the $\gamma_{j(i)}$ equal to zero?)
- Given parallelism, are there differences in "intercepts" across the subpopulations? (Are the τ_i equal to zero?)

These hypotheses can be interpreted in terms of the subpopulation cumulative frequency distributions themselves. If the cumulative logit line segments are parallel, the cumulative frequency distributions have the same proportionate increase in cumulative frequency from one threshold to the next across the subpopulations. Alternatively, parallelism implies that the relative frequency distributions have similar dispersions and skewness once the proportion of each subpopulation with the limiting value is accounted for in the variational model. If there is no difference among the intercepts across subpopulations, the cumulative distributions have the same proportion with a "zero" value for the response variable.

A weighted least squares methodology can be applied to fit the basic model to the data and to generate Wald statistics (Wald, 1943) for hypotheses about the model parameters. Consider an $st \times 1$ vector $\mathbf{L}$ of the observed

cumulative logits obtained from the vector of proportions $\mathbf{p} = [p_{11}, \ldots, p_{1,t+1}, \ldots, p_{s1}, \ldots, p_{s,t+1}]$ through the compounded transformations

$$\mathbf{L} = A_2[log_e(A_1\mathbf{p})].$$

Here, A_1 is the linear operator matrix used to form the vector of cumulative proportions and their complements, $\mathbf{F} = [F_{11}, 1 - F_{11}, \ldots, F_{1t}, 1 - F_{1t}, \ldots, F_{s1}, 1 - F_{s1}, \ldots, F_{st}, 1 - F_{st}]$; $log_e(\cdot)$ takes the natural logarithm of every element in the matrix argument $(\cdot)$; and $A_2 = [1, -1] \otimes I_{s \times t}$, where $\otimes$ denotes the left-hand direct product. Let $V_{\mathbf{L}}$ denote the estimated covariance matrix of dimension $st \times st$ for the elements of $\mathbf{L}$, which can be obtained as the direct matrix product

$$V_{\mathbf{L}} = A_2 D_{\mathbf{F}}^{-1} A_1 V_{\mathbf{p}} A_1' D_{\mathbf{F}}^{-1} A_2',$$

where $D_{\mathbf{F}}$ is a diagonal matrix with elements of $\mathbf{F}$ along the main diagonal and $V_{\mathbf{p}}$ is the standard maximum likelihood estimator for the covariance structure of the observed proportions under a product multinomial distribution.

Following Grizzle, Starmer, and Koch (1969), the linear model

$$E_A\{\mathbf{L}\} = X\mathbf{B}$$

is fitted to the vector of cumulative logits $\mathbf{L}$, where $E_A\{\cdot\}$ denotes the asymptotic expected value of the argument $\{\cdot\}$, X denotes an $st \times g$ model matrix of constants, and $\mathbf{B}$ denotes a $g \times 1$ vector of parameters. Under the hypothesis that $E_A\{\mathbf{L}\} = X\mathbf{B}$, the WLS estimator

$$\mathbf{b} = (X' V_{\mathbf{L}}^{-1} X)^{-1} X' V_{\mathbf{L}}^{-1} \mathbf{L}$$

minimizes the residual error sum of squares,

$$Q = (\mathbf{L} - X\mathbf{b})' V_{\mathbf{L}}^{-1} (\mathbf{L} - X\mathbf{b}).$$

Given that $E_A\{\mathbf{L}\} = X\mathbf{B}$ holds, Q asymptotically follows a χ^2 distribution with $st - g$ degrees of freedom. Thus, Q can be used to test the goodness of fit of the linear model $X\mathbf{B}$ to the vector of cumulative logits.

If the model fits the data, hypotheses of the form

$$H_0: C\mathbf{B} = 0,$$

where the contrast matrix C is a $(d \times g)$ matrix of constants, can be tested using the test statistic

$$Q_C = (C\mathbf{b})'[C(X'V_{\mathbf{L}}^{-1}X)^{-1}C']^{-1}C\mathbf{b}.$$

Under H_0, the test statistic Q_C asymptotically follows a χ^2 distribution with d degrees of freedom. Hypothesis tests of the form $C\mathbf{B} = 0$ suggest that a reduced model with fewer parameters will explain the variation in the observed cumulative logits adequately. The vector of reduced parameters can be estimated using the WLS estimation procedure and predicted values for the cumulative logits obtained as $\hat{\mathbf{L}} = X_R\mathbf{b}_R$, where X_R is the reduced model matrix and $\mathbf{b}_R$ is the estimated reduced parameter vector.

These WLS methods can be implemented using software such as GENCAT (Landis et al., 1976) or the CATMOD procedure within SAS (SAS Institute, Inc., 1985). The nature of the model matrix X, the corresponding parameter vector $\mathbf{B}$, and the contrast matrix C will be illustrated subsequently using the NMCUES physician visit charge data.

Cumulative Logit Analysis Under Complex Sampling

The preceding presentation was based on the assumption of simple random selection. In order to apply the cumulative logit methodology to NMCUES data, the estimation procedures used to obtain the vector of proportion $\mathbf{p}$ and its variance-covariance matrix $V_{\mathbf{p}}$ must take account of the complexity of the stratified multistage probability sample design and weighted observations. These estimation procedures have already been discussed in the section on estimation but are repeated here for the sake of completeness.

In NMCUES, sampling error computing unit codes are available that identify a set of $a_h = 2$ sampling error computing units (SECU's) from each of the $h = 1, \ldots, 69$ sampling error strata. Let n_{ha} denote the number of sample elements within the (ha)th SECU. As previously, w_{hak} is the weight and t_{hak} is the time-adjustment factor assigned to the kth sample element within the (ha)th SECU. Recall that the weight w_{hak} accounts for unequal probabilities of selection and serves as an adjustment for nonresponse and noncoverage.

Consider the indicator variable

$$y_{ijhak} = \begin{cases} w_{hak}t_{hak}, & \text{if the } (hak)\text{th sample element is in the } i\text{th subpopulation and } j\text{th threshold group,} \\ 0, & \text{otherwise.} \end{cases}$$

Weighted estimates of the sample proportions p_{ij} can be obtained as

$$\begin{aligned} p_{ij} &= \Sigma_h \Sigma_a \Sigma_k y_{ijhak} / \Sigma_h \Sigma_a \Sigma_k \Sigma_j y_{ijhak} \\ &= n_{ij} / n_{i\cdot} . \end{aligned}$$

As such, each of the estimated proportions is a ratio of two random variables because the denominator, $n_{i\cdot}$, is not a fixed quantity in the design but a random variable. Estimation of the variances of the p_{ij} and the covariances among them is typically accomplished through the use of Taylor series approximations. For example, the variance of p_{ij} is estimated as

$$\mathrm{var}(p_{ij}) \cong (n_{i\cdot})^{-2}\left[\mathrm{var}(n_{ij}) + (p_{ij})^2\,\mathrm{var}(n_{i\cdot}) - 2p_{ij}\,\mathrm{cov}(n_{ij}, n_{i\cdot})\right],$$

where $\mathrm{var}(n_{ij})$, $\mathrm{var}(n_{i\cdot})$, and $\mathrm{cov}(n_{ij}, n_{i\cdot})$ are the respective estimated variances and covariance. (See, for example, Kish, 1965.)

The variances and covariance of n_{ij} and $n_{i\cdot}$ in $\mathrm{var}(p_{ij})$ are estimated by taking the stratified multistage sample design into account. For NMCUES, in which $a_h = 2$ for all 69 strata (i.e., a paired selection of primary sampling units within strata), these variances and covariance can be estimated as

$$\mathrm{var}(n_{ij}) = \Sigma_h (n_{ijh1} - n_{ijh2})^2,$$
$$\mathrm{var}(n_{i\cdot}) = \Sigma_h (n_{i\cdot h1} - n_{i\cdot h2})^2, \text{ and}$$
$$\mathrm{cov}(n_{ij}, n_{i\cdot}) = \Sigma_h (n_{ijh1} - n_{ijh2})(n_{i\cdot h1} - n_{i\cdot h2}),$$

where $n_{ijha} = \Sigma_k y_{ijhak}$ and $n_{i\cdot ha} = \Sigma_j n_{ijha}$, for $a = 1, 2$.

In a similar manner, the covariances between any two sample proportions p_{ij} and $p_{i'j'}$ can be estimated using a Taylor series approximation:

$$\mathrm{cov}(p_{ij}, p_{i'j'}) = [\mathrm{cov}(n_{ij}, n_{i'j'}) + (p_{ij}p_{i'j'})\,\mathrm{cov}(n_{i\cdot}, n_{i'\cdot}) - p_{ij}\,\mathrm{cov}(n_{i'j'}, n_{i\cdot}) - p_{i'j'}\,\mathrm{cov}(n_{ij}, n_{i'\cdot})] \,/\, (n_{i\cdot}\, n_{i'\cdot}).$$

The covariance terms in $\mathrm{cov}(p_{ij}, p_{i'j'})$ are computed in a manner similar to that indicated for $\mathrm{cov}(n_{ij}, n_{i\cdot})$. Unlike the case in a simple random sampling design, the covariances between proportions from two different subpopulations in a complex sample design are not equal to zero.

These variances and covariances of the NMCUES sample proportions can be estimated using specialized statistical software such as the PSTABLE program within OSIRIS IV (Computer Support Group, 1982), the SESUDAAN package of programs that operate under the SAS system (Shah, 1984), or SUPERCARP and PC CARP (Hidiroglou, Fuller, and Hickman, 1974; Schnell et al., 1986). Thus, the vector **p** and the corresponding covariance matrix $V_{\mathbf{p}}$ can be estimated directly from survey data for a specified set of subpopulations and thresholds. This vector and covariance matrix can then be substituted directly into the cumulative logit methodology. At present, only the GENCAT program (Landis et al., 1976) allows direct input of the vector **p** and the matrix $V_{\mathbf{p}}$ into a WLS analysis. Thus, implementation of the cumulative logit methodology for the complex sample design case requires access to GENCAT.

Physician Visit Charge Illustration

Consider now the problem of developing a variational model to explain differences among the observed cumulative distributions of physician visit charges across subgroups of the population defined by health care coverage groups. Two issues must be addressed in the selection of thresholds on the distribution of physician visit charges: How many thresholds to select and how to choose the values of the thresholds. The number of thresholds to select depends on such features as the size of the group that has the limiting value (i.e., no physician visit charges), the sample size, and the nature of the distribution among the nonlimiting values. Considerations of the size of the limiting value group and the sample size are often related. For example, in the case of a large proportion with the limiting value and a small sample, only a few thresholds may be possible to provide adequate sample size for reliable estimation. On the other hand, in the case of a small proportion with the limiting value and a large sample, a larger number of thresholds can be used. However, if more thresholds are chosen, more parameters are required in the model and the subsequent model development becomes more difficult. Finally, variation in distributions that are unimodal and symmetric in shape can be more adequately explained with fewer thresholds (and parameters) than distributions with other properties. For many types of measures for which cumulative logit analysis is suitable, from three to six thresholds (four to seven categories of response) are adequate.

Suppose we select a set of $t = 4$ thresholds of physician visit charges (Figure 6) yielding an ordinal response variable with $t + 1 = 5$ levels. For obvious reasons, the first threshold corresponds to the group of persons with no charge in 1980. The remaining thresholds can be chosen by a variety of strategies, three of which are briefly described.

Substantive considerations may immediately suggest appropriate threshold values. Even then, from a statistical perspective, the empirical distribution of physician visit charges for each subgroup must be examined to determine whether the substantive thresholds provide adequate sample sizes in each threshold group of each subpopulation. If sample sizes are not adequate, the threshold boundaries may be moved or a smaller number of thresholds may be required.

A second, more empirically based, approach is to divide the distribution of nonlimiting values into equal-sized groups. With $t = 4$ thresholds, the first of which is the limiting value, quartiles on the remainder of the distribution are determined and corresponding thresholds identified. For instance, in Figure 6 approximately 35 percent of the population have the limiting value. Quartiles of the remaining 65 percent of the distribution would occur at approximately the 51st, 67th, and 83d percentiles of the distribution. The corresponding thresholds are at approximately $35, $65, and $135, respectively.

A third model-based approach to threshold selection is to choose thresholds such that the odds ratios involving cumulative proportions for successive pairs of thresholds are identical. In particular, thresholds are chosen so that a constant difference in successive cumulative logits,

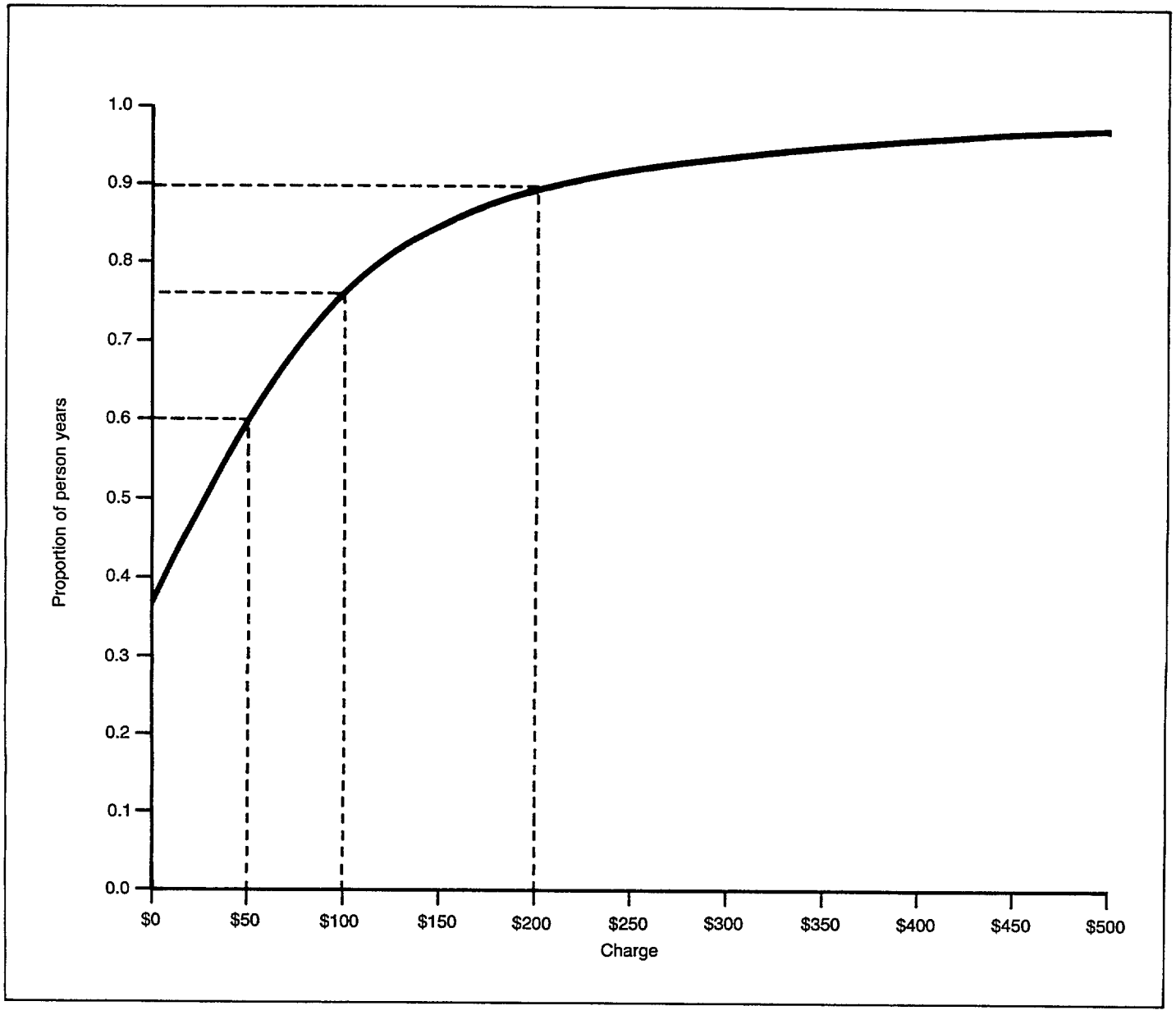


Figure 6

Cumulative frequency distribution of physician visit charges for adults 18–64 years of age, with selected threshold values: United States, 1980

$$log_e[F_{j+1}/(1-F_{j+1})] - log_e[F_j/(1-F_j)] = log_e[F_{j+2}/(1-F_{j+2})] - log_e[F_{j+1}/(1-F_{j+1})],$$

is obtained for $j = 1, \ldots, t - 1$. Let this constant difference (or log odds ratio) be denoted by Ω, and let $\phi = exp(\Omega)$ denote the uniform odds ratio. Then the cumulative probabilities that satisfy the constant odds ratio model can be obtained as

$$F_{j+1} = \phi F_j / (1 - (1 - \phi) F_j).$$

For the limited dependent variable problem, F_1 is the proportion of the population that has the limiting value. Thus, for a fixed value of ϕ, the cumulative probabilities for the remaining thresholds can be solved successively. In Figure 6, $F_1 = 0.350$, and suppose $\phi = 3$. Then $F_2 = 3 \cdot F_1/(1 - (1 - 3)F_1) = 0.618$, which corresponds to approximately $50. Similarly, $F_3 = 3 \cdot F_2/(1 - (1 - 3)F_2) = 0.829$ (approximately $125), $F_4 = 0.936$ (approximately $250), and the three cumulative logit line segments will have identical slopes for the overall distribution (i.e., equal values of α_j). Given this choice of thresholds based on uniform odds ratios for the overall distribution, the subpopulation models can be interpreted in terms of their departures from the overall model. The departure will be reflected in the mean slope parameters (the α_j values) and deviations from the mean slope parameters (the $\gamma_{j(i)}$ values).

The first alternative for the selection of thresholds was employed in this example, and the estimated weighted proportions of person years corresponding to the

four selected thresholds in each of four health care coverage groups are shown in Table 20. The proportion of person-years with no physician visit charge is lowest for those with some type of public coverage and is highest for those with no coverage. Similarly, the proportion with charges less than or equal to $200 (the last threshold value) is lowest for those with some type of public health care coverage and highest for those with no coverage. From visual inspection, the proportions in Table 20 suggest that once the initial differences for the "None" threshold are accounted for, the four health care coverage groups have similar distributions for physician visit charges.

In Figure 7, the results of complex sample design variance and covariance estimation for the proportions shown in Table 20 are illustrated. The lower triangle of the upper left 10 × 10 submatrix of the 20 × 20 covariance matrix for the proportions is given. The first 5 × 5 submatrix is composed of the variances (along the diagonal) and covariances (the off-diagonal terms) for the private coverage group. The next 5 × 5 submatrix along the diagonal corresponds to the public coverage group. The 5 × 5 submatrix on the off diagonal is composed of covariances between the two sets of proportions from two different subpopulations.

The corresponding covariance matrix for the simple random sampling case is shown in Figure 8. By definition, the off-diagonal covariances between the proportions from the two different subpopulations are equal to zero. In addition, the variances along the diagonal in Figure 8 are generally smaller than those in Figure 7. The ratio of variances in Figure 7 to those in Figure 8 is the design effect for each of the 10 proportions in the first two subpopulations. The design effect is greater than 1.0 in most cases.

In terms of the cumulative logit analysis, we can now perform model-fitting under both of these two sample selection assumptions. A comparison of the various test statistics obtained under each set of assumptions will provide an indication of the effect that the complex sample design has on subsequent inference relative to the standard simple random sampling approach. If the vector of sample proportions in Table 20 is used with the full variance-covariance matrix associated with

$$\begin{bmatrix}
0.616 & & & & & & & & & \\
-0.190 & 0.388 & & & & & & & & \\
-0.171 & -0.023 & 0.254 & & & & & & & \\
-0.157 & -0.098 & -0.012 & 0.233 & & & & & & \\
-0.098 & -0.076 & -0.048 & 0.034 & 0.188 & & & & & \\
-0.116 & -0.004 & 0.277 & -0.099 & -0.058 & 2.821 & & & & \\
0.042 & -0.023 & -0.072 & 0.026 & 0.027 & -1.000 & 1.488 & & & \\
-0.014 & 0.082 & -0.046 & -0.055 & 0.033 & -0.258 & -0.365 & 0.642 & & \\
0.015 & -0.012 & -0.043 & 0.053 & -0.013 & -0.805 & 0.137 & -0.165 & 0.916 & \\
0.073 & -0.043 & 0.116 & 0.076 & 0.011 & -0.757 & -0.260 & 0.146 & -0.084 & 0.954
\end{bmatrix} \times 10^{-4}$$

Figure 7

Lower triangle of the complex sample design covariance matrix $V_p(\times 10^{-4})$ for the private and public health care coverage groups

$$\begin{bmatrix}
0.339 & & & & & & & & & \\
-0.133 & 0.292 & & & & & & & & \\
-0.076 & -0.059 & 0.193 & & & & & & & \\
-0.067 & -0.052 & -0.030 & 0.172 & & & & & & \\
-0.064 & -0.049 & -0.028 & -0.025 & 0.166 & & & & & \\
0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 1.527 & & & & \\
0.0 & 0.0 & 0.0 & 0.0 & 0.0 & -0.477 & 1.170 & & & \\
0.0 & 0.0 & 0.0 & 0.0 & 0.0 & -0.314 & -0.207 & 0.840 & & \\
0.0 & 0.0 & 0.0 & 0.0 & 0.0 & -0.304 & -0.201 & -0.132 & 0.818 & \\
0.0 & 0.0 & 0.0 & 0.0 & 0.0 & -0.432 & -0.285 & -0.188 & -0.182 & 1.087
\end{bmatrix} \times 10^{-4}$$

Figure 8

Lower triangle of the simple random sample covariance matrix ($\times 10^{-4}$) for the private and public health care coverage groups

Figure 7, the option 3 analysis described in the last section and in Landis et al. (1982) is obtained. Similarly, if the same vector of proportions and the full variance-covariance matrix associated with Figure 8 is used, the option 2 analysis is obtained. All three options were actually used. The results of the option 3 analysis are presented, and then the option 3 test statistics are compared with those obtained under option 2. (The option 1 results are, in this case, quite similar to those obtained under option 2.)

As noted previously, the vector of cumulative logits $\mathbf{L}$ and $V_{\mathbf{L}}$ can be obtained from $\mathbf{p}$ and $V_{\mathbf{p}}$ through the compounded transformations $A_2[log_e(A_1\mathbf{p})]$. In this instance,

$$A_1 = \begin{bmatrix} 10000 \\ 01111 \\ 11000 \\ 00111 \\ 11100 \\ 00011 \\ 11110 \\ 00001 \end{bmatrix} \otimes I_4$$

and $A_2 = [1, -1] \otimes I_{16}$, where I_d is a $d \times d$ matrix with 1's along the diagonal. The values of the cumulative proportions F_{ij} and the cumulative logits L_{ij} obtained through these transformations are given in Table 21.

In Figure 9, the cumulative logits for physician visit

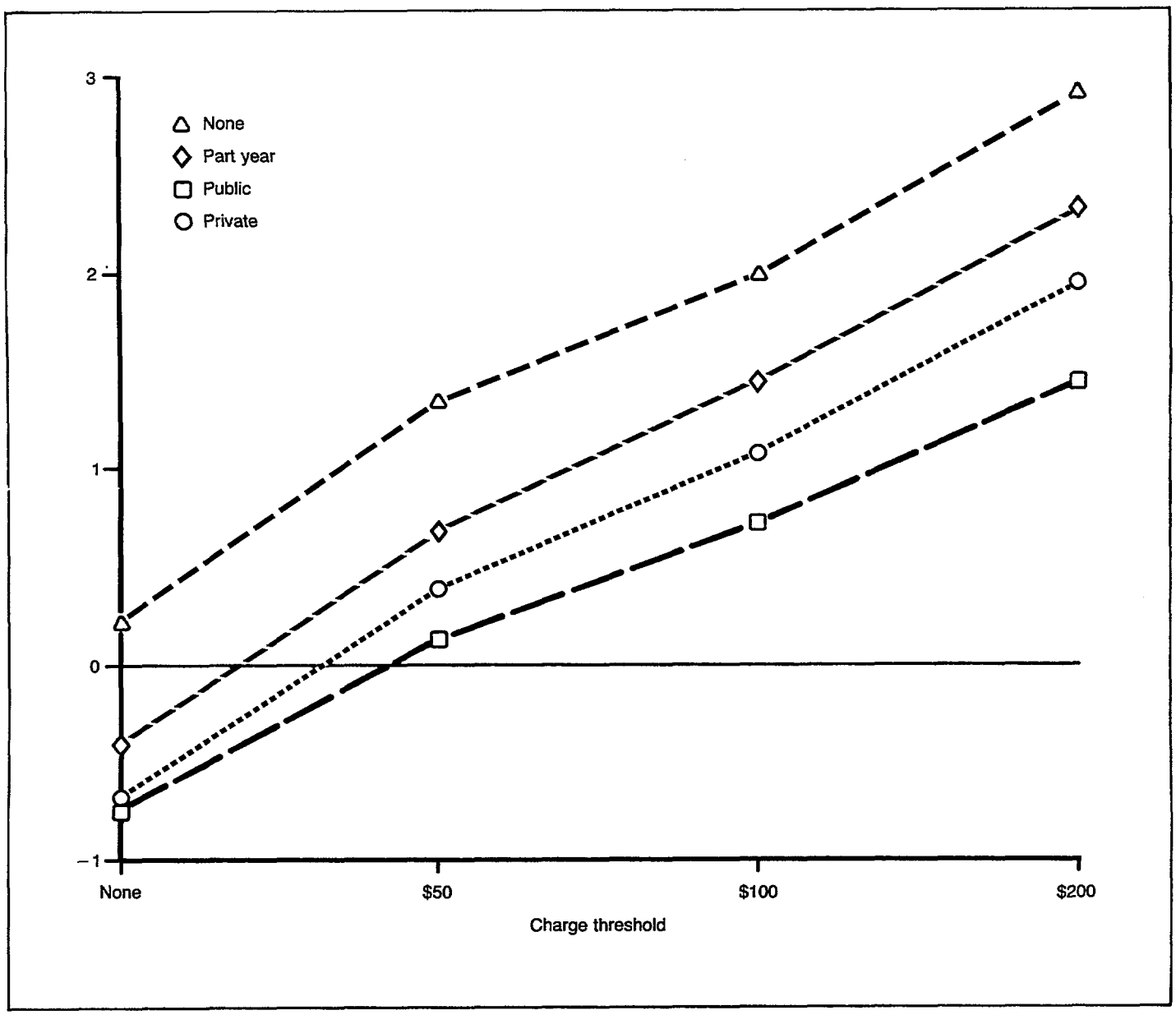


Figure 9

Observed cumulative logits for adults 18–64 years of age, by health care coverage: United States, 1980

charge thresholds across the four health care coverage groups are presented. For each of the three sets of line segments (i.e., from the first to the second threshold, etc.), it appears that the slope for the public health care coverage is smaller than the slope for the other categories. However, the slopes for the other three health care coverage groups appear to be essentially the same. The curves might thus be summarized in terms of a model with different intercepts for each group, equal slopes or increments for each subsequent threshold for three of the coverage groups, and a smaller slope for the public coverage group for each segment. Such a strategy of visual inspection followed by model-fitting may lead to "overfitting" of models and spurious results. Hence it was not adopted here.

In the strategy used, the parameters of a saturated model (a model with as many parameters as cumulative logits) was examined by testing various hypotheses about the intercepts and slopes shown in Figure 9. The saturated model matrix is

$$X_S = \begin{bmatrix} 1\,100\,000\,000\,000\,000 \\ 1\,100\,100\,100\,000\,000 \\ 1\,100\,110\,000\,100\,000 \\ 1\,100\,111\,000\,000\,100 \\ 1\,010\,000\,000\,000\,000 \\ 1\,010\,100\,010\,000\,000 \\ 1\,010\,110\,000\,010\,000 \\ 1\,010\,111\,000\,000\,010 \\ 1\,001\,000\,000\,000\,000 \\ 1\,001\,100\,001\,000\,000 \\ 1\,001\,110\,000\,001\,000 \\ 1\,001\,111\,000\,000\,001 \\ 1\,000\,000\,000\,000\,000 \\ 1\,000\,100\,000\,000\,000 \\ 1\,000\,110\,000\,000\,000 \\ 1\,000\,111\,000\,000\,000 \end{bmatrix}$$

with parameter vector

$$\mathbf{B_S} = \begin{bmatrix} \text{Mean logit for no charges} \\ \text{Health care coverage: Private} \\ \text{Health care coverage: Public} \\ \text{Health care coverage: Part year} \\ \text{Slope 1} \\ \text{Slope 2 increment} \\ \text{Slope 3 increment} \\ \text{Slope 1 differential: Private} \\ \text{Slope 1 differential: Public} \\ \text{Slope 1 differential: Part year} \\ \text{Slope 2 differential: Private} \\ \text{Slope 2 differential: Public} \\ \text{Slope 2 differential: Part year} \\ \text{Slope 3 differential: Private} \\ \text{Slope 3 differential: Public} \\ \text{Slope 3 differential: Part year} \end{bmatrix}$$

The blocks of four rows in X_S correspond to parametrization of the four cumulative logits for each health care coverage group. A reference cell parametrization relative to the subgroup with no health care coverage is used for the intercept effects τ_i and the differential slope effects $\gamma_{j(i)}$. As a result, the first column of 1's corresponds to the cumulative logit at the first threshold for individuals with no health care coverage. Thus, the τ_i and $\gamma_{j(i)}$ can be interpreted as departures from the base level of those with no health care coverage for the remaining three groups. This parametrization allows examination of the effect of alternative forms of health care coverage on physician visit charges relative to having no health insurance coverage during the year.

The fifth, sixth, and seventh columns of X_S correspond to the average slope 1 and the average incremental slopes 2 and 3 for the three line segments between successive thresholds (Figure 9). Each of the three remaining blocks of three columns represents departures for each of the first three coverage groups from the slope for the final group with no health care coverage. The first four columns of the matrix X_S (i.e., the first four parameters of $\mathbf{B_S}$) correspond to the intercepts of the cumulative logit curves at the baseline threshold of no charge. The remaining columns of X_S concern the successive slopes of the cumulative logit curves, represented as "increments" to the preceding threshold cumulative logits, and departures in the slope from the mean for each health care coverage group.

Estimates of the saturated model parameters in $\mathbf{B}$ were obtained by WLS and are shown, together with their estimated standard errors, in Table 22. The three health care coverage coefficients (i.e., intercepts) all appear to be substantial and negative, and it might be concluded that private, public, and part-year coverage groups have lower proportions with no charges than the reference group with no coverage during the year has. The three slope coefficients are all substantial, indicating the obvious monotonically increasing cumulation of cumulative logits from one threshold to the next. The slope differential coefficients associated with public coverage appear to be substantial and negative, and the other slope differential coefficients are small relative to their standard errors. It thus appears that the private and part-year coverage groups have increases in cumulative logits from one threshold to the next that are similar to those for the no coverage group, but the public coverage group has a slower cumulation.

These observations can, of course, be tested more rigorously through the WLS hypothesis testing framework outlined previously. For example, to examine whether the slope differential effects for private and part-year coverage are statistically important, the hypothesis $H_0: C\mathbf{B} = 0$ can be examined where the contrast matrix

$$C = \begin{bmatrix} 0\,000\,000\,100\,000\,000 \\ 0\,000\,000\,000\,000\,100 \\ 0\,000\,000\,010\,000\,000 \\ 0\,000\,000\,000\,000\,010 \\ 0\,000\,000\,001\,000\,000 \\ 0\,000\,000\,000\,000\,001 \end{bmatrix}$$

This hypothesis test will have six degrees of freedom (d.f.) corresponding to the six rows of the contrast matrix C.

The results of several hypothesis tests concerning the saturated model parameters are summarized in Table 23. For comparison purposes, test statistics are reported under two sets of assumptions about the sample selection procedure: (1) the actual complex design and (2) simple random sampling. The data set includes sampling weights to compensate for unequal probabilities of selection and for nonresponse and noncoverage errors. Therefore, the weights were used to compute estimates of proportions for the simple random sampling analysis in order to assure that the complex and simple random sampling analyses were making inferences about the same target population. Analytically, the WLS analyses differ for each sample selection assumption only in the method of computing the covariance matrix. The comparison of results by sample design assumption allows an assessment of the effect of the sample design on inferences about the cumulative distributions. The results obtained under the complex sample design analysis are examined, and then the differences in test statistics obtained under simple random sampling assumptions are observed.

The tests concerning the health care coverage effects for private (Q_C = 170.09, 1 d.f.), public (Q_C = 103.60, 1 d.f.), and part-year (Q_C = 46.69, 1 d.f.) coverage groups indicate that intercepts (i.e., the proportion with no charge) for each of these groups differ from the intercept for the no coverage group. The test for parallelism, which hypothesizes that the variation among the four curves can be explained by nonzero but identical slopes for each group (i.e., all the slope differential effects simultaneously equal to zero), has a statistically significant test statistic (Q_C = 34.89, 9 d.f.). Hence, the four curves are not parallel. The last sets of hypothesis tests examine the individual differential slope effects $\gamma_{j(i)}$ to determine the source of the lack of parallelism. For each slope, the public coverage group shows a significant departure from parallelism (Q_C = 6.96, 6.14, and 8.42 for slopes 1, 2, and 3, respectively, each with 1 d.f.). A reduced model in which the private, part-year, and no coverage groups have parallel cumulative logit distributions; the public group departs from parallelism; and all four groups have different proportions with no charge can be expected to fit the data well.

The ratios of Q_C values for each of the sample design options shown in the last column of Table 23 indicate that simple random sampling inference is, on average, liberal compared with inferences based on the actual sample design. However, in this case, despite the larger test statistic values for the simple random sampling inferences, similar conclusions about the importance of parameters in the model are drawn under either sample selection assumption.

These results are consistent with results for design effects in the sample survey literature. The ratio of $Q_{C(s)}$ to $Q_{C(c)}$ cannot be interpreted as a design effect (except, perhaps, for statistics with d.f. = 1). In this instance, the design effects for the sample proportions in Table 20 range from 0.70 to 1.85, with an average design effect of 1.39 for private, 1.18 for public, 1.27 for part-year, and 0.95 for no health care coverage. The average of the Q_C ratios in Table 23 is 1.07. The finding that the Q_C ratios are smaller than design effects for the subpopulation proportions, but on average still greater than 1.0, is consistent with findings of Kish and Frankel (1974), who observed that the effect of the sample design diminishes as the statistic becomes more "complicated."

A reduced model matrix with differing intercepts and parallelism for three of the four health care coverage groups,

$$X_R = \begin{bmatrix} 1\,100\,000\,000 \\ 1\,100\,100\,000 \\ 1\,100\,110\,000 \\ 1\,100\,111\,000 \\ 1\,010\,000\,000 \\ 1\,010\,100\,100 \\ 1\,010\,110\,010 \\ 1\,010\,111\,001 \\ 1\,001\,000\,000 \\ 1\,001\,100\,000 \\ 1\,001\,110\,000 \\ 1\,001\,111\,000 \\ 1\,000\,000\,000 \\ 1\,000\,100\,000 \\ 1\,000\,110\,000 \\ 1\,000\,111\,000 \end{bmatrix}$$

was fit to the observed vector of cumulative logits, $\mathbf{L}$. The lack-of-fit statistic under the complex sample design assumptions was Q = 2.26 with 6 degrees of freedom (p = 0.89). On the other hand, under simple random sampling inference, Q = 2.18 (p = 0.90). Thus, in either case, it can be concluded that X_R provides a satis-

factory characterization of the variation in these sample logits.

In Table 24 the estimated parameter values and standard errors for the reduced model are presented. The negative intercept coefficients indicate that the predicted cumulative logit intercepts for the private, public, and part-year coverage groups are less than the intercept for the no coverage group (i.e., $\hat{\mu} = 0.238$). Hence, the proportions with no charge in the private, public, and part-year coverage groups are smaller than the proportion for the no coverage group. The slope coefficients are all positive, indicating increasing cumulative logits from one threshold to the next, and under the reduced model, the slopes for the private, part-year, and no coverage groups are the same. The slope differentials for the public coverage group are negative, indicating that this group cumulates more slowly than the other three coverage groups. In other words, persons aged 18–64 with public health care coverage tend to have proportionately higher physician charges than persons in the other groups have.

The parameter estimates in Table 24 can be used to obtain various predicted values under the reduced model to aid interpretation. For example, the predicted cumulative logits and cumulative proportions can be calculated directly from the parameter estimates. Predicted cumulative proportions can also be obtained from the predicted cumulative logits as $\hat{F}_{ij} = [1 + exp(-\hat{L}_{ij})]^{-1}$. Both types of predicted quantities are shown in Table 25. These values reflect the differing intercepts and parallelism features of the reduced model.

Predicted odds ratios can also be calculated from the predicted cumulative logits. For example, the predicted ratio of private group odds to no coverage group odds that a charge greater than the jth threshold will be incurred is computed as $exp(\hat{L}_{4j} - \hat{L}_{1j})$. The predicted ratio of each coverage group's odds to the odds for the no coverage group is given in Table 26. Under the parallelism model, the odds of incurring a charge greater than each threshold are the same for each threshold for the private and part-year coverage groups. However, the odds of incurring a higher charge at higher thresholds increase faster for the public coverage group than for the no coverage group. Thus, the lack of parallelism can be illustrated in terms of both the predicted logits and predicted odds ratios.

Concluding Remarks

The estimation of the complex sample design covariance matrix for the cumulative logits is more complicated than that required under simple random sampling assumptions. In this instance, the variance and covariance calculations are based on 69 differences in the paired sampling error computations, which are more expensive to compute than the simple random sampling approach. However, available sampling error software can estimate the covariance matrixes required as input to the WLS procedure conveniently and without unreasonable expense. Further, the effect of the complex sample design on inferences must be considered. Although the illustration for physician visit charges has demonstrated only modest design effects, one must take the design into account at some point in a set of analyses to determine at least the general size of the effects. Although a sequential survey analysis strategy can be followed, as suggested in Landis et al. (1982), the increasing availability of sampling error estimation software makes application to survey data of existing methods taking the sample design into account more feasible.

The cumulative logit modeling strategy provides a means to compare frequency distributions across population subgroups that is not limited to a comparison of single measures of central tendency (e.g., means, medians). The method allows the examination of variation across the entire distribution of the response variable. Further, this methodology provides a single model for limited dependent variables. Although not illustrated here, it can also be used to compare frequency distributions for dependent variables that do not have a limiting value and for measures that have nonstandard distributions, such as extremely skewed distributions.

In the illustration, only a single predictor was employed, but several predictors can be considered by cross-classifying the categories of the predictors to form subpopulations. The model testing strategy can then be used to examine the nature of main and interaction effects among the predictors for intercepts, as well as for slope differentials, across the subpopulations formed by the cross-classification. In addition, the intercept and slope differential model parameters can be specified in a variety of ways to provide more insightful inference about the relationships among the subpopulation frequency distributions. Besides the reference cell parametrization illustrated here, the standard ANOVA sum-to-zero and, for intervally scaled predictors, orthogonal polynomial (e.g., linear, quadratic, and cubic) parametrizations can also be used. For ordinally scaled predictors, scores can be assigned to the categories and incorporated into the analysis as well. Finally, the methodology can be generalized in terms of the response variable through a cross-classification of two or more dependent variables whose distributions have been categorized by suitable thresholds. Williams and Grizzle (1972) suggest models for multivariable responses in cumulative logit analysis.

Thus, the cumulative logit method can be applied to a wide range of problems in the analysis of frequency distribution data involving multiple factors and responses with a variety of parametrizations of the model. The WLS framework for categorical data analysis provides a basis for incorporating a complex sample design into the analysis, and the results can be interpreted in terms of predicted frequencies, cumulative logits, or odds ratios.

References

Agresti, A.: *Analysis of Ordinal Categorical Data*. New York. John Wiley & Sons, Inc., 1985.

Andersen, R., and Anderson, O. W.: *A Decade of Health Services*. Chicago. University of Chicago Press, 1967.

Andersen, R., Lion, J., and Anderson, O. W.: *Two Decades of Health Services, Social Survey Trends in Use and Expenditure*. Cambridge, Mass. Ballinger Publishing Co., 1976.

Anderson, O. W., and Feldman, J. J.: *Family Medical Costs and Voluntary Health Insurance, A Nationwide Survey*. New York. McGraw-Hill Book Co., Inc., 1956.

Anderson, O. W., Colette, P., and Feldman, J. J.: *Changes in Family Medical Expenditures and Voluntary Health Insurance*. Cambridge, Mass. Harvard University Press, 1963.

Berki, S. E., Lepkowski, J. M., Wyszewianski, L., et al.: High-volume and low-volume users of health services, United States, 1980. *National Medical Care Utilization and Expenditure Survey*. Series C, No. 2. DHHS Pub. No. 86–20402. National Center for Health Statistics, Public Health Service. Washington. U.S. Government Printing Office, 1985.

Bhapkar, V. P., and Koch, G. G.: Hypotheses of "no interaction" in multidimensional contingency tables. *Technometrics* 10:107–123, 1968.

Casady, R. J.: The survey design and estimation methodology for the NMCUES, in *Proceedings of the 19th National Meeting of the Public Health Conference on Records and Statistics*. DHHS Pub. No. (PHS) 81–1214. National Center for Health Statistics, Public Health Service. Hyattsville, Md., 1983, pp. 181–185.

Cohen, S. B., Burt, V. L., and Jones, G. K.: Efficiencies in variance estimation for complex survey data. *The American Statistician* 40:157–164, 1986.

Computer Support Group, Institute for Social Research: *OSIRIS IV User's Guide*. Ann Arbor, Mich. Institute for Social Research, The University of Michigan, 1982.

Cox, B.: The weighted sequential hot-deck imputation procedure, in *Proceedings of the Section on Survey Research Methods, American Statistical Association*. 1980, pp. 721–726.

Dempster, A. P., and Rubin, D. B.: Introduction, in W. G. Madow, I. Olkin, and D. Rubin, eds., *Incomplete Data in Sample Surveys, Volume 2, Theory and Bibliographies*. New York. Academic Press, 1983.

Duan, N., Manning, W. G., Jr., Morris, C. N., and Newhouse, J. P.: *A Comparison of Alternative Models for the Demand for Care*. Santa Monica, Calif. The Rand Corporation, 1982.

Falk, I. S., Klem, M. C., and Sinai, N.: The incidence of illness and the receipt of costs of medical care among representative families, experiences in twelve consecutive months during 1928–31. *Publications of the Committee on the Costs of Medical Care*. No. 25. Chicago. University of Chicago Press, 1933.

Ford, B. L.: An overview of hot-deck procedures, in W. G. Madow, I. Olkin, and D. Rubin, eds., *Incomplete Data in Sample Surveys, Volume 2, Theory and Bibliographies*. New York. Academic Press, 1983.

Frankel, M. R.: *Inference from Complex Samples*. Ann Arbor, Mich. Institute for Social Research, 1974.

Freeman, D. E., Freeman, J. L., Brock, D., and Koch, G. G.: Strategies in the multivariate analysis of data from complex surveys, II, An application to the United States National Health Interview Survey. *International Statistical Review* 44:317–330, 1976.

Grizzle, J. E., Starmer, C. F., and Koch, G. G.: Analysis of categorical data by linear models. *Biometrics* 25:489–503, 1969.

Harlan, W. R., Murt, H. A., Thomas, J. W., et al.: Incidence, utilization, and costs associated with acute respiratory conditions, United States, 1980. *National Medical Care Utilization and Expenditure Survey*. Series C, No. 4. DHHS Pub. No. 86–20404. National Center for Health Statistics, Public Health Service. Washington. U. S. Government Printing Office, Sept. 1986.

Harlan, W. R., Thomas, J. W. Murt, H. A., et al.: Health care utilization and costs associated with acute cardiovascular conditions, United States, 1980. *National Medical Care Utilization and Expenditure Survey*. Series C, No. 7. To be published.

Heckman, J. J.: The common structure of statistical models of truncation, sample selection and limited dependent variables, and a simple estimator for such models. *Annals of Economic and Social Measurement* 5(4):475–492, 1976.

Hidiroglou, M. A., Fuller, W. A., and Hickman, R. D.: *SUPER CARP*. Ames, Iowa. Department of Statistics, Iowa State University, 1974.

Kalton, G.: *Compensating for Missing Survey Data*. Ann Arbor, Mich. Survey Research Center, 1983.

Kalton, G., and Kasprzyk, D.: Imputing for missing survey responses, in *Proceedings of the Section on Survey Research Methods, American Statistical Association*. 1982, pp. 22–31.

Kish, L.: *Survey Sampling*. New York. John Wiley & Sons, Inc., 1965.

Kish, L., and Frankel, M. R.: Inference from complex samples. *Journal of the Royal Statistical Society* 36:1–37, 1974.

Koch, G. G., and Lemeshow, S.: An application of multivariate analysis to complex sample survey data. *Journal of the American Statistical Association* 67:780–782, 1972.

Koch, G. G., Freeman, D. H., and Freeman, J. L.: Strategies in the multivariate analysis of data from complex surveys. *International Statistical Review* 43:59–78, 1975.

Landis, J. R., Lepkowski, J. M., Eklund, S. A., and Stehouwer, S. A.: A statistical methodology for analyzing data from a complex survey, the first National Health and Nutrition Examination Survey.

Vital and Health Statistics. Series 2, No 92. DHHS Pub. No. 82–1366. Public Health Service. Washington. U.S. Government Printing Office, 1982.

Landis, J. R., Stanish, W. M., Freeman, J. L., and Koch, G. G.: A computer program for the generalized chi-square analysis of categorical data using weighted least squares (GENCAT). *Comput Programs Biomed* 6:196–231, 1976.

LaVange, L., Iannacchione, V., and Garfinkel, S.: An application of logistic regression methods to survey data, predicting high cost users of medical care, in *Proceedings of the Section on Survey Research Methods, American Statistical Association*. 1986.

McCullagh, P.: Regression models for ordinal data. *Journal of the Royal Statistical Society*. Series B, 42:109–142, 1980.

Mohadjer, L., Morganstein, D., Chu, A. and Rhoads, M.: Estimation and analysis of survey data using SAS procedures—WESVAR, NASSREG, and NASSLOG, in *Proceedings of the Section on Survey Research Methods, American Statistical Association*. 1986.

Murt, H. A., Parsons, P. E., Harlan, W. R., et al.: Disability, utilization, and costs associated with musculoskeletal conditions, United States, 1980. *National Medical Care Utilization and Expenditure Survey*. Series C, No. 5. DHHS Pub. No. 86–20405. National Center for Health Statistics, Public Health Service. Washington. U.S. Government Printing Office, Sept. 1986.

Parsons, P. E., Lichtenstein, R., Berki, S. E., et al.: The costs of illness, United States, 1980. *National Medical Care Utilization and Expenditure Survey*. Series C, No. 3. DHHS Pub. No. 86–20403. National Center for Health Statistics, Public Health Service. Washington. U.S. Government Printing Office, Apr. 1986.

Perrot, G. S., Tibbets, C., and Britton, R. H.: The National Health Survey, scope and methods of the nationwide canvass of sickness in relation to its social and economic setting. *Public Health Rep* 54:1663–1687, Sept. 1939.

Research Triangle Institute: *Public Use Data Tape Documentation, National Medical Care Utilization and Expenditure Survey, 1980*. National Center for Health Statistics. Hyattsville, Md., Aug. 1983.

Rust, K.: Variance estimation for complex estimators in sample surveys. *Journal of Official Statistics* 1:381–397, 1986.

Santos, R.: Effects of imputation on regression coefficients, in *Proceedings of the Section on Survey Research Methods, American Statistical Association*. 1981, pp. 140–145.

SAS Institute, Inc.: *SAS User's Guide, Statistics,* 1985 ed. Cary, N.C. SAS Institute Inc., 1985.

Schnell, D., Sullivan, G., Kennedy, W. J., and Fuller, W. A.: *P C CARP: Variance Estimation for Complex Surveys*. Ames, Iowa. Statistical Research Laboratory, Iowa State University, 1986.

Shah, B. V.: Software for survey data analysis. *The American Statistician* 38:68, 1984.

Tobin, J.: Estimation of relationships for limited dependent variables. *Econometrica* 26:24–36, 1958.

Wald, A.: Tests of statistical hypotheses concerning several parameters when the number of observations is large. *Transactions of the American Mathematical Society* 54:426–482, 1943.

Williams, O. D., and Grizzle, J. E.: Analysis of contingency tables having ordered response categories. *Journal of the American Statistical Association* 67:55–63, 1972.

List of Detailed Tables

1. Target average population for the National Medical Care Utilization and Expenditure Survey, by sex, race and age: 1980 55
2. Revisions to time-adjusted weight and time-adjustment factor for newborn records: National Medical Care Utilization and Expenditure Survey, 1980 56
3. Revisions to hospital charges and source-of-payment amounts in hospital stay file records: National Medical Care Utilization and Expenditure Survey, 1980 56
4. Revisions to total and out-of-pocket hospital charges in person file records: National Medical Care Utilization and Expenditure Survey, 1980 57
5. Revisions to health care coverage codes: National Medical Care Utilization and Expenditure Survey, 1980 57
6. Health care coverage codes for persons under 65 years of age, by type and source of coverage: National Medical Care Utilization and Expenditure Survey, 1980 58
7. Health care coverage codes for persons 65 years of age and over, by type and source of coverage: National Medical Care Utilization and Expenditure Survey, 1980 58
8. Revisions to bed-disability days and restricted-activity days and unaltered values for hospital nights in person file records: National Medical Care Utilization and Expenditure Survey, 1980 59
9. Revisions to hospital admission and discharge dates and related information in hospital stay file and person file records: National Medical Care Utilization and Expenditure Survey, 1980 62
10. Revisions to poverty status code in person file records: National Medical Care Utilization and Expenditure Survey, 1980 62
11. Revisions to person file records for persons reporting deliveries: National Medical Care Utilization and Expenditure Survey, 1980 62
12. Revisions to hospital stay file records for persons reporting deliveries: National Medical Care Utilization and Expenditure Survey, 1980 63
13. Revisions to person file records for person identifying number 5031 (duplicate hospital record): National Medical Care Utilization and Expenditure Survey, 1980 63
14. Rate of imputation, sample size, mean, standard errors, and square root of design effect for 3 disability measures and 2 data types: National Medical Care Utilization and Expenditure Survey, 1980 64
15. Sample size, mean, standard errors, square root of design effect, and element standard deviation for total charge for outpatient department visit, by data type: National Medical Care Utilization and Expenditure Survey, 1980 64
16. Rate of imputation, sample size, mean, and standard errors for total charge for outpatient department visits for 4 family income groups and 3 estimation methods: National Medical Care Utilization and Expenditure Survey, 1980 65
17. Values of *roh* and $\hat{s}^2$ for standard error formula for estimated means: National Medical Care Utilization and Expenditure Survey, 1980 65
18. Estimated logistic regression coefficients and odds ratios for use of hospital care for all persons, by selected characteristics and indicators: United States, 1980 67
19. Sample frequency distribution, by threshold level and subpopulation 68
20. Proportion of person years for persons 18–64 years of age, by physician visit charge threshold and selected types of health care coverage: United States, 1980 68
21. Cumulative proportion and cumulative logit of person years for persons 18–64 years of age, by physician visit charge threshold and selected types of health care coverage: United States, 1980 68

22. Estimates of parameters and their standard errors under the saturated model X_S: United States, 1980 69
23. Analysis of variance for cumulative logit analysis for persons 18–64 years of age under 2 sample design options for physician visit charges, by selected types of health care coverage: United States, 1980 69
24. Estimates of parameters and their standard errors under the reduced model X_R: United States, 1980 70
25. Predicted cumulative proportion and predicted cumulative logit of person years for persons 18–64 years of age, by physician visit charge threshold and selected types of health care coverage under reduced model X_R: United States, 1980 70
26. Predicted odds ratios under reduced model X_R for odds of incurring charges greater than the given threshold value: United States, 1980 70

Table 1

Target average population for the National Medical Care Utilization and Expenditure Survey, by sex, race, and age: 1980

Age	Total	Male			Female		
		All races	Black	White and other	All races	Black	White and other
		Number in thousands					
All ages	226,368	109,362	12,311	97,051	117,006	14,242	102,763
Under 1 year	6,868	3,600	551	3,049	3,268	491	2,777
1–4 years	13,098	6,516	957	5,559	6,582	1,034	5,548
5–9 years	16,872	8,746	1,294	7,452	8,126	1,194	6,932
10–14 years	18,260	9,231	1,328	7,904	9,029	1,440	7,589
15–16 years	8,584	4,572	647	3,925	4,012	576	3,436
17–19 years	12,436	6,142	890	5,253	6,293	843	5,451
20–24 years	20,783	9,848	1,028	8,820	10,936	1,394	9,542
25–29 years	18,902	9,372	1,003	8,369	9,529	1,266	8,264
30–34 years	17,125	8,251	812	7,439	8,874	975	7,898
35–44 years	25,599	12,351	1,161	11,190	13,248	1,570	11,678
45–49 years	10,850	5,303	511	4,793	5,547	629	4,918
50–54 years	11,669	5,592	496	5,096	6,077	580	5,497
55–64 years	21,304	10,077	842	9,235	11,227	1,020	10,207
65–74 years	15,319	6,586	615	5,971	8,733	779	7,954
75 years and over	8,699	3,174	177	2,997	5,525	452	5,073

Table 2

Revisions to time-adjusted weight and time-adjustment factor for newborn records: National Medical Care Utilization and Expenditure Survey, 1980

Person identifying number	Revised value: Time-adjusted weight	Revised value: Time-adjustment factor
1675	10,415	0.8825
1793	1,450	0.1776
2120	10,525	0.7896
3343	16,181	0.9590
3418	12,173	0.9153
3488	9,970	0.9235
3722	13,933	0.9809
3865	10,168	0.8607
4082	8,822	0.9863
4876	5,326	0.5519
5137	11,591	0.8770
5431	15,902	0.8443
7247	10,273	0.8470
7310	12,439	0.9836
7515	13,793	0.9809
7710	15,678	0.9508
7721	11,163	0.7787
8430	8,667	0.6639
8434	2,741	0.2869
8729	10,148	0.5902
8741	13,648	0.7268
8812	10,383	0.4863
8838	3,645	0.2322
8841	4,438	0.3852
8875	1,125	0.1066
8895	1,306	0.0792
10517	14,659	0.9071
12170	21,589	0.9891
12182	12,064	0.8197
12463	13,470	0.9153
12479	14,618	0.9126
12637	10,836	0.9699
13014	12,751	0.9645
13226	11,150	0.8989
13492	9,395	0.8962
13658	12,165	0.9454
13931	11,300	0.7678
14144	10,710	0.8634
14253	10,959	0.9809
14874	11,523	0.9290
15026	10,641	0.8689
15218	12,426	0.8443
15516	10,222	0.7732
15937	12,822	0.9699
15940	15,915	0.8525
16575	12,132	0.9781
16589	10,111	0.9645
16615	12,353	0.9344
16628	13,072	0.8470
16654	36	0.0027
16663	12,767	0.8770
16820	13,848	0.9563
17041	15,887	0.9918
17390	13,661	0.8852
17634	3,948	0.2377
17643	9,595	0.9153
17746	17,973	0.9098
17818	8,335	0.7951
17936	14,420	0.8251
17957	8,678	0.6284
17965	7,691	0.6803
17987	9,392	0.6393
18107	5,557	0.5301
18111	7,354	0.4208
18131	4,033	0.3251
18277	4,169	0.3361
18305	10,235	0.6967
18355	2,189	0.1311

Table 3

Revisions to hospital charges and source-of-payment amounts in hospital stay file records: National Medical Care Utilization and Expenditure Survey, 1980

Person identifying number and hospital record number	Revised value: Total hospital charges	Revised value: Source-of-payment amount: First	Revised value: Source-of-payment amount: Second	Revised value: Source-of-payment amount: Third
1570/230	1,164.05	232.81	931.24	...
17594/2956	1,000.00	333.33	333.33	333.34
11313/1982	9,000.00	1,000.00	8,000.00	0
1282/198	683.35	180.00	503.35	...
4920/832	1,000.00	15.00	985.00	...
3015/496	9,000.00	4,500.00	4,500.00	...

Table 4

Revisions to total and out-of-pocket hospital charges in person file records: National Medical Care Utilization and Expenditure Survey, 1980

Person identifying number	Revised value			
	Total hospital charges	Hospital out-of-pocket charges	Total charges	Total out-of-pocket charges
1570	1,164	0	2,962	415
17594	5,768	1,923	6,273	2,116
11313	12,650	1,208	17,918	1,902
1282	2,796	180	3,292	293
4920	1,000	15	1,781	15
3015	12,993	1,446	15,404	1,986

Table 5

Revisions to health care coverage codes: National Medical Care Utilization and Expenditure Survey, 1980

Person identifying number	Original code	Revised code
65 years and over		
409	Any other coverage	Medicare and private
1203	Any other coverage	Medicare and other public
1532	Any other coverage	Medicare and private
1876	Any other coverage	Medicare and private
2134	No coverage	Medicare only
2924	Any other coverage	Medicare and private
3159	No coverage	Medicare only
3218	Any other coverage	Medicare and other public
3299	Any other coverage	Medicare and private
3403	Any other coverage	Medicare and private
3928	Any other coverage	Medicare and private
3982	Any other coverage	Medicare and private
3983	Any other coverage	Medicare and private
5031	No coverage	Medicare only
5102	Any other coverage	Medicare and private
5207	Any other coverage	Medicare and private
5579	Any other coverage	Medicare and private
6172	Any other coverage	Medicare and other public
6365	Any other coverage	Medicare and private
6831	No coverage	Medicare only
7407	Any other coverage	Medicare and private
8091	Any other coverage	Medicare and other public
9330	No coverage	Medicare only
9445	Any other coverage	Medicare and private
9546	Any other coverage	Medicare and other public
9547	Any other coverage	Medicare and private
10537	Any other coverage	Medicare and private
10733	Any other coverage	Medicare and other public
11262	Any other coverage	Medicare and private
11791	Any other coverage	Medicare and private
11835	No coverage	Medicare and private
12133	No coverage	Medicare only
12322	Any other coverage	Medicare and private
12529	Any other coverage	Medicare and private
12790	Any other coverage	Medicare and other public
12830	No coverage	Medicare and private
13130	Any other coverage	Medicare and private
13685	Any other coverage	Medicare and private
13980	No coverage	Medicare only
14569	Any other coverage	Medicare and private
15302	Any other coverage	Medicare and private
15314	Any other coverage	Medicare and private
16540	No coverage	Medicare only
17011	Any other coverage	Medicare and private
17869	No coverage	Medicare only
17913	No coverage	Medicare only
Under 65 years		
6017	No coverage	Other public program only[1]

[1]Persons under 65 years of age who were covered by Medicare only were coded as "other public program only."

Table 6

Health care coverage codes for persons under 65 years of age, by type and source of coverage: National Medical Care Utilization and Expenditure Survey, 1980

Source of coverage	Type of coverage						
	Only 1 source				Mixed[2]		
	All year						
	Private	Medicaid	Other public[1]	Part year[1]	Private and public	More than 1 public[3]	None
	Code						
Private	P.81:1	P.81:3	P.81:3	P.82:2	P.81:1,2	P.81:3	P.81:3
Public:							
Medicaid	P.75:3	P.75:1	P.75:3	P.75:2	P.75:1,2	P.75:1,2	P.75:3
Other public:							
Medicare	P.68:5	P.68:5	P68:2,3	...	P.68:2,3	P.68:2,3	P.68:5
CHAMPUS/CHAMPVA[4]	P.69:3	P.69:3	P.69:1	P.69:2	P.69:1,2	P.69:1,2	P.69:3
Indian Health Service	P.70:3	P.70:3	P.70:1	P.70:2	P.70:1,2	P.70:1,2	P.70:3
Other	P.76:3	P.76:3	P.76:1	P.76:2	P.76:1,2	P.76:1,2	P.76:3

[1]Persons covered by only 1 source must be coded as "not covered" for all but 1 source. Therefore, persons in this category will have 1 of the "covered" codes listed, and all other sources will be coded "not covered."
[2]It is assumed that most but not all individuals in this category will be covered by at least 1 source all year.
[3]Persons in this category will have more than 1 of the public sources listed.
[4]Civilian Health and Medical Program of the Uniformed Services or Civilian Health and Medical Program of the Veterans' Administration.

NOTE: Codes refer to label and code numbers for variables in the public use files (Research Triangle Institute, 1983).

Table 7

Health care coverage codes for persons 65 years of age and over, by type and source of coverage: National Medical Care Utilization and Expenditure Survey, 1980

Source of coverage	Medicare			No Medicare	
	1 or more other sources				
	Private	Public[1]	Only	Any other source[2]	None
			Code		
Private	P.81:1,2	P.81:3	P.81:3	P.81:1,2	P.81:3
Public:					
Medicare	P.68:1	P.68:1	P.68:1	P.68:4	P.68:4
Other public:					
Medicaid	P.75:3	P.75:1,2	P.75:3	P.75:1,2	P.75:3
CHAMPUS/CHAMPVA[3]	P.69:3	P.69:1,2	P.69:3	P.69:1,2	P.69:3
Indian Health Service	P.70:3	P.70:1,2	P.70:3	P.70:1,2	P.70:3
Other	P.76:3	P.76:1,2	P.76:3	P.76:1,2	P.76:3

[1]Persons in this category may have more than 1 of the other public sources listed below.
[2]Persons in this category may have more than 1 source but must be coded as "not covered" by Medicare.
[3]Civilian Health and Medical Program of the Uniformed Services or Civilian Health and Medical Program of the Veterans' Administration.

NOTE: Codes refer to label and code numbers for variables in the public use files (Research Triangle Institute, 1983).

Table 8

Revisions to bed-disability days and restricted-activity days and unaltered values for hospital nights in person file records: National Medical Care Utilization and Expenditure Survey, 1980

Person identifying number	Bed-disability days		Restricted-activity days		Hospital nights
	Original value	Revised value	Original value	Revised value	
	Number				
147	4	6	50	52	6
155	0	2	0	2	2
243	4	18	111	125	18
370	0	1	1	2	1
405	8	21	87	100	21
429	10	11	28	29	11
545	15	38	15	38	38
789	18	20	46	48	20
812	0	7	0	7	7
928	14	22	14	22	22
1037	14	31	14	31	31
1038	0	2	17	19	2
1244	11	15	11	15	15
1419	0	37	0	37	37
1451	0	1	2	3	1
1463	0	1	93	94	1
1532	0	38	0	38	38
1623	0	3	0	3	3
1628	34	35	277	278	35
1686	25	48	25	48	48
1778	22	34	84	96	34
1817	7	10	12	15	10
1874	53	81	65	93	81
2492	10	11	73	74	11
2497	4	5	4	5	5
2649	5	7	12	14	7
2785	11	21	267	277	21
3106	117	123	117	123	123
3140	14	22	14	22	22
3218	4	59	21	76	59
3340	0	3	0	3	3
3384	8	14	8	14	14
3417	0	3	0	3	3
3475	3	5	3	5	5
3539	103	107	103	107	107
3549	0	1	11	12	1
3574	4	7	4	7	7
3626	37	39	67	69	39
3864	0	4	0	4	4
4020	2	50	2	50	50
4111	12	29	14	31	29
4521	3	4	38	39	4
4574	11	15	24	28	15
4698	3	4	8	9	4
4789	5	7	53	55	7
4847	0	1	5	6	1
4970	9	11	11	13	11
5031	73	139	130	196	139
5045	0	9	0	9	9
5077	8	9	22	23	9
5098	16	17	16	17	17
5136	0	4	4	8	4
5172	8	9	22	23	9
5289	1	2	1	2	2
5320	21	22	195	196	22
5340	9	10	78	79	10
5468	18	44	32	58	44
5685	17	44	87	114	44

Table 8—Con.

Revisions to bed-disability days and restricted-activity days and unaltered values for hospital nights in person file records: National Medical Care Utilization and Expenditure Survey, 1980

Person identifying number	Bed-disability days		Restricted-activity days		Hospital nights
	Original value	Revised value	Original value	Revised value	
	Number				
5723	0	2	0	2	2
5763	0	10	0	10	10
5772	5	6	22	23	6
5877	2	3	2	3	3
6010	0	94	0	94	94
6116	0	6	2	8	6
6487	1	3	1	3	3
6673	9	21	122	134	21
6735	20	31	110	121	31
6797	9	10	17	18	10
6880	5	6	22	23	6
6937	8	9	86	87	9
7046	62	70	280	288	70
7076	0	5	0	5	5
7149	0	17	0	17	17
7202	1	22	34	55	22
7299	2	61	2	61	61
7359	106	129	110	133	129
7417	4	5	89	90	5
7533	2	3	7	8	3
7637	7	8	7	8	8
7706	0	6	24	30	6
7797	0	3	52	55	3
7800	0	15	14	29	15
7910	23	84	85	146	84
8047	0	1	2	3	1
8258	6	7	6	7	7
9057	3	4	3	4	4
9350	0	2	0	2	2
9445	92	99	122	129	99
9477	35	38	101	104	38
9531	4	9	4	9	9
9547	10	12	25	27	12
9563	9	11	52	54	11
9597	55	68	155	168	68
9645	49	55	75	81	55
9807	14	15	41	42	15
9904	3	4	3	4	4
10070	5	6	5	6	6
10458	6	7	26	27	7
10513	3	5	5	7	5
10536	2	20	2	20	20
10550	7	16	8	17	16
10598	4	5	11	12	5
10720	1	2	2	3	2
10837	12	20	101	109	20
10853	3	5	55	57	5
10896	3	4	19	20	4
11010	20	26	20	26	26
11051	3	4	3	4	4
11127	0	3	0	3	3
11170	0	3	63	66	3
11183	3	4	34	35	4
11250	217	247	217	247	247
11305	15	17	123	125	17
11343	1	2	1	2	2
11627	3	21	3	21	21
11781	2	4	12	14	4
11996	2	10	2	10	10

Table 8—Con.

Revisions to bed-disability days and restricted-activity days and unaltered values for hospital nights in person file records: National Medical Care Utilization and Expenditure Survey, 1980

Person identifying number	Bed-disability days		Restricted-activity days		Hospital nights
	Original value	Revised value	Original value	Revised value	
	Number				
12122	43	52	43	52	52
12133	10	11	169	170	11
12202	12	17	20	25	17
12543	0	8	5	13	8
12747	22	48	24	50	48
12764	20	27	185	192	27
12830	0	4	7	11	4
12845	7	22	7	22	22
12859	0	2	21	23	2
13013	4	5	7	8	5
13105	2	4	2	4	4
13145	0	1	0	1	1
13196	7	8	28	29	8
13219	2	3	2	3	3
13396	9	10	9	10	10
13489	0	3	0	3	3
13556	21	32	38	49	32
13799	3	18	3	18	18
13831	14	15	43	44	15
14107	0	2	0	2	2
14254	13	24	39	50	24
14332	1	2	31	32	2
14545	21	24	257	260	24
14649	12	17	144	149	17
14913	6	7	7	8	7
15001	1	2	1	2	2
15014	4	5	44	45	5
15097	2	4	18	20	4
15135	8	9	15	16	9
15270	3	5	25	27	5
15295	15	24	18	27	24
15299	0	3	3	6	3
15311	30	51	36	57	51
15339	27	28	33	34	28
15393	1	2	1	2	2
15410	39	49	71	81	49
15411	0	6	1	7	6
15423	0	28	28	56	28
15465	0	2	0	2	2
15479	44	50	65	71	50
15494	4	5	29	30	5
15526	4	9	75	80	9
15983	0	12	0	12	12
16088	2	3	2	3	3
16165	0	2	20	22	2
16257	2	64	2	64	64
16304	28	43	42	57	43
16419	2	4	2	4	4
16586	0	2	0	2	2
16605	30	71	30	71	71
16647	7	8	21	22	8
16713	1	2	9	10	2
16870	20	22	20	22	22
17400	8	12	16	20	12
17503	1	6	1	6	6
17578	5	7	73	75	7
17765	0	5	4	9	5
17826	0	3	0	3	3

Table 9

Revisions to hospital admission and discharge dates and related information in hospital stay file and person file records: National Medical Care Utilization and Expenditure Survey, 1980

Person identifying number	Hospital stay file			Person file		
	Date of admission (revised)	Date of discharge (revised)	Hospital nights (revised)	Hospital nights (revised)	Bed-disability days	Restricted-activity days
	Code			Number		
11176	80040	80043	3	3	3	8
11445	80262	80270	8	8	8	14
7354	80260	80262	2	2	2	12
37722	80060	80063	3	3	3	5

Table 10

Revisions to poverty status code in person file records: National Medical Care Utilization and Expenditure Survey, 1980

Person identifying number	Unaltered continuous *poverty status value*	Revised categorical *poverty status code*
13809	1.24	2
13940	1.88	5
15402	2.42	7
17529	1.46	3

Table 11

Revisions to person file records for persons reporting deliveries: National Medical Care Utilization and Expenditure Survey, 1980

Revised variable	Person identifying number					
	17578	17987	17826	17827	3712	11720
Number of inpatient physician visits	‡	2	0	1	‡	‡
Number of hospital stays	1	1	0	1	‡	‡
Number of hospital nights	4	3	0	3	‡	‡
Inpatient physician charges	198	198	999999	125	‡	‡
Hospital stay charges	‡	0	999999	994	‡	‡
Total charges	843	209	999999	2,017	‡	‡
Inpatient physician out-of-pocket charges	45	45	‡	0	‡	‡
Hospital stay out-of-pocket charges	‡	0	‡	0	‡	‡
Total out-of-pocket charges	99	56	‡	‡	‡	‡
Sex code	‡	‡	‡	‡	2	2

NOTE: ‡ = *data entry unaltered.*
999999 = not applicable.

Table 12

Revisions to hospital stay file records for persons reporting deliveries: National Medical Care Utilization and Expenditure Survey, 1980

Person identifying number	Revision
17826	Record 2986 deleted
17827	Record 2986 transferred from 17826
8471	No condition at admission = 2
8922	No condition at admission = 2
18392	No condition at admission = 2
18107	No condition at admission = 2 on record 3057
17578	Record 2953 deleted
17987	Record 2953 transferred from 17578
13809	Number of conditions at admission = 9; first condition = 46

Table 13

Revisions to person file records for person identifying number 5031 (duplicate hospital record): National Medical Care Utilization and Expenditure Survey, 1980

Revised variable	Revised value
Number of bed-disability days	73
Number of restricted-activity days	130
Number of inpatient physician visits	10
Number of hospital stays	2
Number of hospital nights	71
Inpatient physician charges	4,377
Hospital stay charges	20,965
Total charges	26,026
Inpatient physician out-of-pocket charges	851
Hospital stay out-of-pocket charges	380
Total out-of-pocket charges	1,707
Hospital record numbers deleted	850 & 851

NOTE: Hospital record numbers 848 and 851 matched Medicare estimates from the Health Care Financing Administration. Deletion of these two records would be technically better than deletion of record numbers 850 and 851. If 848 and 851 were deleted instead of 850 and 851, hospital charges would equal $21,145. The University of Michigan deleted record numbers 850 and 851 as a result of miscommunication with the National Center for Health Statistics.

Table 14

Rate of imputation, sample size, mean, standard errors, and square root of design effect for 3 disability measures and 2 data types: National Medical Care Utilization and Expenditure Survey, 1980

Disability measure and data type	Rate of imputation	Sample size	Mean	Standard error: Simple random sampling	Standard error: Complex	Square root of design effect
Bed-disability days						
All data	0.08	17,123	5.268	0.1269	0.1540	1.21
Real data	...	15,777	5.228	0.1319	0.1599	1.21
Work-loss days						
All data	0.12	13,069	3.696	0.1220	0.1629	1.34
Real data	...	11,537	3.574	0.1277	0.1716	1.34
Work-loss days in bed						
All data	0.16	13,069	1.568	0.0518	0.0592	1.14
Real data	...	10,970	1.578	0.0568	0.0652	1.15
Cut-down days						
All data	0.08	17,123	6.881	0.1697	0.3343	1.97
Real data	...	15,724	6.639	0.1735	0.3322	1.91
Restricted-activity days						
All data	0.18	17,123	13.805	0.2573	0.4716	1.83
Real data	...	14,049	13.064	0.2742	0.4658	1.70

NOTE: Estimates in this table are presented for illustrative purposes. Calculations were made prior to modifications described in the section on public use data files.

Table 15

Sample size, mean, standard errors, square root of design effect, and element standard deviation for total charge for outpatient department visit, by data type: National Medical Care Utilization and Expenditure Survey, 1980

Data type	Sample size	Mean	Standard error: Simple random sampling	Standard error: Complex	Square root of design effect	Element standard deviation
All data	9,529	$51.61	$1.018	$1.914	1.88	$ 99.4
Real data only	4,688	52.27	1.430	2.936	2.05	97.9
Real data:						
Not donor	929	48.53	2.117	3.935	1.86	64.5
Donor once	2,798	55.76	1.982	3.386	1.71	104.8
Donor twice	841	49.37	3.579	4.879	1.36	103.8
Donor 3 or more times	120	28.97	7.987	11.64	1.46	87.6
Imputed data	4,841	50.98	1.447	2.323	1.60	100.7
Real data (adjusted weights)	4,688	51.80	1.470	3.000	2.04	100.7

NOTE: Estimates in this table are presented for illustrative purposes. Calculations were made prior to modifications described in the section on public use data files.

Table 16

Rate of imputation, sample size, mean, and standard errors for total charge for outpatient department visits for 4 family income groups and 3 estimation methods: National Medical Care Utilization and Expenditure Survey, 1980

Family income group and estimation method	Rate of imputation	Sample size	Mean	Standard error: Simple random sampling	Standard error: Complex
Less than $5,000:					
All data	0.76	1,403	$48.21	$2.132	$3.325
Real data	...	331	40.13	4.086	5.850
Real data (adjusted weights)	...	331	38.80	3.795	4.859
$5,000–$11,999:					
All data	0.58	2,389	50.34	1.849	5.111
Real data	...	11,015	48.30	2.710	6.272
Real data (adjusted weights)	...	1,015	46.99	2.756	6.313
$12,000–$34,999:					
All data	0.43	4,390	51.85	1.623	2.444
Real data	...	2,518	52.95	2.039	3.594
Real data (adjusted weights)	...	2,518	51.46	2.038	3.487
$35,000 or more:					
All data	0.39	1,347	56.90	2.889	4.182
Real data	...	824	60.48	3.673	5.998
Real data (adjusted weights)	...	824	64.82	4.176	6.817

Table 17

Values of *roh* and $\hat{s}^2$ for standard error formula for estimated means: National Medical Care Utilization and Expenditure Survey, 1980

Estimator	*roh*[1]	$\hat{s}^2$
Mean charge per person		
All charges:		
Ambulatory visits	0.029644	2.4952×10^9
Hospital stays	0.029644	6.1652×10^{10}
Physician visits	0.029644	6.1914×10^8
Total	0.029644	7.2407×10^{10}
Emergency room visits	0.029644	9.9816×10^7
Prescribed medications	0.029644	9.6458×10^7
Hospital outpatient visits	0.031367	7.6646×10^8
Independent provider visits	0.031367	2.6559×10^7
Hospital outpatient visits (nonphysician provider)	0.031367	4.2419×10^8
Physician visit (nonphysician provider)	0.031367	5.3375×10^7
Dental and other medical expenses	0.031367	8.8305×10^7
Charges paid out of pocket:		
Ambulatory visits	0.029644	2.4323×10^8
Hospital stays	0.029644	2.4068×10^9
Physician visits	0.029644	1.0745×10^8
Total	0.029644	3.5873×10^9
Emergency room visits	0.029644	1.0038×10^7
Prescribed medications	0.029644	4.5416×10^7
Hospital outpatient visits	0.031367	8.6571×10^6
Independent provider visits	0.031367	2.4996×10^8
Hospital outpatient visits (nonphysician provider)	0.031367	2.5341×10^7
Physician visit (nonphysician provider)	0.031367	6.7847×10^8
Dental and other medical expenses	0.031367	3.8943×10^8

See footnote at end of table.

Table 17—Con.

Values of *roh* and $\hat{s}^2$ for standard error formula for estimated means: National Medical Care Utilization and Expenditure Survey, 1980

Estimator	*roh*[1]	$\hat{s}^2$
Mean charge per user		
All charges:		
Ambulatory visits	0.043633	3.0423×10^9
Hospital stays	0.043633	3.0044×10^{11}
Physician visits	0.043633	1.1955×10^9
Total	0.043633	8.7587×10^{10}
Emergency room visits	0.043633	3.3067×10^8
Prescribed medications	0.043633	1.2535×10^8
Charges paid out of pocket:		
Ambulatory visits	0.043633	2.9046×10^8
Hospital stays	0.043633	1.6296×10^{10}
Physician visits	0.043633	1.5871×10^8
Total	0.043633	5.3877×10^9
Emergency room visits	0.043633	7.5825×10^7
Prescribed medications	0.043633	6.2806×10^7
Mean charge per visit		
All charges:		
Ambulatory visits	0.018777	3.7690×10^7
Hospital stays	0.018777	8.4926×10^{11}
Physician visits	0.018777	2.4686×10^7
Emergency room visits	0.018777	9.7896×10^8
Prescribed medications	0.018777	6.7348×10^5
Charges paid out of pocket:		
Ambulatory visits	0.018777	8.8152×10^6
Hospital stays	0.018777	9.4998×10^{10}
Physician visits	0.018777	9.2576×10^6
Emergency room visits	0.018777	1.1109×10^8
Prescribed medications	0.018777	7.8309×10^5
Mean visits per user		
Ambulatory visits	0.048246	1.4117×10^6
Hospital stays	0.048246	4.3009×10^3
Physician visits	0.048246	4.4788×10^5
Emergency room visits	0.048246	7.9937×10^3
Prescribed medications	0.048246	1.3402×10^6
Mean visits per person		
Ambulatory visits	0.048246	1.6398×10^6
Hospital stays	0.048246	1.0029×10^4
Physician visits	0.048246	5.5650×10^5
Emergency room visits	0.048246	1.6024×10^4
Prescribed medications	0.048246	1.6651×10^6
Mean percent of charges paid out pocket		
Ambulatory visits	0.051674	2.3071×10^3
Hospital stays	0.011724	1.7959×10^2
Prescribed medications	0.056569	2.7935×10^3
Dental and other medical expenses	0.053301	2.6150×10^3
Other means		
Mean length of hospital stay	0.013098	8.5018×10^5
Mean bed days	0.023772	7.6885×10^6
Mean work-loss days	0.026868	5.2013×10^6
Mean restricted-activity days	0.058349	3.4354×10^7
Mean functional limitation score	0.050066	4.9489×10^4
Mean number of surgical procedures	0.0	1.4628×10^8

[1]Synthetic measure of intraclass homogeneity.

Table 18

Estimated logistic regression coefficients and odds ratios for use of hospital care for all persons, by selected characteristics and indicators: United States, 1980

Independent variable[1]	Regression coefficient: Estimate	Regression coefficient: Adjusted standard error	Ratio of coefficient to standard error	Odds ratio: Estimate	Odds ratio: 95-percent confidence interval: Lower limit	Odds ratio: 95-percent confidence interval: Upper limit
Constant	− 2.8445	...	...	...	...	...
Characteristic						
Sex:						
(Male)[2]	...	...	...	...	...	...
Female	0.0766	0.0529	1.4493	1.0796	0.9734	1.1974
Race:						
White and other	0.0984	0.1057	0.9313	1.1034	0.8970	1.3573
(Black)[2]	...	...	...	...	...	...
Perceived health status:						
Excellent	− 0.2296	0.0567	− 4.0476	0.7949	0.7112	0.8883
(Good)[2]	...	...	...	...	...	...
Fair	0.9616	0.1114	8.6335	2.6159	2.1029	3.2541
Poor	1.5169	0.2040	7.4350	4.5581	3.0557	6.7990
Age:						
Under 35 years	0.1510	0.0609	2.4792	1.1630	1.0321	1.3105
(35–54 years)[2]	...	...	...	...	...	...
55–74 years	0.1051	0.0789	1.3319	1.1108	0.9516	1.2966
75 years and over	0.5187	0.1379	3.7606	1.6798	1.2819	2.2013
Poverty level:[3]						
Less than 2.00	0.1483	0.0600	2.4716	1.1599	1.0312	1.3046
2.00–4.99	− 0.0728	0.0441	− 1.6502	0.9298	0.8528	1.0138
5.00–6.99	− 0.1191	0.0754	− 1.5793	0.8877	0.7657	1.0291
7.00 or more	...	...	...	...	...	...
Third-party coverage:						
Multiple public	1.1897	0.1886	6.3066	3.2861	2.2704	4.7562
Single public	0.6484	0.1173	5.5286	1.9125	1.5197	2.4067
Private and public	0.7989	0.1154	6.9209	2.2231	1.7730	2.7875
Private only	0.3468	0.0788	4.4020	1.4145	1.2121	1.6507
(None or other)[2]	...	...	...	...	...	...
Region of residence:						
(Northeast)[2]	...	...	...	...	...	...
North Central	0.2066	0.1031	2.0045	1.2295	1.0046	1.5047
South	0.1186	0.0866	1.3703	1.1259	0.9502	1.3341
West	− 0.0944	0.0991	− 0.9523	0.9099	0.7492	1.1050
Usual source of care:						
Yes	− 0.0608	0.0629	− 0.9668	0.9410	0.8319	1.0645
(No)[2]	...	...	...	...	...	...
Indicator						
Healthy females 17–44 years:						
Yes	0.4387	0.0813	5.3970	1.5507	1.3223	1.8185
(No)[2]	...	...	...	...	...	...
Poor persons with poor health:						
Yes	− 0.5032	0.1596	− 3.1529	0.6046	0.4422	0.8266
(No)[2]	...	...	...	...	...	...

[1]Proportion reduction of error = 5.1 percent.
[2]Reference cell.
[3]Analysis of variance parametrization.

Table 19

Sample frequency distribution, by threshold level and subpopulation

Subpopulation	Threshold level 1	2	. . .	t	$t+1$	Total
1	n_{11}	n_{12}	. . .	n_{1t}	$n_{1,t+1}$	n_1
2	n_{21}	n_{22}	. . .	n_{2t}	$n_{2,t+1}$	n_2
.	.	.	. . .	.	.	.
.	.	.	. . .	.	.	.
.	.	.	. . .	.	.	.
s	n_{s1}	n_{s2}	. . .	n_{st}	$n_{s,t+1}$	n_s

Table 20

Proportion of person years for persons 18–64 years of age, by physician visit charge threshold and selected types of health care coverage: United States, 1980

Type of coverage	Physician visit charge threshold: None	$50	$100	$200	More than $200	Sample size
Private	0.336	0.260	0.149	0.130	0.125	6,579
Public	0.321	0.212	0.140	0.135	0.192	1,429
Part year	0.399	0.265	0.144	0.103	0.089	978
None	0.555	0.236	0.089	0.068	0.052	876

Table 21

Cumulative proportion and cumulative logit of person years for persons 18–64 years of age, by physician visit charge threshold and selected types of health care coverage: United States, 1980

Type of coverage	Physician visit charge threshold: None	$50	$100	$200
		Cumulative proportion		
Private	0.336	0.596	0.745	0.875
Public	0.321	0.533	0.673	0.808
Part year	0.399	0.664	0.808	0.911
None	0.555	0.791	0.880	0.948
		Cumulative logit		
Private	− 0.682	0.387	1.071	1.947
Public	− 0.747	0.134	0.721	1.436
Part year	− 0.409	0.680	1.438	2.323
None	0.223	1.334	1.989	2.908

Table 22

Estimates of parameters and their standard errors under the saturated model X_S: United States, 1980

Description	Parameter	Estimate	Standard error
Mean	μ	0.223	0.057
Intercept:			
Private	τ_1	− 0.905	0.069
Public	τ_2	− 0.970	0.095
Part year	τ_3	− 0.632	0.093
Slope 1	β_1	1.111	0.069
Slope 2 increment	β_2	0.655	0.076
Slope 3 increment	β_3	0.919	0.130
Slope 1 differential:			
Private	$\gamma_{1(1)}$	− 0.041	0.077
Public	$\gamma_{1(2)}$	− 0.230	0.087
Part year	$\gamma_{1(3)}$	− 0.022	0.102
Slope 2 differential:			
Private	$\gamma_{2(1)}$	− 0.013	0.118
Public	$\gamma_{2(2)}$	− 0.298	0.120
Part year	$\gamma_{2(3)}$	0.082	0.132
Slope 3 differential:			
Private	$\gamma_{3(1)}$	− 0.056	0.163
Public	$\gamma_{3(2)}$	− 0.501	0.173
Part year	$\gamma_{3(3)}$	− 0.048	0.199

Table 23

Analysis of variance for cumulative logit analysis for persons 18–64 years of age under 2 sample design options for physician visit charges, by selected types of health care coverage: United States, 1980

Type of coverage	Degrees of freedom	Simple random sampling, weighted		Complex sampling		$Q_{C(s)}/Q_{C(c)}$
		$Q_{C(s)}$	p	$Q_{C(c)}$	p	
Total	15	6,905.48	<0.01	6,372.77	<0.01	1.08
Health care coverage	3	170.11	<0.01	198.02	<0.01	0.86
Private	1	154.31	<0.01	170.09	<0.01	0.91
Public	1	120.07	<0.01	103.60	<0.01	1.16
Part year	1	44.93	<0.01	46.69	<0.01	0.96
Parallelism	9	35.67	<0.01	34.89	<0.01	1.02
Slope 1:						
Private	1	0.30	0.58	0.29	0.59	1.06
Public	1	7.34	<0.01	6.96	<0.01	1.05
Part year	1	0.05	0.82	0.05	0.83	1.19
Slope 2:						
Private	1	0.02	0.90	0.01	0.91	1.33
Public	1	6.89	<0.01	6.14	0.01	1.12
Part year	1	0.41	0.52	0.38	0.54	1.08
Slope 3:						
Private	1	0.13	0.72	0.12	0.73	1.10
Public	1	9.09	<0.01	8.42	<0.01	1.08
Part year	1	0.06	0.80	0.06	0.81	1.11

Table 24

Estimates of parameters and their standard errors under the reduced model X_R: United States, 1980

Description	Parameter	Estimate	Standard error
Mean	μ	0.238	0.049
Intercept:			
Private	τ_1	− 0.921	0.060
Public	τ_2	− 0.979	0.089
Part year	τ_3	− 0.608	0.086
Slope 1	β_1	1.075	0.023
Slope 2 increment	β_2	0.692	0.021
Slope 3 increment	β_3	0.877	0.024
Slope 1 differential: Public	$\gamma_{1(2)}$	− 0.200	0.057
Slope 2 differential: Public	$\gamma_{2(2)}$	− 0.309	0.062
Slope 3 differential: Public	$\gamma_{3(2)}$	− 0.454	0.085

Table 25

Predicted cumulative proportion and predicted cumulative logit of person years for persons 18–64 years of age, by physician visit charge threshold and selected types of health care coverage under reduced model X_R: United States, 1980

Type of health care coverage	Physician visit charge threshold			
	None	$50	$100	$200
	Cumulative proportion			
Private	0.336	0.597	0.747	0.877
Public	0.323	0.533	0.672	0.810
Part year	0.409	0.669	0.802	0.907
None	0.559	0.788	0.881	0.947
	Cumulative logit			
Private	− 0.683	0.392	1.084	1.961
Public	− 0.740	0.134	0.717	1.449
Part year	− 0.370	0.704	1.397	2.274
None	0.238	1.313	2.005	2.882

Table 26

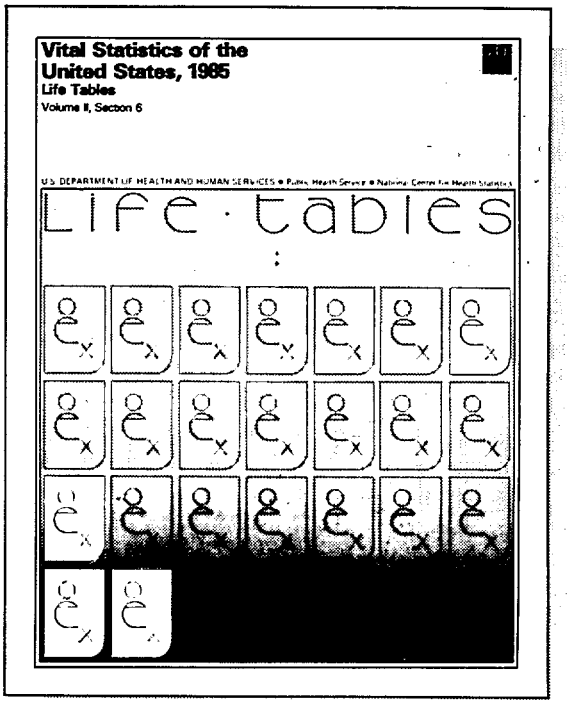
Predicted odds ratios[1] under reduced model X_R for odds of incurring charges greater than the given threshold value: United States, 1980

Comparison	Physician visit charge threshold			
	None	$50	$100	$200
Private: No coverage	2.51	2.51	2.51	2.51
Public: No coverage	2.66	3.25	3.62	4.19
Part year: No coverage	1.84	1.84	1.84	1.84

[1] $exp(\hat{L}_{4j} - \hat{L}_{ij})$, for $i = 1,2,3$ and $j = 1,2,3,4$.

Available Only from the Government Printing Office...

Certified Life Tables



The **NATIONAL CENTER FOR HEALTH STATISTICS** has just released for sale the 1985 life expectancy tables. The life tables, certified as official documents, are utilized by lawyers as legal evidence, insurance analysts, actuaries, pension planners, demographers, and other researchers. The publication includes current and trend data by age, sex, and race from 1900 to 1985, survivors to specified ages, and median length of life. The booklet is available only from the Government Printing Office at a cost of $1.50 per copy. Use the handy order form below to purchase your copy today.

nchs

Publication Order Form

Order processing code: * 6247

Mail to: Superintendent of Documents
Government Printing Office
Washington, D.C. 20402

☐ **YES,** please send me _____ copies of **Vital Statistics of the United States, 1985, Life Tables**
GPO Stock Number-017-022-01027-5 Price $1.50

The total cost of my order is $_____. Foreign orders please add an additional 25%.
Prices include regular domestic postage and handling and are good through December, 1988. After that date, please call Order and Information Desk at (202) 783-3238 to verify prices.

Please Type or Print

_____
(Company or personal name)

_____
(Additional address/attention line)

_____
(Street address)

_____
(City, State, ZIP Code)

( )
(Daytime phone including area code)

Please choose method of payment:

☐ Check payable to the Superintendent of Documents

☐ GPO Deposit Account ☐☐☐☐☐☐☐–☐

☐ VISA, MasterCard or Choice Account

☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐

_____
(Signature)

_____
(Credit card expiration date)

Thank you for your order! Thank you for your order! Thank you for your order! Thank you for your order! Thank you for your order!

Our warehouses here at the Government Printing Office contain more than 16,000 different Government publications. Now we've put together a catalog of nearly 1,000 of the most popular books in our inventory. Books like *Infant Care, National Park Guide and Map, The Space Shuttle at Work, Federal Benefits for Veterans and Dependents, Merchandising Your Job Talents*, and *The Back-Yard Mechanic.* Books on subjects ranging from agriculture, business, children, and diet to science, space exploration, transportation, and vacations. Find out what the Government's books are all about. For your **free copy** of our new bestseller catalog, write—

New Catalog
Post Office Box 37000
Washington, D.C. 20013

937 Bestsellers

Department of Health and Human Services
Otis R. Bowen, M.D., Secretary

Health Care Financing Administration
William L. Roper, M.D., Administrator

Office of Research and Demonstrations
Judith B. Willis, Director

Office of Research
J. Michael Fitzmaurice, Ph.D., Director

Division of Program Studies
Carl Josephson, Director

Survey Studies Branch
Herbert A. Silverman, Ph.D., Chief

Public Health Service
Robert E. Windom, M.D.,
Assistant Secretary for Health

National Center for Health Statistics
Manning Feinleib, M.D., Dr.P.H., Director

Office of Vital and Health Statistics Systems
Peter L. Hurley, Associate Director

Division of Health Interview Statistics
Owen T. Thornberry, Jr., Ph.D., Director

Utilization and Expenditure Statistics Branch
Robert A. Wright, Chief