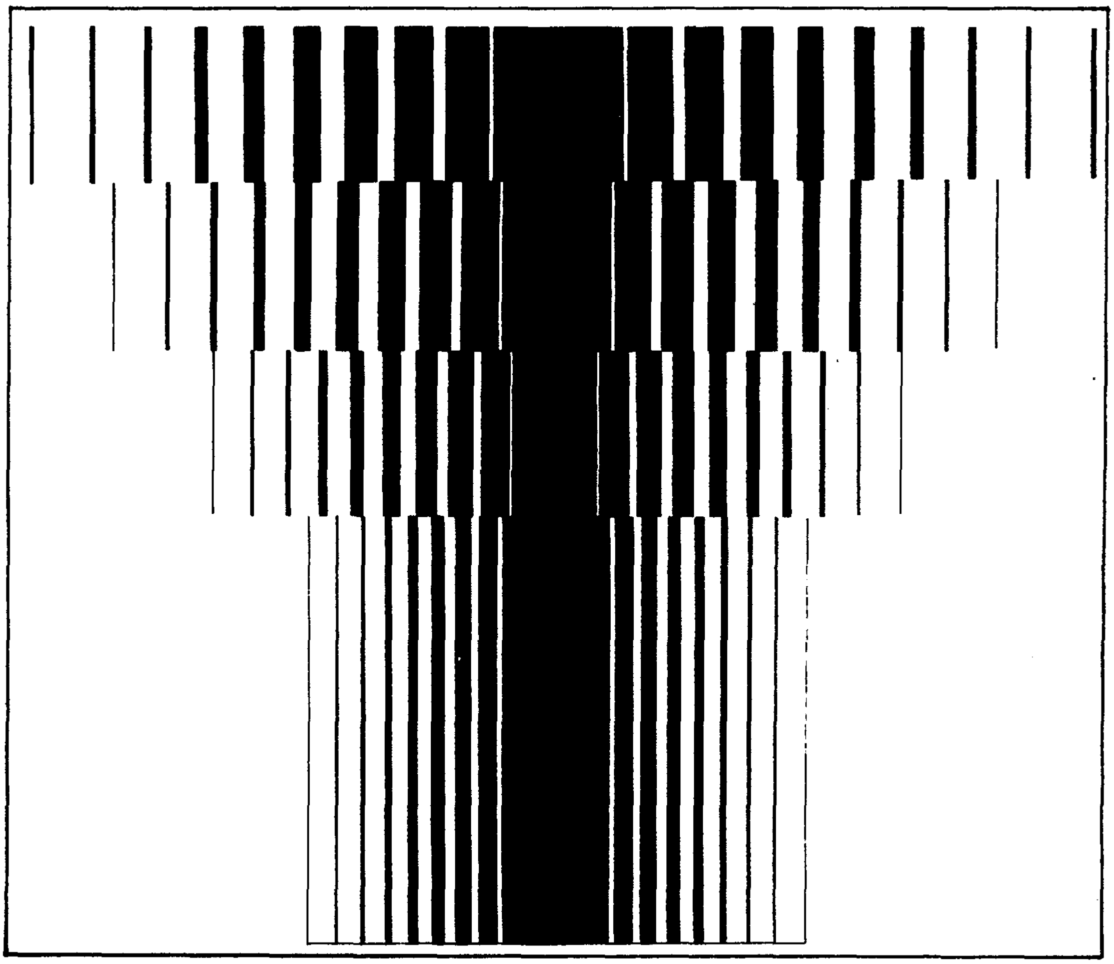
High-Volume and Low-Volume Users of Health Services United States, 1980

Series C, Analytical Report No. 2



U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES

Published by
Public Health Service
National Center for Health Statistics

November 1985

Suggested Citation

Berki, S. E., Lepkowski, J. N., 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, Analytical Report No. 2. DHHS Pub. No. 86-20402. National Center for Health Statistics, Public Health Service. Washington. U.S. Government Printing Office, Nov. 1985.

Library of Congress Cataloging in Publication Data

High-volume and low-volume users of health services.
(Series C, Analytical report; no. 2) (DHHS publication no. 86-20402)
Written by: Sylvester E. Berki and others. Bibliography: p.
1. Medical Care—United States—Utilization—Statistics.
2. Public health—United States—Statistics. 3. Health surveys—United States. I. Berki, Sylvester E. II. National Center for Health Statistics (U.S.) III. Series: National medical care utilization and expenditure survey. Series C, Analytical report; no. 2. IV. Series: DHHS publication; no. 86-20402. [DNLM: 1. Health Services—utilization—United States. 2. Health Surveys—United States. W 84 AA1 H6]
RA410.7.H54 1985 362.1'0973'021 85-18844
ISBN 0-8406-0324-X

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 primarily responsible for data processing.

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

Foreword

This report is a contribution to the literature on health services research and health economics. Specifically, it was designed to investigate differences in the use of health care. There is a great deal of variation in the use of health care among members of the U.S. population with consequent wide variation in the expenditures for that care. This report should lead to improved understanding of the variation and provide data for those making public policy.

The research, analysis, and publication was conducted under a contract (No. 282–83–2119) between the Division of Health Interview Statistics, the National Center for Health Statistics division responsible for the National Medical Care Utilization and Expenditure Survey, and the School of Public Health at the University of Michigan.

Mary Grace Kovar, Dr.P.H.,
Special Assistant for Data Policy
and Analysis, Office of Interview
and Examination Statistics Program

Contents

List of Text Tables

List of Figures

Symbols

---	Data not available
...	Category not applicable
-	Quantity zero
0.0	Quantity more than zero but less than 0.05
*	Test statistic is significant at 0.05 level
**	Test statistic is significant at 0.01 level

High-Volume and Low-Volume Users of Health Services: United States, 1980

by Sylvester E. Berki, James N. Lepkowski, Leon Wyszewianski, J. Richard Landis, M. Lou Magilavy, Catherine G. McLaughlin, and Hillary A. Murt, University of Michigan

Executive Summary

Data from the National Medical Care Utilization and Expenditure Survey of 1980 are used to examine the characteristics of high-volume users of health care services, contrasting them with low-volume users and those who used no services at all. The three major types of medical care services examined are hospital inpatient care, ambulatory visits, and prescribed medications. Low users were defined, respectively, as those who during the year had either one or two hospital days, one nondental visit to a physician or nonphysician, and one prescribed medicine acquisition. High users were those with, respectively, 17 or more hospital days, 20 or more visits, and 25 or more prescribed medicine acquisitions.

A very small percent of the U.S. civilian noninstitutionalized population and of those who used services at all during the year consume a large percent of services in each of the three service types. High users of inpatient hospital care constitute 1.7 percent of the civilian noninstitutionalized population and 15 percent of persons hospitalized during the year, yet they used 54.4 percent of all hospital days used by the reference population. High users of ambulatory services constitute 4.5 percent of the reference population and only 5.7 percent of all users of ambulatory services, yet they accounted for 32.3 percent of all ambulatory visits. For prescribed medications, only 3.7 percent of the civilian noninstitutionalized population are high users, comprising 5.9 percent of all users, but they account for 32.9 percent of all prescription acquisitions. At the other extreme, low users of ambulatory care visits represent 17 percent of the reference population, and 21 percent of all users of such care, but only 3.3 percent of all visits.

High users share certain characteristics. They are more likely than low users to be older and poorer, to have poorer health status and more medical conditions, and are more likely to have functional limitations.

Both univariate and multivariable analyses show that the most important distinguishing characteristics of high users of any of the three medical services are poor health status, severe functional limitations, and the presence of multiple medical conditions—most importantly cancer, cardiac disorders, musculoskeletal diseases, respiratory diseases, and injuries and poisonings.

Almost all high-volume users of every category of service (88 percent for hospital days, 89 percent for ambulatory visits, and 94 percent for prescribed medications) had at least three different diagnostic conditions reported during the year.

The likelihood of being a high user of hospital inpatient services, the most expensive component of health care, is increased not by the comprehensiveness of health care coverage but by the need for health care, measured both by reported health status and by the presence of disabling or life threatening conditions. The evidence is persuasive that high-volume hospital use is associated with severe illness, functional limitation, and death. Of persons with no functional limitation only 0.5 percent were high hospital inpatient care users, a percent that increased to 23 percent for those most severely limited in their activities, and to 54 percent of those who died during the year.

Among the high-volume users of hospital inpatient services, 58 percent had 6 or more separate diagnostic conditions, and 53 percent had their health status reported to be fair or poor. This is in sharp contrast to persons who did not experience any hospital episodes during the year, among whom only 11 percent had their health status reported to be fair or poor.

The results strongly suggest that demographic variables such as race, sex, and even age are related to hospital use only to the extent that they are associated with other factors such as reported health status, functional limitations, and health care coverage.

High users of ambulatory services tend to differ from high users in the two other categories. Although they also tend to have the characteristics of poor reported health, severe illness, and the greater likelihood of at least some functional limitation found among high users of hospital inpatient care and prescription medications, they tend to do so to a lesser extent. Therefore, efforts to reduce high use in one category of service are not likely to yield comparable results in other categories.

Across all three service categories, low users were found to be similar to nonusers in terms of age, perceived health status, and presence of activity limitation. However, for hospital days low users were found to be different from nonusers with respect to health care coverage, suggesting that health care coverage continues to be important in determining access to hospitalizations.

Although specific patterns vary for each service category, in general low-volume users were found to have significantly different characteristics from high-volume users, especially on need-related variables: They had higher perceived health status, less functional limitation, and fewer medical conditions.

The findings on the characteristics of high users of medical care services are sobering: 54 percent of all hospital days, 32 percent of all ambulatory visits, and 33 percent of all prescribed medications reported for the U.S. civilian noninstitutionalized population in 1980 were consumed by a very small percent of that population. These high users of medical care services were predominantly sick, functionally limited in their activities, or dying. In an effort to contain costs, approaches to reduce the incidence of conditions that lead to high use of medical care resources need to be considered. Alternative treatment modalities and institutional structures to reduce to costs of management of conditions that cannot be prevented should also be explored.

NOTE: We are grateful for the support we received during all stages of the preparation of this document both from our colleagues at The University of Michigan and from the staff of the National Center of Health Statistics. At The University of Michigan, Sharon Stehouwer contributed greatly to the initial analyses of NMCUES data and to identification and correction of several problems encountered in the data base. Kenneth E. Guire was responsible for the overall preparation and correction of data files on which this report is based. P. Ellen Parsons provided much appreciated assistance on many aspects of the preparation of this report. Valuable consultation on matters related to use of medications was given by Dr. Duane Kirking. Quality secretarial support in the preparation of the many tables included in the report was provided by Katherine Metcalf. At the University's Institute for Social Research, Nan Collier developed software for calculating sampling errors, and Judy Connors performed many of the analyses for generating sampling errors for national estimates.

We also received continual support and guidance from the National Center for Health Statistics and in particular from our project officer, Dr. Mary Grace Kovar, Special Assistant for Data Policy and Analysis. We are indebted to Dr. Robert J. Casady, Chief of the Statistical Methods Staff, for writing the major section in Appendix I, describing the survey design and estimation methodology for NMCUES. Robert Wright and Michelle Chyba quickly responded to all our queries about problems we encountered with the data. Editors in the Publications Branch provided valuable assistance during all stages of the report, especially preparation of the detailed tables.

Introduction

Overview

This report examines the characteristics of high-volume users of health care services, contrasting them with low-volume users and those who used no services at all. The data for this study are from the public use files for the National Househould Survey component of the National Medical Care Utilization and Expenditure Survey (NMCUES). Interviews for the survey were conducted between early 1980 and mid-1981. The survey and public use files are described further in the section on "Sources of Data." The next subsection briefly summarizes relevant past work, followed by a statement of the objectives and the scope of this report.

Background

The small percent of the population that in any given year makes extensive use of health services has been of continuing interest to policymakers and analysts, because this group accounts for a disproportionate share of total health expenditures and includes individuals who are at high risk for financially catastrophic health care expenses. Yet to date only modest efforts have been made to estimate the number and other characteristics of high-volume users in the United States. The most ambitious of these have focused on estimating, for high-cost illnesses—the incidence, cost, and contributions of different payers. Although based on a variety of sources of data, such efforts rely most heavily on the two major national surveys of health care services use, the National Health Interview Survey (NHIS) and the periodic surveys conducted by the Center for Health Administration Studies and National Opinion Research Center at the University of Chicago (Trapnell, 1977; Congressional Budget Office, 1977; Birnbaum, 1978a and 1978b). Other studies, more limited in scope, focus on subgroups defined by the source of data on which the study is based, such as third-party payers (Forthofer et al., 1982; Congressional Budget Office, 1982; Anderson and Knickman, 1984) or hospital records (Schroeder et al., 1979; Zook and Moore, 1980; and Kobrinski and Matteson, 1981).

Additional information on high-volume users is available from studies not directly concerned with this population but which nevertheless include patient groups who tend to be high-volume users, including those with cancer (Cancer Care, 1973; Eldred et al., 1977), stroke (Weinfield, 1981), spinal cord injuries (Webb et al., 1977; Anderson et al., 1980), those in intensive care units (Cullen et al., 1976; Budetti et al., 1981; Detsky et al., 1981), and the terminally ill (Bloom and Kissick, 1980; Gibbs and Newman, 1982; Lubitz and Prihoda, 1984; McCall, 1984; Scitovsky, 1984).

This report is based on recent data from a national survey and provides national estimates of the number and characteristics of high-volume users. It also contrasts high-volume users with low-volume users and those who used no services at all.

Objectives of the Report

Relying on the database generated by NMCUES, this report investigates the characteristics of high users of health services and compares such high users with other categories of users, in particular low users. The report relies on a comprehensive set of univariate analyses of the major categories of use: Hospital inpatient care, ambulatory care, and prescribed medications. Inpatient hospital use is further analyzed by the application of multivariable regression methods, designed as a major step in modeling and as an indication of the potential fruitfulness of applying these approaches to the analysis of other major categories of service.

This report presents national estimates of the number and characteristics of high users of health services within the civilian noninstitutionalized population of the United States during 1980 for each of three categories of service: Ambulatory care, inpatient hospital care, and prescription medications. The report also illuminates the characteristics of high-volume users by contrasting high-users with low-users and nonusers of services.

Detailed national estimates of major policy-relevant attributes of high-volume and low-volume users are presented:

- Number and percent of the total population each group represents.
- Proportions of total volume of services each group accounted for.
- Average charges per person and use levels per person.
- Demographic characteristics.
- Sources of payment for the care.
- Medical characteristics.

The multivariable analysis of inpatient hospital use presents the independent effects of variables hypothesized to be related to high-volume use, holding statistically constant the effects of other variables.

As indicated in the section on "Limitations of the Data," characteristics of the information available on charges made it necessary to limit both types of analysis to use of services rather than incorporating both use of service and charges.

Sources of Data

The National Medical Care Utilization and Expenditure Survey (NMCUES)

Between February 1980 and April 1981, data on 17,123 persons in 6,798 families were collected at approximately 3-month intervals for a total of five interviews: Two personal interviews followed by two telephone interviews, and a final personal interview. At the conclusion of the first interview, survey participants were provided with a specially designed calendar-diary for recording data about medical events and costs in preparation for subsequent rounds of interviewing. Prior to each interview after the first, interviewees were sent a summary sheet with all medical events and costs reports in previous interviews. Specific details concerning the sample design and data collection are outlined in Appendix I.

This report is based on data in the public use tapes, which consists of six files: The person, medical visit, dental visit, hospital stay, prescribed medicines and other medical expenses, and condition files. The person file has one record for each of the 17,123 responding eligible persons with data describing the person's demographic characteristics, health care coverage, employment, income, and usual source of care; numbers of visits, hospitalizations, and other medical events reported for 1980; total charges for each category of care; and limitations and disabilities, including identification of conditions. Data from the other five files, which have more detailed information about events summarized in the person file, can be linked to records in the person file through a unique identification number assigned to each person.

Limitations of the Data

This section describes some of the limitations of NMCUES data that determined the scope of this report and that need to be taken into account in interpreting its results.

Reporting of Data on Charges

Analyses here reported focus on the *use* of services rather than on charges or costs. The emphasis is on comparisons of different levels of use, particularly low versus high. The principal reason for emphasizing use of services is that NMCUES data contain a number of improbably low values of total annual charges for ambulatory visits, prescribed medicines, and hospital stays. When total annual charges for *all services* were examined for those with charges greater than 0, the lowest 5 percent were between $1.00 and $21.00. It is likely that for many of these cases the reported data do not reflect the actual charges for the services received. These reported charges are more likely to be out-of-pocket expenses incurred, in particular the deductibles and copayments paid by those who have insurance, or expenses not covered by Medicaid. Inclusion of such cases in the low-charge category distorts the characteristics of that class of individuals, especially since the number of such misreported cases appears to be quite large. Similarly, certain high charge hospitalizations may also be excluded.

The problems posed by out-of-pocket expenses being misreported as total charges are compounded by the high rate of nonresponse on questions about charges, resulting in imputed charges for about 36 percent of all hospital admissions and 26 percent of ambulatory visits. This complicates the task of characterizing certain categories of cases: Because imputations of charges were based on such variables as age and type of provider but not on diagnosis, incongruous pairings of charges with diagnoses may result.

The combination of these biases from incorrect responses and nonresponses on charges indicated the desirability of focusing these analyses on high and low users defined in terms of volume of services and not in terms of charges.

Exclusion of the Institutionalized Population

Because the institutionalized population was not included in the NMCUES sample frame, all data presented here on high users of care apply to the U.S. civilian noninstitutionalized population only. This is consistent with many previous studies that had a similar focus. It must be recognized, however, that most, if not all, of those in health care institutions are high users. For this reason the picture presented here is not complete.

Reporting of Diagnoses and Medical Conditions

In NMCUES a person was noted as having a particular medical condition only when the condition caused some type of restriction of activity or resulted in an ambulatory visit, hospitalization, purchase of a prescribed medicine, or other encounter with the medical care system. Thus conditions that did not cause any restriction in activity or days in bed, or did not lead to a visit to a doctor, for example, were not recorded. For conditions that were reported, the diagnostic accuracy depends both upon the information provided by the health care provider and on the respondent's ability and willingness to accurately convey this information to the interviewer.

For each medical encounter recorded in the survey, respondents were allowed to report up to four medical conditions. The public use files show that (a) approximately 10 percent of medical visits had two or more conditions recorded, (b) multiple conditions are listed for about 12 percent of all hospital stays, and (c) 4 percent of the prescribed medication records have two or more conditions recorded. The NMCUES survey instrument does not designate which of the underlying conditions is the "principal diagnosis" or the primary reason for each medical encounter. Because a principal diagnosis is not identified for each medical event, it is impossible to determine to which condition health services use should be attributed when multiple conditions are reported. This poses a problem in developing accurate estimates of the extent to which a given illness is associated with given levels of utilization of health care resources.

Operationalization of Variables

Each of the variables included in this study is briefly defined in Appendix III. The three key service categories on which this analysis focuses are defined as follows:

- *Hospital care use* is measured in terms of number of days spent in the hospital. It should be noted that persons who were admitted to and discharged from a hospital on the same date, i.e., they did not stay overnight in the hospital, are counted as having zero hospital days.
- *Ambulatory care use* is defined as the number of all nondental visits made to physicians as well as to nonphysicians, whether in a private office, a hospital outpatient department, or emergency room.
- *Prescribed medicine use* is measured by the number of acquisitions, that is, the total number of times a person had a prescription filled, regardless of whether it was an initial filling or a refill of a prescription.

The other variables used in the analysis refer to demographic and other characteristics, such as age, sex, income, education, health care coverage, medical conditions, perceived health status, and functional disability. Their definitions, given in Appendix III, are consistent with those in other reports that are based on the same NMCUES public use files. The procedures for applying weights to all these variables to derive national estimates are described in Appendix V.

Definition of Low- and High-Volume Use

No commonly agreed upon definitions of low and high use of health services exist. Several categories of definitions can be identified in past studies, and within categories the thresholds specified vary as well. In many instances the definitions are in terms of costs. When defined in terms of an absolute dollar amount, the threshold for annual inpatient expenses has ranged, in studies done in the late 70's and early 80's, from $3,000 to $10,000 (see Birnbaum, 1978; Schroeder et al., 1979; Congressional Budget Office, 1982). Other definitions are based on the distribution of cases, focusing on the top 5 percent (e.g. Lubitz and Prihoda, 1982), or on a more elaborate statistical definition such as the specification of "outliers" under Medicare's Prospective Payment System.

In this study, four levels were defined for each category of use analyzed: Low-volume and high-volume as well as zero and intermediate use, to provide a fuller contrast for the low and high categories on which the study focuses. The following two criteria were specified *a priori* for the determination of cut points:

1. The number of cases included in the high and low categories must be sufficiently large to permit meaningful comparisons across characteristics of interest. Based on the analyses anticipated, 300 was the estimate of the minimum number needed in the high and the low categories for each service.
2. To the extent possible, the cut points should isolate the extremes of the distribution of use of services that account for disproportionate shares of the total, and are therefore often the subject of study. That includes, in particular, the top 5 percent, which tends to account for one-third to one-half of total volume, and the bottom 20 to 25 percent, which accounts for 3 to 5 percent of total volume.

The requirement that there be at least 300 cases in each category was found to be the binding constraint in most instances, except for the low-use categories of service for ambulatory care and prescribed medicine acquisitions, where even taking the lowest possible nonzero number, one visit and one acquisition, yielded several thousand cases in each instance.

Each threshold is discussed next, and the values and characteristics of thresholds are shown in Table A. (Percents shown in Table A are based on *unweighted* counts and therefore differ from similar percents given in Table 1–29, which are all based on weighted data.)

Table A
Thresholds of low and high use for each type of service: United States, 1980

Type of service and level of use	Threshold (T)	Number in the category (unweighted count)	Percent of all service users	Percent of total service volume
Hospital days				
Low	$0<T\leq2$	416	19.9	3.3
High	$T\geq17$	324	15.5	55.5
Ambulatory care visits				
Low	$T=1$	2,927	21.7	3.4
High	$T\geq20$	754	5.6	31.8
Prescribed medicine acquisitions				
Low	$T=1$	2,318	21.8	3.0
High	$T\geq25$	637	6.0	33.3

Hospital Days

Low use of hospital days was defined as including those who had either 1 or 2 days of hospitalization. This specifically excludes cases classified as having had one or more hospitalizations but zero nights in the hospital: An examination of such cases showed that they are likely to have involved either ambulatory surgery or other, mostly major, therapeutic and diagnostic procedures performed on an outpatient basis. Because they do not involve the use of an inpatient bed, it would not be appropriate to include them in the category of low-volume hospitalizations. There were only 151 cases with one hospital day; therefore, 2-day users had to be added to meet the requirement for an unweighted count of at least 300. The 416 cases with 1 or 2 hospital days represented nearly 20 percent of all users, but only 3.3 of total hospital days used by those in the sample.

High Use is represented by all cases with 17 or more hospital days. There were 324 such cases, and they accounted for 15 percent of all users of services. Ideally, no more than 7 or 8 percent of all users would have been captured to be consistent with similar studies in which the focus is on the top 5 percent who account for one-third of all the use. However, hospitalization is a relatively rare event, and even in a national sample of more than 17,000 persons, fewer than 2,100 had any hospitalization. So to meet the minimum of 300 cases, a cut point had to be selected that is lower than desirable.

Ambulatory Care Visits

Low use of ambulatory care visits was defined as one visit. Although that is obviously the lowest nonzero number that can be selected, it captures 2,927 cases, or 22 percent of all users, far more than the minimum 300 cases required.

High use was taken to be 20 or more visits during the year, which is four times the mean number of visits in the sample. The 754 cases that are captured with this definition reflect the familiar top 5 percent of users, which accounts for about one-third of total volume.

Prescribed Medicine Acquisitions

Low use for prescribed medicines was set at one acquisition, which, as was the case for ambulatory visits, is the lowest nonzero number that can be selected. That yielded a group of 2,318 low users who represented 22 percent of all users and 3 percent of total volume, which is very similar to the low-use category for ambulatory visits.

High use was defined as 25 or more acquisitions, yielding 637 cases, which, like the high-use category for ambulatory visits, also comes very close to the 5 percent of users who represented one-third of total volume.

Univariate Analysis

This section presents the descriptive analyses of zero-, low-, intermediate-, and high-volume groups within three major medical service categories: Inpatient, ambulatory, and prescription drugs. Because the objective is to describe high-volume users of care and to contrast them with low-volume users, and with individuals who did not use any services at all during 1980, the percents in all detailed tables (Tables 1–30) are relative to level of use categories. Thus, for example, they show the percent of high users of ambulatory care who were female, rather than the percent of females who were high users.

Distribution of Use in Each Service Category

A principal reason for exploring the distribution of use for each category of medical service is to determine the extent to which services are disproportionately consumed by groups of individuals in the civilian noninstitutionalized population.

The findings reflect disproportionate use of services within each of the three volume levels: Low, intermediate, and high (Table 1). High-volume users of inpatient hospital care represented 1.7 percent of the total civilian noninstitutionalized population, 15 percent of all individuals who used hospital inpatient services, but they accounted for 54.3 percent of all hospital days, and 45.2 percent of inpatient hospital charges during 1980. Further evidence of this disproportion is the very high percent of the total direct health care costs accounted for by this small group of individuals. The inpatient hospital charges of $39.4 billion generated by high-volume users represented 25.6 percent of the Nation's estimated total health care bill of $153.9 billion based on NMCUES data for the civilian noninstitutional population in 1980 (Parsons et al., 1985). At the other end of the continuum, low-volume users of hospital inpatient services (those with 1 or 2 hospital days) constituted 20.5 percent of all users. This low-volume user group comprised 2.4 percent of the total civilian noninstitutionalized population; it accounted for 3.4 percent of the total volume of hospital days used and generated 6.0 percent of total hospital inpatient charges for the population.

The disproportion in the relationship between the small percent of high-volume users of inpatient hospital services and the high percent of the volume of use is illustrated graphically with the Lorenz Curve. To obtain such a curve, users of inpatient care are arrayed by level of use from lowest to highest, and divided into fourths, fifths, deciles, or other fractions, and the percent of total use represented by each fraction is obtained. The Lorenz Curves, displayed in Figures 1–3, plot the cumulative percent of total volume of services accounted for by the cumulative percent of users. Figure 1 is the Lorenz Curve for users of inpatient hospital care. If each user had used the same number of days, the curve would be the 45-degree dotted line shown on the figure, and, for example, 50 percent of users would account for 50 percent of total hospital days. The greater the area between the 45-degree line and the curve, the greater the differences in use between the lowest and highest use categories, and therefore, the larger the share of total volume of use attributable to a small group of users. The Lorenz Curve for hospital use thus illustrates that high-volume users constituted only 15 percent of all users, but accounted for 54.3 percent of all hospital days used by the civilian noninstitutionalized population. At the other end of the continuum, the Lorenz Curve also illustrates that low-volume users represented a substantial fraction of all users, 20.5 percent,

Figure 1
Lorenz Curve of hospital days: United States, 1980

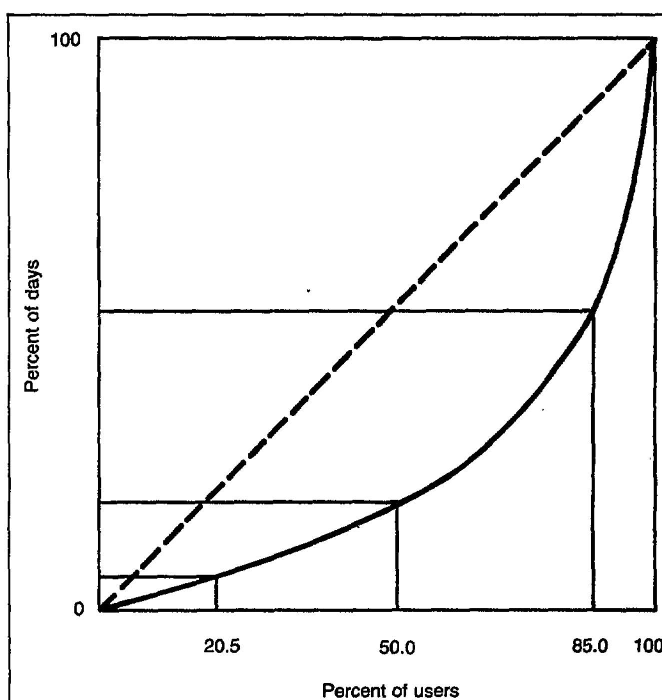


Figure 2
Lorenz Curve of ambulatory care visits: United States, 1980

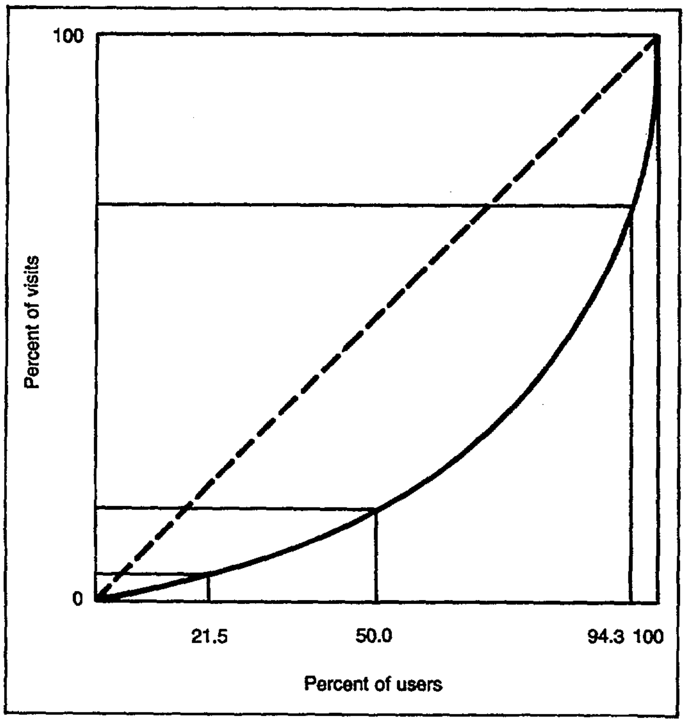
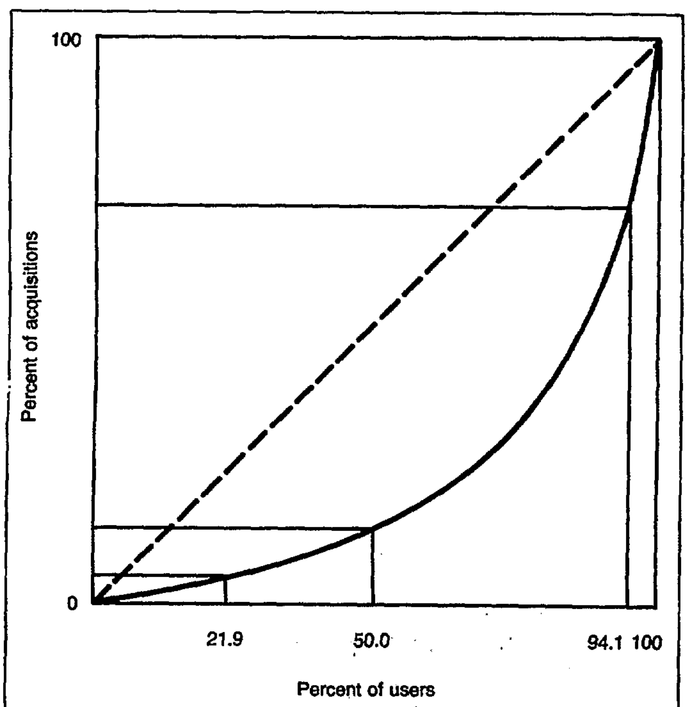


Figure 3
Lorenz Curve of prescribed medicine acquisitions: United States, 1980



but accounted for only 3.4 percent of the total volume of hospital days used by this population.

The disproportionate distribution of ambulatory visits is also striking. Of all ambulatory visits, 32.3 percent were generated by 4.5 percent of the civilian noninstitutionalized population, constituting only 5.7 percent of all users of ambulatory visits. But this group of high-volume users, making 37 visits on average during 1980, accounted for 28.7 percent of ambulatory-related charges (Table 1). Of all users, 21.5 percent made only one visit during 1980. This group of low-volume users represented 17.0 percent of this population but accounted for only 3.3 percent of all visits and 3.6 percent of all ambulatory-related charges (Figure 2).

The distributional pattern of use of prescribed medicine acquisitions is similar to ambulatory visits. Approximately 3.7 percent of the population was categorized as high-volume users of prescription medications, with an average of 41.3 acquisitions per year. They constituted only 5.9 percent of all individuals in this population who had at least one prescription filled in 1980, yet they accounted for 32.9 percent of all prescribed medicine acquisitions and 34.2 percent of all prescription-related charges. A relatively small percent (13.7 percent) of the civilian noninstitutionalized population had only one prescription filled during 1980. This group of low-volume users accounted for only 3.0 percent of all acquisitions and 2.6 percent of all prescription charges, as shown in Table 1 and illustrated in Figure 3.

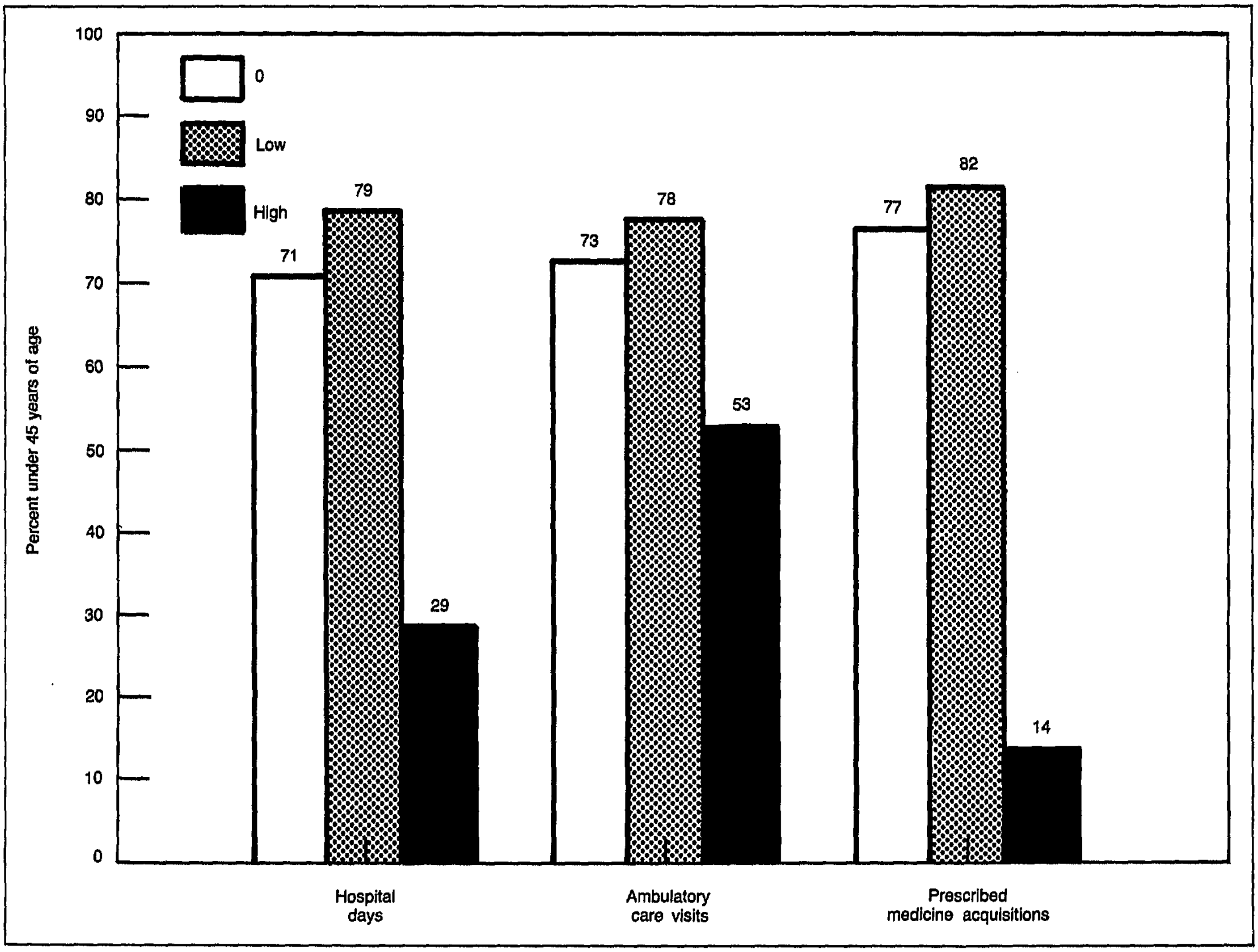
Demographic Characteristics

When the age distribution of high users of each of the three service categories considered here is examined, the percent of persons under 45 years of age is significantly higher among low users than among high users both for hospital days and prescribed medications (Figure 4). For example, 79 percent of low-volume users of inpatient services are under 45 years of age, compared with 29 percent of high-volume users. The difference is even more dramatic for prescription acquisitions, where 82 percent of low-volume users were under 45 years of age, as against only 14 percent of high users. Although the difference in the percent of persons under 45 years of age who were high and low users of ambulatory services is also statistically significant, it is not nearly as large as the differences observed for the other two services.

The age distribution within each category of use differs between males and females (Table 2). Among females categorized as low-volume users of hospital inpatient services (those with 1- or 2-day stays), 53.9 percent were between 17 and 44 years of age. In contrast, only 36.1 percent of males who were low-volume users of inpatient care were 17–44 years of age. This difference in the age distribution of male and female low-volume users is probably due in large part to deliveries that occur among women in this age group. Differences in the age distribution between males and females among high-and low-volume users of both ambulatory services and prescribed medications (Tables 3 and 4) are not nearly as large as the difference in hospital days.

There are major differences in the sex composition within low- and high-volume user groups for certain services. Low-volume users of ambulatory services are nearly equally divided between males and females. However, females comprised 48.0 percent of low-volume users in this service category,

Figure 4
Percent of persons under 45 years of age in 0-, low-, and high-use categories, by type of service: United States, 1980



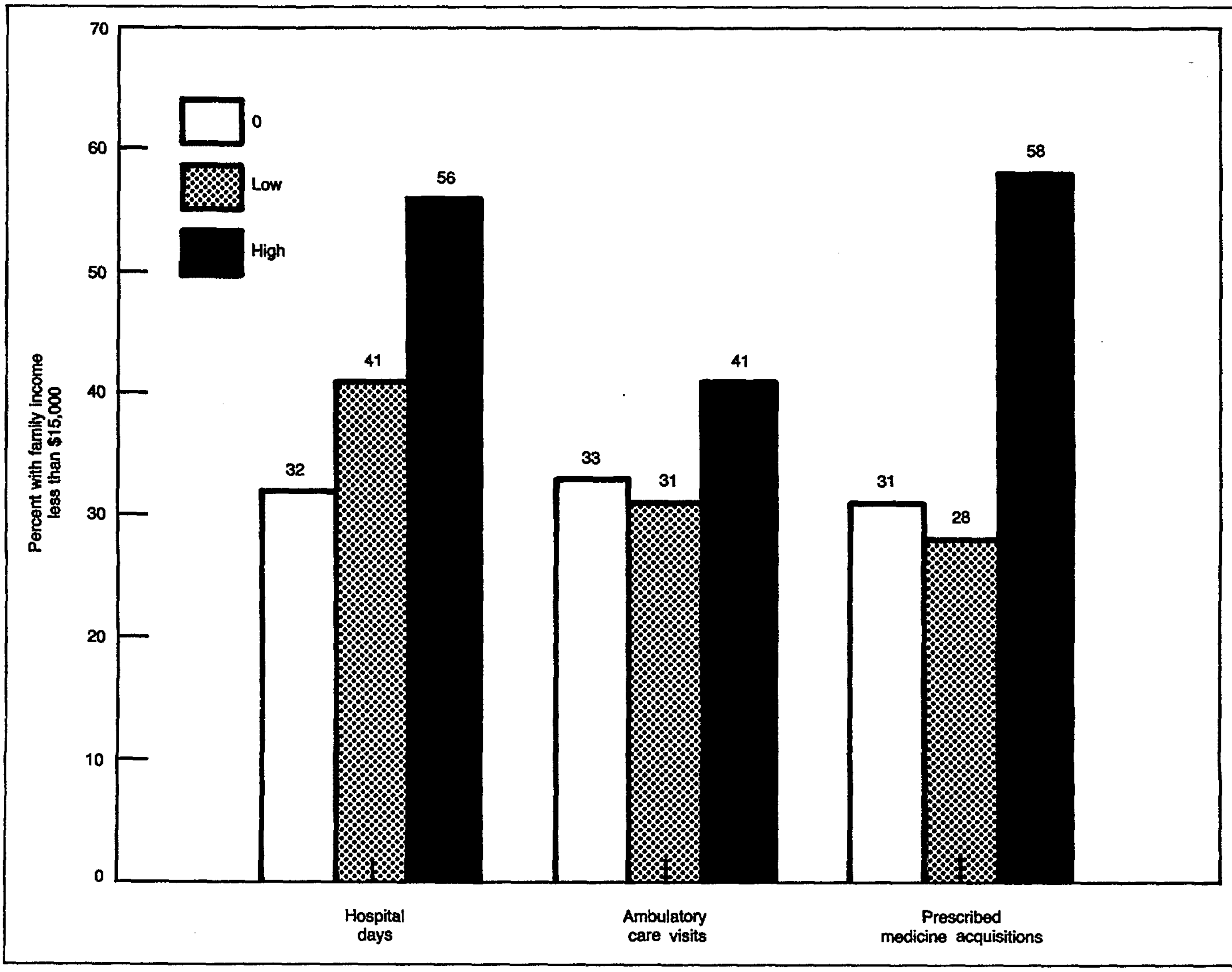
which is significantly less than the percent of females in the civilian noninstitutionalized population (51.8 percent). When high-volume users of ambulatory care are considered, however, nearly two-thirds of high-volume users were female, significantly more than their percent of the reference population. Similar sex differences can be seen between low- and high-volume users of prescribed medications. Given the greater representation of females among high-volume users of both ambulatory services and prescriptions, it is worth noting the absence of a similar difference in the sex composition of high-volume users of hospital days: For that service category the percents of male and female high-volume users were approximately the same.

There are small differences in racial composition between low- and high-volume users in each of the three categories of service (Tables 2–4). The difference is statistically significant only for ambulatory visits, where 12.8 percent of low users were black in contrast to only 8.8 percent of high users (Table 3).

In each of the three categories of service, persons classified as high-volume users were more likely to have family incomes less than \$15,000 than low-volume users were (Figure 5). Among high users of hospital days, for example, 56 percent belonged to families with an annual income of less than \$15,000, compared with 41 percent of low-volume users. This is particularly a reflection of the fact that the elderly are more likely to be in the lower income groups and to be higher users of hospital inpatient services.

It is notable that high-volume users of inpatient hospital services living in families with an annual income of less than \$15,000 comprised 2.9 percent of the civilian noninstitutionalized population at that income level, but high-volume hospital users in families with an annual income of \$35,000 or more were only 1.2 percent of the civilian noninstitutionalized population at that economic level. Thus the proportion of the relatively poor who were high users was more than twice the proportion of those relatively well off. The difference is most striking for the category of prescribed

Figure 5
Percent of persons with family incomes less than $15,000 in 0-, low-, and high-use categories, by type of service: United States, 1980



medicine acquisitions, where 58 percent of high-volume users were in families with incomes less than $15,000.

A measure of income that takes family size into account is the relation of the family's income to the poverty level. When the relative poverty status of high-volume users is considered, a picture emerges that is similar to the one described above for family income: A substantial number of high-volume users were in the category closest to the poverty line (Tables 5–7). Using 200 percent of the poverty level as a benchmark, 46.9 percent of high-volume users of hospital care fell below this level compared with 38.5 percent of low users (Table 5). It is also notable that although 46.9 percent of high users of hospital days were among the relatively poor, the corresponding proportion in the total civilian noninstitutionalized population was only 31.4 percent. By contrast, in the category of ambulatory visits the percent of persons falling below 200 percent of the poverty level is approximately the same across all levels of use (Table 6).

When levels of use are investigated by the four census regions of the United States, no distinct patterns emerge (Table 8). It is interesting to note that there is no regional variation in the relative frequencies in any of the use categories once the population's regional distribution is taken into account (Table 8).

Some regional patterns in the relative use levels of ambulatory care visits and of prescribed medications do emerge. In particular, the West had the highest percent of high users of ambulatory care, 28.6 percent (Table 9).

Prescribed medicine use categories also show some regional variation, with residents of the South comprising 40.2 percent of high users, even though only 31.2 percent of the civilian noninstitutionalized population resided there (Table 10).

Health Care Coverage

Nearly 95 percent of those who are 65 years of age or over are covered by Medicare, whereas a very small percent of those *under* 65 years of age are beneficiaries of that program. Therefore, to facilitate the analysis of use levels by health care coverage, the population was divided into two groups, those under 65 years of age and those 65 years of age or over.

In the civilian noninstitutionalized population under 65 years of age, significantly different patterns by use level and health care coverage emerge in all three types of service: Hospital days, ambulatory care visits, and prescribed medicine acquisitions (Tables 11–13). Of this population, 65 percent may be called the fully privately insured, that is, individuals who are covered for the entire year by private insurance only. Yet among high users of hospital days, this group comprised only a little more than half (51.6 percent). This is in stark contrast to those covered by one or more public programs, all of whom are disproportionately represented in the high hospital use category. For example, 9.7 percent of high-volume users of hospital services were covered by two or more public programs, yet they constituted only 2.2 percent of the civilian noninstitutionalized population under 65 years of age. This group, in other words, was found 4.4 times as frequently among high hospital users as in the population. Public programs that provide health care coverage for these persons are Medicaid, Medicare, and various State and local government welfare programs. Because poverty and medical indigency are the principal eligibility criteria for Medicaid and State and local programs, and the presence of certain disabilities and end-stage renal disease are the principal conditions that qualify persons under 65 years of age as Medicare beneficiaries, the population group covered by one or more public programs is likely to be sick or disabled. This is reflected in their relatively very high use of hospital days. This finding is reinforced in the multivariable analysis, where coverage by public programs is shown to be a powerful predictor of hospital use.

The disproportionalities obtain as well for those who have mixed private-public coverage, with this group comprising 17.4 percent of high hospital users but only 8.0 percent of the reference population.

Those not covered at all during the year by any program and those who have insurance during only part of the year are at the other end of the spectrum, with the direction of the disproportionalities reversed. Thus, 8.7 percent of the civilian noninstitutionalized population under 65 years of age had no health care coverage at all during the year, and only 2.8 percent of high hospital users fell into this category. In other words, while individuals covered by two or more public programs were found 4.4 times as frequently among high users as they are in the reference population, those without any health care coverage were found in the high-use category only a third as frequently as in the reference population.

It may be noted that the patterns of health care coverage among high users of hospital services are not replicated in the low hospital use category, with some exceptions. For example, those without any coverage for the entire year are found in the low-use category, as they have been in the high-use category, only about half as frequently as their relative proportions of the civilian noninstitutionalized population.

These patterns strongly suggest that health care coverage, and especially coverage by public programs, plays an important role in determining levels of hospital use, an implication that is further explored in the next section on the multivariate analysis of high use of hospital care.

Relationships found between levels of hospital use and health care coverage are replicated but in a more attenuated form for the various use levels of ambulatory visits (Table 12).

The patterns for the use of prescribed medications are almost the same as those found for hospital inpatient services (Table 13). Those covered by two or more public programs (2.2 percent of the reference population) and those covered by the combination of private and public programs (8.0 percent of the reference population), constituted 9.0 percent and 20.0 percent respectively, of the high users of prescribed medications. Thus individuals in these health care coverage categories were found among high users of prescribed medications at a rate 2.5 to 4 times greater than their frequency in the civilian noninstitutionalized population.

Within the civilian noninstitutionalized population 65 years of age or over, two-thirds are covered *both* by Medicare and private insurance (Table 14). This population group is represented among the hospital-use categories roughly in proportion to their population share, with 68.1 percent of high, 71.3 of intermediate, and 72.3 percent of low hospital users comprised of individuals with Medicare and private coverage. Those with Medicare coverage only, 16.3 of the reference population, were found proportionately somewhat less frequently among high hospital users, 11.2 percent, but this difference is not statistically significant. On the other hand, those who were covered both by Medicare and at least one other public program, generally Medicaid, constituted 20.1 percent of the high hospital user category, almost two times their reference population proportion (11.3 percent).

The patterns found in hospital use for the civilian population 65 years of age or over is replicated almost exactly when the use category of ambulatory services is considered (Table 15). Those with Medicare only are found disproportionately less often, and those with Medicare and any other type of coverage, disproportionately more often, in the high ambulatory use category.

These relationships, found both for hospital and ambulatory care, are replicated once again when the use of prescribed medicines is investigated (Table 16). Only 2.1 percent of high users were represented by the 4.4 percent of the reference population not covered by Medicare. Those who are covered by Medicare and private insurance comprised 71.9 percent of the high users of prescribed medications and only 66.8 percent of the reference population. Those covered both by Medicare and one other public program constituted 11.3 percent of the reference population and 14.3 percent of high users of prescribed medicines. In other words, only

10.7 percent of those with Medicare only were found in the high prescribed medicine use category, while 16.6 percent of those who were covered both by Medicare and private insurance were high users of prescribed medication.

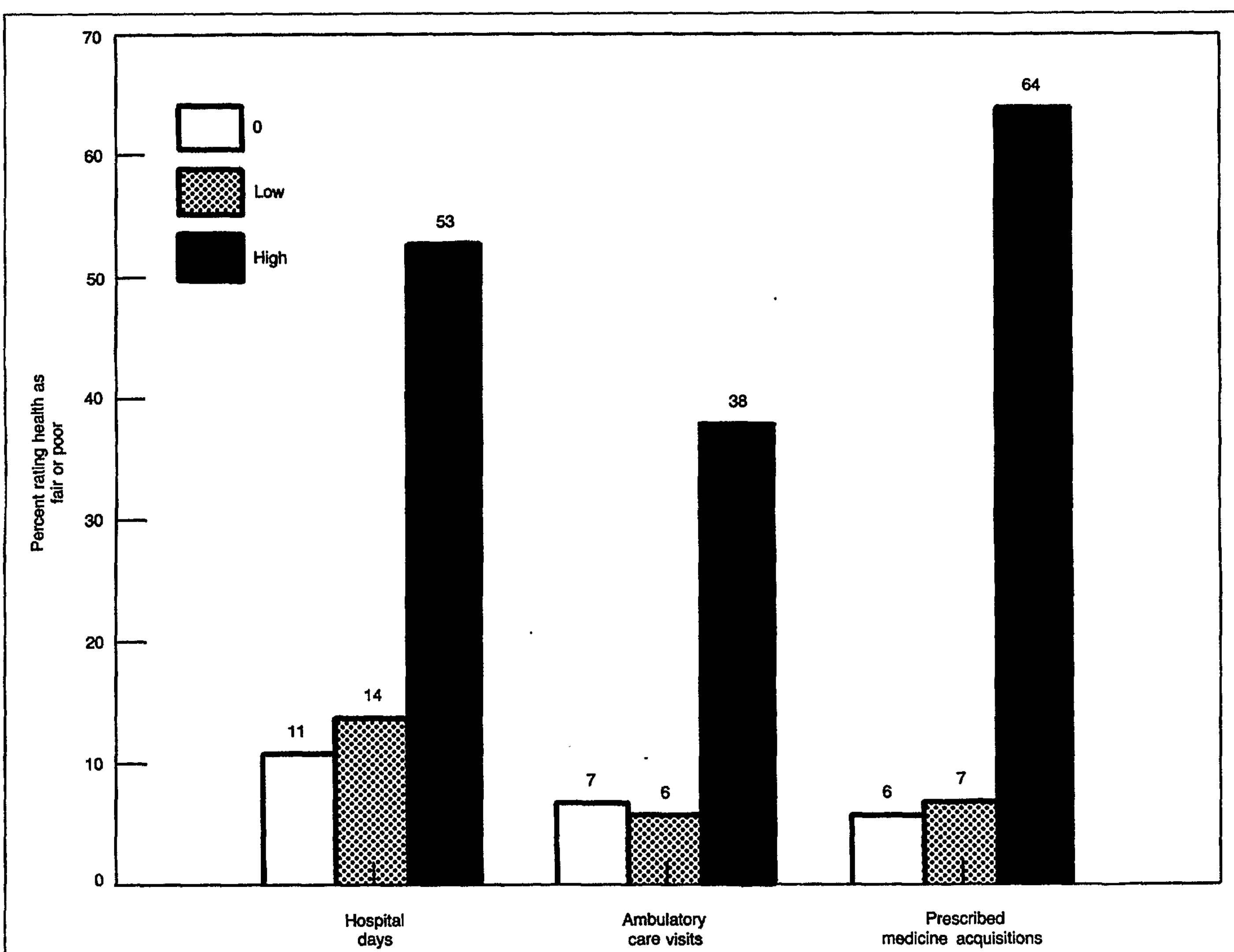
Health Status and Functional Limitations

There is a close and striking relationship between perceived health status and level of use across all three services considered. Although health status was reported to be fair or poor for only 12.9 percent of the civilian noninstitutionalized population, that percent was 52.6 for those in the high-use category for hospital care, 37.8 percent in the high ambulatory use category, and 63.9 percent for high users of prescribed medications (Figure 6 and Tables 17-19). This is in sharp contrast with those who had no hospital days, 89.1 percent of whom had a reported health status of excellent or good, as did 86.1 of low users of hospital care.

The relationship between perceived health status and level of use is particularly noteworthy since health status was reported at the beginning of the survey year, in the first of five waves of interveiws. In single wave surveys, information both on health status and use of service is elicited within the same time frame, and thus reported health status in such surveys may reflect the use of services, that is, be postdictive rather than predictive. In NMCUES this problem is essentially overcome by having health status reported at the beginning of the year, with the utilization data measuring levels of use during the entire year. The predictive power of perceived health status as measured in NMCUES is clearly demonstrated in the multivariable analysis of hospital inpatient services, where it is found to be one of the two most important predictors both of any hospitalization at all and of the high use of hospital inpatient services.

It might be noted that 47.4 percent of high users of hospital inpatient services also had a reported health status of excellent or good. The apparent contradiction is explained

Figure 6
Percent of persons who rated their health as fair or poor in 0-, low-, and high-use categories, by type of service: United States, 1980



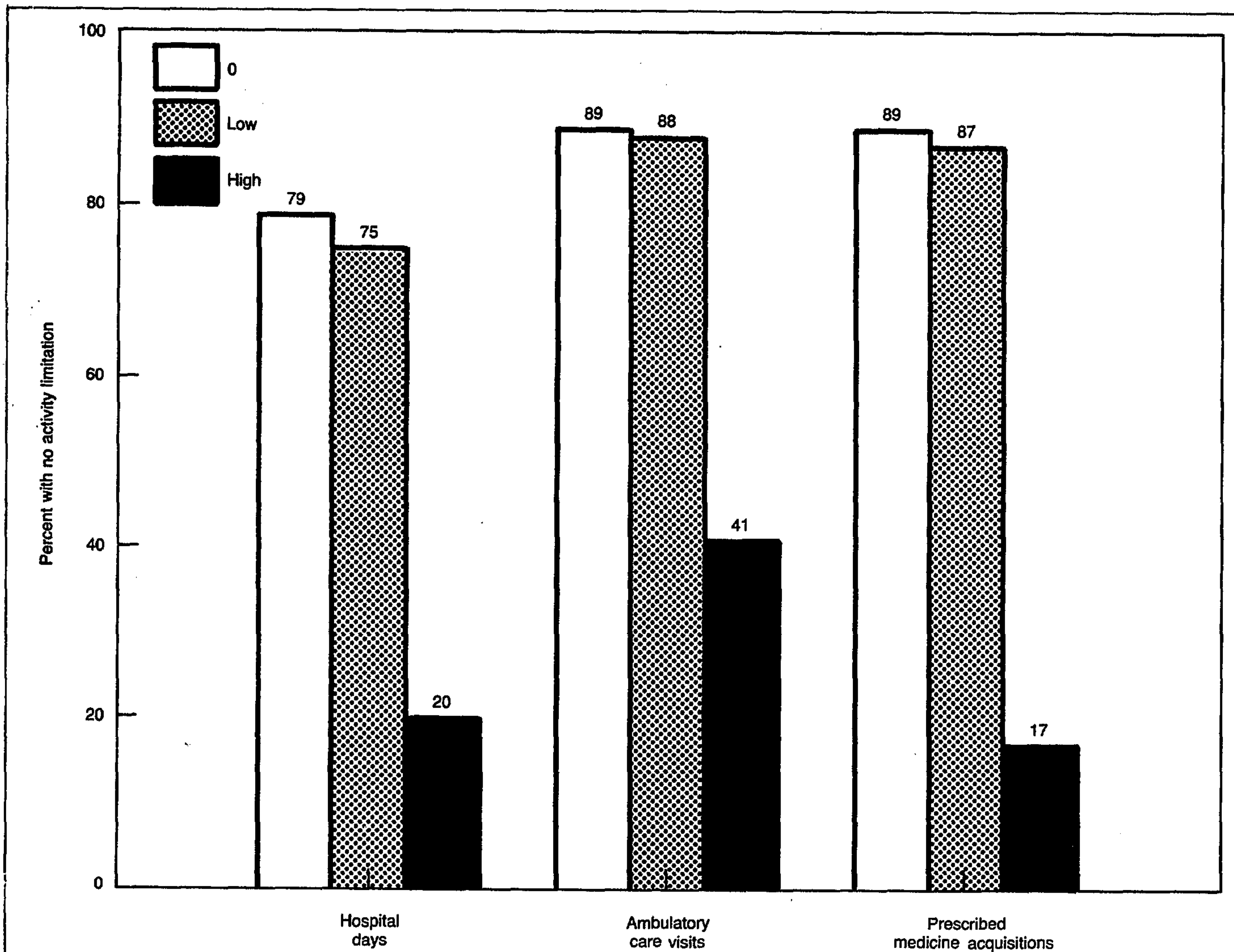
by the fact that high users of hospital care as a whole constituted a relatively small percent of the civilian noninstitutionalized population (1.7 percent), but 87.0 percent of that population had its health status rated as good or excellent. Thus even a very small percent of those whose health status had been rated good or excellent and who were high users of hospital care represent a very large percent of that small category. Indeed, only about one-half of a percent (0.55 percent) of those whose health status was rated excellent were in the high hospital use category, but they represented 16.2 percent of this high-use cateogry. Among those whose health status was rated to be good, only 1.45 percent were high users of hospital inpatient services, but they represented 31.2 percent of this use category. On the other hand, among the 3.6 percent of the civilian noninstitutionalized population that had its health status reported to be poor, 13.85 percent were in the high hospital inpatient use category.

The picture of distributional disparity in the use of hospital services, as well as the validity of the reported data, is reinforced when nonusers of hospital inpatient services are considered. Of the group that had no reported hospital inpatient services in 1980, 89.1 percent had a reported health status of good or excellent; in contrast, only 11.0 percent of this no-use group had their health status reported as fair or poor. Thus it is clear, as one might expect, that the high hospital use category is composed of relatively sicker people whether compared with those who used no hospital inpatient services or with those who used them at low rates.

The patterns observed for hospital use are also found across use levels of prescribed medicines. Those whose health status was rated as fair or poor comprised 63.9 percent of the high prescribed medicine use category (Table 19). In this high-use category, health status was reported as either excellent or good for 36.1 percent, in contrast to 94.1 percent in the zero-use category. Similarly, only 5.9 percent of nonusers of prescribed medications had their health status reported as poor or fair.

Figure 7

Percent of eligible persons reporting no activity limitation in 0-, low-, and high-use categories, by type of service: United States, 1980



The relationships between perceived health status and the level of ambulatory care use, while similar, are weaker. The 37.8 percent of high users of ambulatory visits whose reported health status was fair or poor (Table 18), was lower than the corresponding percent for hospital days and for prescribed medicine acquisitions (Figure 6). Still, for ambulatory visits the association between higher levels of use and lower health status ratings is quite explicit. For example, in the low-use category, 58.8 percent had a reported health status of excellent, and for 34.7 percent it was rated as good, but only 6.5 percent of this use level had a health status rating of fair or poor. Much the same pattern obtains among zero users. In the intermediate-use category a small shift occurs, with 47.0 percent with a reported health status of excellent, 38.1 percent with good, and 15 percent who had a reported health status of fair or poor.

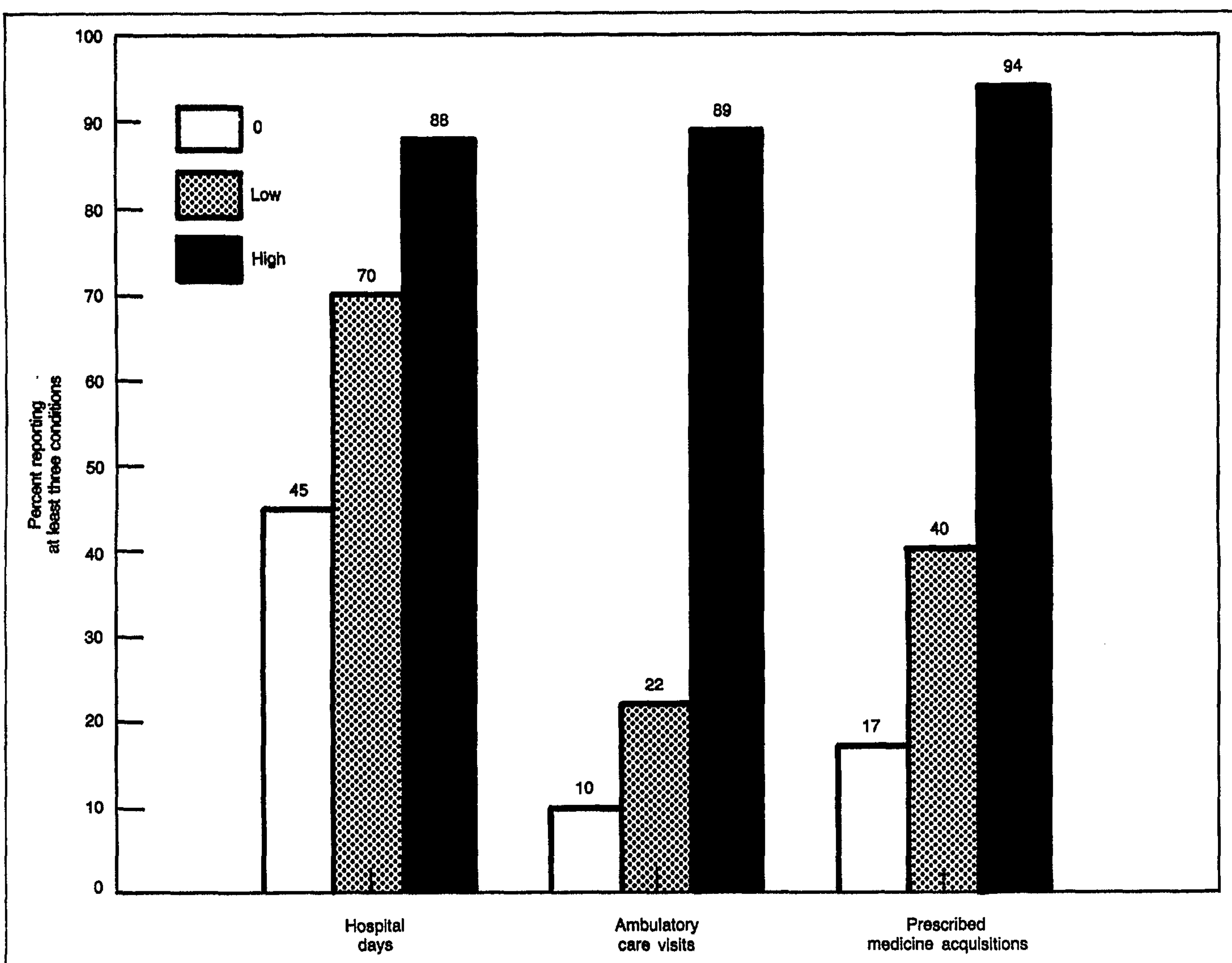
Because data on activity limitation was collected only for persons 17 years of age and older who survived the sample year, the prevalence of activity limitation among those under 17 years of age is not known. The appropriate basis for relating levels of use to degree of activity limitation, therefore, is the U.S. civilian noninstitutionalized population 17 years of age and over who lived through 1980. On this basis, the pattern for functional limitations mirrors closely that for health status: The percent of persons with no activity limitation is large among low users and zero users and significantly smaller among high users. (Figure 7 and Tables 20-22). In this reference population, 79.4 percent and 74.5 percent of zero and low users of hospital days, respectively, had no activity limitation, compared with 20 percent of high users in this service category. As was also the case for perceived health status, the differences related to activity limitation are smallest for ambulatory visits.

Diagnostic Conditions

Almost all high-volume users in every category of service (88.1 percent for hospital days, 88.6 percent for ambulatory visits, and 94.1 percent for prescribed medications) reported having had at least three different diagnostic conditions during

Figure 8

Percent of persons reporting at least three conditions during the year in 0-, low-, and high-use categories, by type of service: United States, 1980



the year (Figure 8 and Tables 23–25). The percent of individuals with three or more conditions was significantly lower in the low-use and zero-use categories for all three medical services, but most specifically for ambulatory care. In ambulatory care only 9.5 percent of nonusers and 22.5 percent of low users reported having three or more unique conditions (Table 24). The 45 percent of nonusers and 70 percent of low users of hospital days who had three or more medical conditions reported represent higher proportions than the percents reported within the same use categories for ambulatory visits and prescriptions (9.5 and 17 percent, respectively, for nonusers). It is important to keep in mind, however, that the zero and low users of hospital services represented 91 percent of the total civilian noninstitutionalized population, and a substantial number of these individuals made at least one ambulatory visit or had one prescription filled for which they could list a medical condition. Consequently, having three or more different conditions is probably not as useful a distinguishing characteristic of high-volume use of hospital inpatient services as it is for ambulatory visits or prescribed medicine acquisitions.

It is particularly notable that 22.9 percent of high users of inpatient hospital service reported having 10 or more different conditions during the year, the same number of conditions (10 or more) reported by only 2.9 percent of the civilian noninstitutionalized population. Ten or more conditions were also reported by 20.9 percent of the high users of ambulatory visits and 24.5 percent of the high users of prescribed medications. Thus, this very small proportion of the civilian noninstitutionalized population with 10 or more conditions (2.9 percent) constituted about one-fourth of the high users of each of the three medical services.

To identify the major condition groups most likely to be associated with each level of hospital use, the number of conditions was tallied across all hospitalizations in each use level and then classified by broad diagnostic categories (Table 26).

The following conditions were among those afflicting the greatest percent of high users of hospital days: Diseases of the circulatory system (this disease category was associated with at least one hospitalization for 32.8 percent of high users); injury and poisoning (21.8 percent); neoplasms (16.7 percent), diseases of the respiratory system (16.1 percent), diseases of the musculoskeletal system and connective tissue (15.1 percent); and diseases of the digestive system (12.8 percent).

Among low users of hospital days, the leading category was diseases of the respiratory system (15 percent), followed by diseases of the genitourinary system (also 15 percent), injury and poisoning (13.5 percent), diseases of the nervous system (6.9 percent), and complications of pregnancy and childbirth (6.9 percent). Thus, diseases of the respiratory system and injury and poisoning figure prominently both among high users and low users of hospital care.

Although it is often thought that surgery is associated with high use of hospital inpatient services, the frequency of surgery does not differentiate high users of hospital days from low users: 49.2 percent of high users and 51.0 percent of low users of hospital days had no surgery in 1980 (Table 27).

Decedents

Four-tenths of one percent of the civilian noninstitutionalized population covered in NMCUES died during the survey year. This 0.4 percent of the population accounted for 12.1 percent of all high users of hospital days. The very high use of hospital inpatient services by those who died during the year is illustrated by the fact that whereas only 1.5 percent of the reference population who survived the year were among high users of hospital inpatient services, more than half, 54.2 percent, of those who died had high hospital use (Table 30). That hospital care is used intensively by those who are dying is reinforced by the finding that whereas 71.7 percent of those who died during the year were either in the intermediate- or high-use category, only 4.4 percent were in the low-use category. This pattern is consistent with previous reports by Lubitz and Prihoda (1984) and McCall (1984).

Patterns of use of prescribed medications and of ambulatory care by those who died during the year are similar but less striking. Thus, among persons who survived the year, only 3.6 percent were high users of prescribed medications (Table 32) and 4.4 percent were high users of ambulatory care (Table 31); of those who died, however, almost a fourth had high use of these services.

Hospital Use and Ambulatory Visits

Use of inpatient hospital services is often given greater attention than the use of other services because the costs of hospital care overshadow those of other services and because use of inpatient care is thought to trigger use of ambulatory and other services as well. That appears to be true when use levels of ambulatory visits associated with hospital days are examined (Table 28). In fact, regardless of the level of use of hospital days, among those who had any hospitalizations, 94 percent were in the intermediate- or high-use categories of ambulatory visits. The only difference is that one-third of high users but only 6 percent of low users of hospital days were high users of ambulatory visits. An examination of the small number of cases with hospitalizations but no ambulatory visits showed that they include a disproportionate number of deaths and institutionalizations as well as of data values that appear to be inconsistent. It is also important to note, however, that in the relationship between inpatient and ambulatory use, about two-thirds of high users of ambulatory visits had *no* hospitalizations.

Summary of Univariate Analyses

Comparisons of high and low users, as well as nonusers, within the three major categories of health care services by their selected demographic socioeconomic and health characteristics showed both differences and similarities.

Low users versus nonusers—Low users are quite similar to nonusers in several respects. Across all three service categories considered, the percent of individuals among both low and nonusers who are under 45 years of age, had their health rated as poor or fair, and had no activity limitation is similar.

On the other hand, low users of hospital inpatient services differ from nonusers with respect to the critical characteristics of income and health care coverage. A larger percent of low users of hospital inpatient services had family incomes of less than $15,000 compared with nonusers. Moreover, low users were more likely than nonusers to have health care coverage through public programs or coverage by one or more sources. Another characteristic on which low users differ from nonusers is the number of conditions: Whereas the two kinds of users did not differ on their reported health status or activity limitation, they did differ on the percent with three or more conditions (see Figure 8).

Low and high use across service categories—High users of ambulatory services were similar to low users and nonusers with respect to age (Figure 4) and private insurance coverage. High users, however, tended to be poorer (Figure 5), to have their health status rated fair or poor more frequently (Figure 6), and to report one or more activity limitations more often (Figure 7) than nonusers and low users of ambulatory services.

High users of hospital days differed markedly from low users. High users of inpatient hospital services compared with low users tend to be considerably older (Figure 4), relatively poorer (Figure 5), belong to families headed by someone without a high school diploma, and have their health rated as fair or poor (Figure 6).

High users of hospital services are five times more likely to report more than 10 unique medical conditions than low users. Moreover, high and low users of hospital inpatient services also differ in the percent of conditions represented by various disease categories (see Table 26). Diseases of the circulatory system as well as neoplasms, for example, are found more frequently among high than among low users.

Multivariable Analysis of High Use of Hospital Care

The preceding analyses have focused almost exclusively on relationships between two variables at a time. In many cases it is clear that interpretation of the relationship between two variables depends upon a third, or even several other variables that have not been considered explicitly. For example, high use of hospital care is associated with being below 200 percent of the poverty level in the univariate relationships examined earlier. However, poor persons also are more likely to be less healthy and older. It is important to determine whether the relationship between poverty level and use of hospital care is direct or whether that relationship reflects, at least partially, the effects of age and ill health.

One way to approach such an analysis is to examine tables in which multiple factors are considered simultaneously. Unfortunately, multivariable tables quickly become cumbersome because of their size and complexity, and interpretation of these tables is difficult without an underlying model characterizing the relationships among the variables. To understand more fully the complex relationships between a dependent variable, such as high use of a given service, and several predictors of the type examined in the bivariate tables presented previously, a regression modeling strategy was pursued.

Because hospital care consumes a major portion of the health care expenditures each year in the United States, a predictive model was developed for high use of hospital days. The objective of the model was to assess what factors determine whether or not individuals will be high users of hospital care.

High use of hospital care depends, of course, on an individual first being hospitalized, and then on the individual being hospitalized for a long period of time, or alternatively, having a series of shorter hospitalizations. A model that simply examines factors associated with extensive use of hospital inpatient care, without also examining factors influencing hospitalization in the first place, may fail to uncover important effects of such factors because they are overwhelmed by other factors that influence whether hospitalization occurs, but not the high use or length of hospitalizations. For this reason, high use of hospital care was investigated in a two-part model. The first part focuses on factors associated with whether an individual in the civilian noninstitutionalized population was hospitalized during 1980. These factors were examined in a multiple logistic regression framework. The second part investigates, also through multiple logistic regression, factors hypothesized to be predictive of high use, given that a person was hospitalized at least once in 1980.

In the first of the two models, the dependent variable is equal to one if the person was hospitalized in 1980 and zero otherwise; in the second model, the dependent variable is equal to one if the person had 17 or more days in the hospital in 1980 (which was defined as high use) and zero otherwise. Given the dichotomous nature of these dependent measures, a standard ordinary least squares regression method would yield inefficient estimators, particularly because in both models the dependent variable takes on one value much more frequently than the other: Zero, which corresponds to "not hospitalized" in one model and "not a high user" in the other, is the likely value for more than 85 percent of cases in both models. Based on these considerations, logistic regression methods were used to develop the predictive models. The assumptions and methods of logistic regression are discussed in Appendix V.

Logistic regression methods are conceptually similar to standard regression methods, but the interpretation of the coefficients is somewhat different. For a set of predictors $X_1, X_2, \ldots, X_p$, partial logistic regression coefficients are estimated that represent the unit change in the logarithm of the odds of being one of the two values of the dependent variable, given a unit change in the predictor. For example, for the dependent variable in the first model (i.e., hospitalization, yes or no), the coefficients in the logistic regression denote the change in the logarithm of the odds that a person will be hospitalized for a unit change in each of the predictor variables, all other predictor variables remaining fixed at some mean value.

The logistic regression coefficients can also be converted to odds ratios which may provide a more direct assessment of the importance of the factors in predicting hospitalization and high use of hospital care. The odds ratio indicates the extent to which persons in a particular group are more or less likely to be hospitalized, and, if hospitalized, to use 17 or more hospital days, than persons in the reference group when all the other factors are fixed at an average value. Both the logistic regression coefficients and the odds ratios derived from them are presented in the subsequent section.

Model for Predicting Use of Hospital Care

There are a large number of factors which previous discussion in this report and the medical care literature suggest as predictors of hospitalization during a given time period.

It would be difficult to construct a model that incorporates every factor that has been examined because the model estimates would likely be unstable as a result of high correlations among many of the potential predictors. A set of predictors or independent variables was selected which includes the most important predictors indicated in the two-way analyses presented previously. These predictor variables include the following:

1. Demographic factors often cited for their association with use of hospital care: Age, sex, and race.
2. Characteristics that represent financial access to inpatient care: Health care coverage (see Appendix V for a definition of the categories used) and income (represented by poverty level which adjusts income for family size).
3. Having a usual source of care, a characteristic believed to facilitate seeking and receiving care.
4. Perceived health status, which reflects the need for care; because perceived health status was reported at the beginning of the survey period, it is a useful predictor of health care use.
5. Region of the United States, a proxy for differences in supply characteristics across the country.
6. Two indicators to identify subgroups of the population that are high users of health care: An indicator for females between 17–44 years of age with good or excellent reported health status (who can be expected to use a substantial amount of inpatient care not because of sickness but for deliveries and other reasons related to reproduction), and an indicator for persons who are below twice the poverty level and reported health status as fair or poor (who are expected to be more likely to use inpatient services, all else being equal, because of the reinforcing processes that are thought to exist between sickness and poverty, this reinforcement results in less access to care initially and greater likelihood of subsequent complications requiring more drastic interventions later).

Variables were created for these measures to use in a logistic regression model to predict hospitalization during 1980. Intervally scaled measures (e.g., age, poverty level) were divided into categories and indicator variables were created as predictors for each of the measures. For all predictors except poverty level, the indicator variables were created by designating one category as the reference category and by defining the indicators for the remaining categories as equal to one for persons in the category and as equal to zero otherwise. Thus, logistic regression coefficients for these indicator variables denote the unit change in the logarithm of the odds of being hospitalized relative to the reference group for that particular indicator.

An indicator variable for poverty level was constructed using the usual Analysis of Variance parameterization. The logistic regression coefficients for poverty level categories denote the departure of the logarithm of the odds ratio for that category from the overall mean logarithm of the odds ratio.

Functional limitation status is a variable that might influence hospitalization but was not included in the model. Those who have substantial functional limitations may be more likely to require inpatient care when other, unimpaired persons, would not. One problem with using this variable derives from its having been measured toward the end of the survey period, and therefore it potentially reflects as much the effect as the cause of hospitalization. Even though, among those who reported a functional limitation and were hospitalized during the year, three-quarters reported that the limitation lasted more than 12 months, the direction of casuality cannot be established unambiguously. A second problem arises because functional limitations were measured only for those who were alive at the end of the survey period and were 17 years of age and over. Consequently, including functional limitation in the model would have meant excluding from consideration those who were under 17 years of age or who died during the year.

Rather than limit the analysis to individuals 17 years of age or over and persons alive the entire year, two separate logistic regression analyses were conducted. In the first, all the data were used, but functional limitation was not used as a predictor. In the second, the data were limited to persons 17 years of age or over and alive during the year, and functional limitation was added to the basic set of predictors to provide insight about the importance of this variable for predicting hospitalization.

The logistic regression coefficients from the analysis of the model using all the data are presented in Table B. The ratio of the coefficient to its adjusted standard error can be used to assess whether a given coefficient is statistically significantly different from zero. The odds ratios are also presented as an indication of whether the factor level of interest is associated with a lower probability of hospitalization (i.e., an odds ratio less than one and a negative coefficient) or is associated with a higher probability of hospitalization (i.e., an odds ratio greater than one and a positive coefficient) than the probability of the reference group for the factor. Because the coefficients are partial regression coefficients in a multiple regression model, they reflect the effect of the factor level given that all the other factors in the model are set at an average value, and thus held statistically constant.

Neither sex nor race are significant predictors of hospitalization at the 5 percent level of statistical significance once all other factors are controlled, because the ratio of the coefficients to adjusted standard errors for those factors are between the limits of -1.96 and $+1.96$. The odds ratios and their confidence limits for sex and race show no significant increased or decreased risk of hospitalization, because the odds ratios are indistinguishable from 1.0.

With respect to age, persons 75 years of age and over and those under 35 years of age had significantly more hospitalizations than persons 35–54 years of age. Persons 55–74 years of age did not have significantly greater risk of being hospitalized than the reference age group.

The two sets of factors that stand out by the size of their coefficients are health care coverage and perceived health status. The types of coverage that are likely to be the most comprehensive—having coverage through two public programs or one public and one private—are those most strongly

Table B

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

Independent variables	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	...	...	...	...	...
Sex						
(Male)	...	...	...	...	...	...
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)	...	...	...	...	...	...
Perceived health status						
Excellent	−0.2296	0.0567	−4.0476	0.7949	0.7112	0.8883
(Good)	...	...	...	...	...	...
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)	...	...	...	...	...	...
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 status[1]						
Below 200 percent of poverty level	0.1483	0.0600	2.4716	1.1599	1.0312	1.3046
200–499 percent of poverty level	−0.0728	0.0441	−1.6502	0.9298	0.8528	1.0138
500–699 percent of poverty level	−0.1191	0.0754	−1.5793	0.8877	0.7657	1.0291
(700 percent of poverty level or more)	...	...	...	...	...	...
Health care 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)	...	...	...	...	...	...
Region of residence						
(Northeast)	...	...	...	...	...	...
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)	...	...	...	...	...	...
Healthy females 17–44 years of age						
Yes	0.4387	0.0813	5.3970	1.5507	1.3223	1.8185
(No)	...	...	...	...	...	...
Poor persons with poor health						
Yes	−0.5032	0.1596	−3.1529	0.6046	0.4422	0.8266
(No)	...	...	...	...	...	...

[1]Analysis of variance parameterization.

NOTE: Proportional reduction of error = 5.1 percent.

related to hospitalization. Similarly, having a reported health status that is fair or poor has a strong positive relation to hospitalization, whereas excellent health status is negatively related to it. However, given that persons with Medicare and private insurance coverage are likely to be 65 years of age or over, and that the indicator variables for age in the model do not separately identify persons 65 years of age and over, the strength of the relationship between health care coverage and hospitalization shown in the model should be interpreted cautiously. The strong relationship in the model may be due to an age effect (i.e., becoming 65 years of age) rather than to the effects of health care coverage on decisions about seeking hospital care.

Two other factors that have strong positive relationships with hospitalization are the two indicators for population groups hypothesized as more likely to be hospitalized: Females 17–44 years of age with good or excellent reported health status, and persons with poverty status less than twice the poverty level and reporting poor health status. Healthy females 17–44 years of age have a 55-percent increased risk of hospitalization compared with other persons, while persons below twice the poverty level with poor perceived health status actually have nearly a 40-percent lower risk of being hospitalized. This latter finding is in the opposite direction of that expected and may be due to the controlling influence of the other factors in the model, particularly those related to health care coverage; or it may be due to access barriers not measured here.

Residence in the North Central Region of the country and poverty status below twice the poverty level both have small positive and significant relationships with hospitalization. By contrast, having a usual source of care had no significant effect.

Thus the best predictors of hospitalization are as follows:

- The factor associated with financial access to health care, health care coverage.
- The factor associated with need for health care, perceived health status.
- Membership in two important population subgroups, healthy females 17–44 years of age and persons in poor families who report poor health status.

These results suggest that the demographic variables, race, sex, and even age are related to hospitalization only to the extent that they are associated with such other factors as health care coverage or health status.

The importance of the factors considered in the model results presented in Table B may be examined by calculating the estimated probability from the model that an individual with particular characteristics would have been hospitalized in 1980. For example, consider a white male 35–54 years of age in good health with an average income (i.e., between 200 and 499 percent of poverty level), private insurance coverage, with a usual source of medical care, and living in the Northeast Region. The probability that such a person would be hospitalized may be estimated by adding the constant coefficient and other appropriate factor coefficients to obtain the logarithm of the odds of hospitalization. In this case, the logarithm of the odds is computed as the sum $(-2.8445 + 0.0984 - 0.0728 + 0.3468 - 0.0608 = -2.5329$. The logarithm of the odds can be transformed to the odds of hospitalization by exponentiation, $e^{-2.5328} = 0.0794$, and the odds converted to the predicted probability by computing $0.0794 / (1 + 0.0794) = 0.074$. That is, the hypothetical white male with the specified characteristics has a predicted probability of being hospitalized of 7.4 percent.

The relative importance of the different factors in the model can be examined by changing the characteristics of the hypothetical individual and estimating the probability for an individual with the new characteristics. For example, suppose that the hypothetical white male had reported poor health status instead of good health status. The predicted probability of hospitalization for an individual with all the same characteristics as the original hypothetical white male except for poor health status can be computed as previously shown. The sum of appropriate coefficients (i.e., the estimated logarithm of the odds) is

$$-2.8445 + 0.0984 + 1.5169 - 0.0728 + 0.3468 - 0.0608 = -1.0160.$$

Exponentiating, the estimated odds are $e^{-1.0160} = 0.3620$, which is converted to the predicted probability as

$$0.3620 / (1 + 0.3620) = 0.2658.$$

That is, poor health status relative to good health status has increased the risk of hospitalization for the hypothetical person by 3.35 times (i.e., 0.2658 / 0.0794). This increase is reflected in the large odds ratio for the poor health status category relative to the good health status category in Table B.

Alternatively, consider the risk of hospitalization for a white female with the same characteristics as the original hypothetical white male (i.e., 35–54 years of age, middle income, private insurance, good health status, a usual source of care, residence in the Northeast Region). The predicted probability of hospitalization for the female is computed in the same three steps as outlined for the hypothetical male:

(1) summation of coefficients to obtain the logarithm of the odds as

$$-2.8445 + 0.0984 + 0.0766 - 0.0728 + 0.3468 - 0.0608 = -2.4563;$$

(2) exponentiation to obtain the odds as $e^{-2.4563} = 0.0858$; and computation of the predicted probability as

$$0.0858 / (1 + 0.0858) = 0.0790.$$

Thus, the predicted probability of hospitalization for this female is 7.9 percent, somewhat higher than for the hypothetical male because the partial regression coefficient for the female group is positive. In addition, if the female were actually 35–44 years of age (and hence the indicator for females 17–44 years of age applies), the predicted probability of hospitalization is 11.7 percent. The interested reader can readily construct predicted probabilities for other individuals using the coefficients in a similar manner. A more detailed description of the estimation procedure for predicted probabilities is given in Appendix III.

These methods for estimating predicted probabilities are useful for interpreting the relative importance of the factors examined in the model, but there are several limitations to the general use of these logistic models. For one, the predicted probabilities for a hypothetical individual do not necessarily define patterns of hospitalization because the hypothetical person defined may represent a small portion of the population. In addition, variables are included in the model which are not statistically significant because findings in the medical care literature indicate that these variables represent important predictors of hospitalization. Nonetheless, including these variables in the model most likely leads to less efficient estimates than could be obtained from a model without them.

Finally, the model in Table B explains a small, but useful, amount of variation in the probability that a person was hospitalized in 1980. The model accounts for 5.5 percent of the error remaining if the mean probability of 12.2 percent was used as a predicted value for every person in the sample. Thus, there is still a large amount of error that may be explained by other factors that have not been included in the model. Those other factors may reduce the importance of the factors considered in this model and substantially change the predicted probabilities for hypothetical individuals considered here.

To assess the goodness of fit of the model further, the predicted probability of being hospitalized was computed for every person in the dataset. The distribution of the predicted probabilities was then examined for users and nonusers of hospital care to determine whether the model was able to discriminate between these two groups. Figure 9 shows the relative frequency distribution of the predicted probabilities by deciles for users and nonusers of hospital care.

A good fit of the model to the data would be observed if the predicted probabilities under the model were small (e.g., less than 50 percent) for nonusers and large for users. The results in Figure 9 suggest that the model does provide some discrimination between users and nonusers, although perhaps not as much as would be desired. In particular, 53 percent of the nonusers had predicted probabilities less than 10 percent, although only 29 percent of the users had predicted probabilities in that range. Similarly, 91 percent of the nonusers had predicted probabilities less than 20 percent compared with 71 percent of the users. Conversely, only 3 percent of the nonusers had predicted probabilities greater than 0.30, with 13 percent of the users had predicted probabilities greater than that value.

The ability of the model to discriminate users and nonusers through the assignment of predicted probabilities close to one and zero, respectively, is evident from the distribution in Figure 9. At the same time, though, the low proportion of the unexplained error accounted for in the model is also demonstrated by the lack of clear discrimination of users and nonusers by the predicted probabilities.

The goodness of fit of the model may be examined further by considering the extent to which the model improves the prediction of use or nonuse. Without the model, one could classify all persons in the sample as users or as nonusers. If all are classified as nonusers, then because 11.5 percent of the population are users, 88.5 percent of the persons would be predicted correctly as nonusers. With the model, a predicted probability that the person is a user can be obtained. By designating selected values of the predicted probabilities as thresholds (such that persons with a predicted value above the threshold are predicted users and those below are predicted nonusers), the ability of the model to predict use or nonuse can be examined.

Table C presents the probability of correctly classifying users, nonusers, and all persons when thresholds corresponding to deciles of predicted probabilities are used to define predicted use and nonuse groups. For example, if the threshold were chosen to be 0.1, then persons with predicted probabilities greater than 0.1 would be predicted users and the remainder are predicted nonusers. Using this threshold value, 70.6 percent of the users are correctly predicted to be users, while only 52.7 percent of the nonusers are correctly predicted. Overall, 54.7 percent of the persons in the sample would be predicted correctly using the 0.1 threshold for predicted probabilities.

The predicted probabilities for users and nonusers in Table C correspond to the sensitivity and specificity of a test procedure based on thresholds at different levels. The probability of a correct prediction for users is the sensitivity of the test at a given threshold, and the probability of a correct prediction for nonusers is the specificity of the test.

As the threshold value increases in Table C, the probability of a correct prediction drops for users and increases for nonusers. For all persons, the probability increases up to the 0.5 threshold and then decreases slightly to the prevalence of nonuse (i.e., 88.5 percent). In other words, the threshold of 0.5 would provide the largest number of correct predictions over all the tests given in Table C, but this finding is dominated by the fact that 88.5 percent of all persons were nonusers.

Table C

Probability of correct prediction by deciles of predicted probabilities for all persons, users, and nonusers: United States, 1980

Threshold of predicted probability	Probability of correct prediction		
	All persons	Users	Nonusers
0.0	0.115	1.000	0.000
0.1	0.547	0.706	0.527
0.2	0.836	0.282	0.908
0.3	0.873	0.124	0.970
0.4	0.882	0.047	0.990
0.5	0.886	0.015	0.998
0.6	0.885	0.002	1.000
0.7	0.885	0.001	1.000
0.8	0.885	0.000	1.000
0.9	0.885	0.000	1.000
1.0	0.885	0.000	1.000

Figure 9

Relative frequency distribution of predicted probabilities for users and nonusers of hospital care: United States, 1980

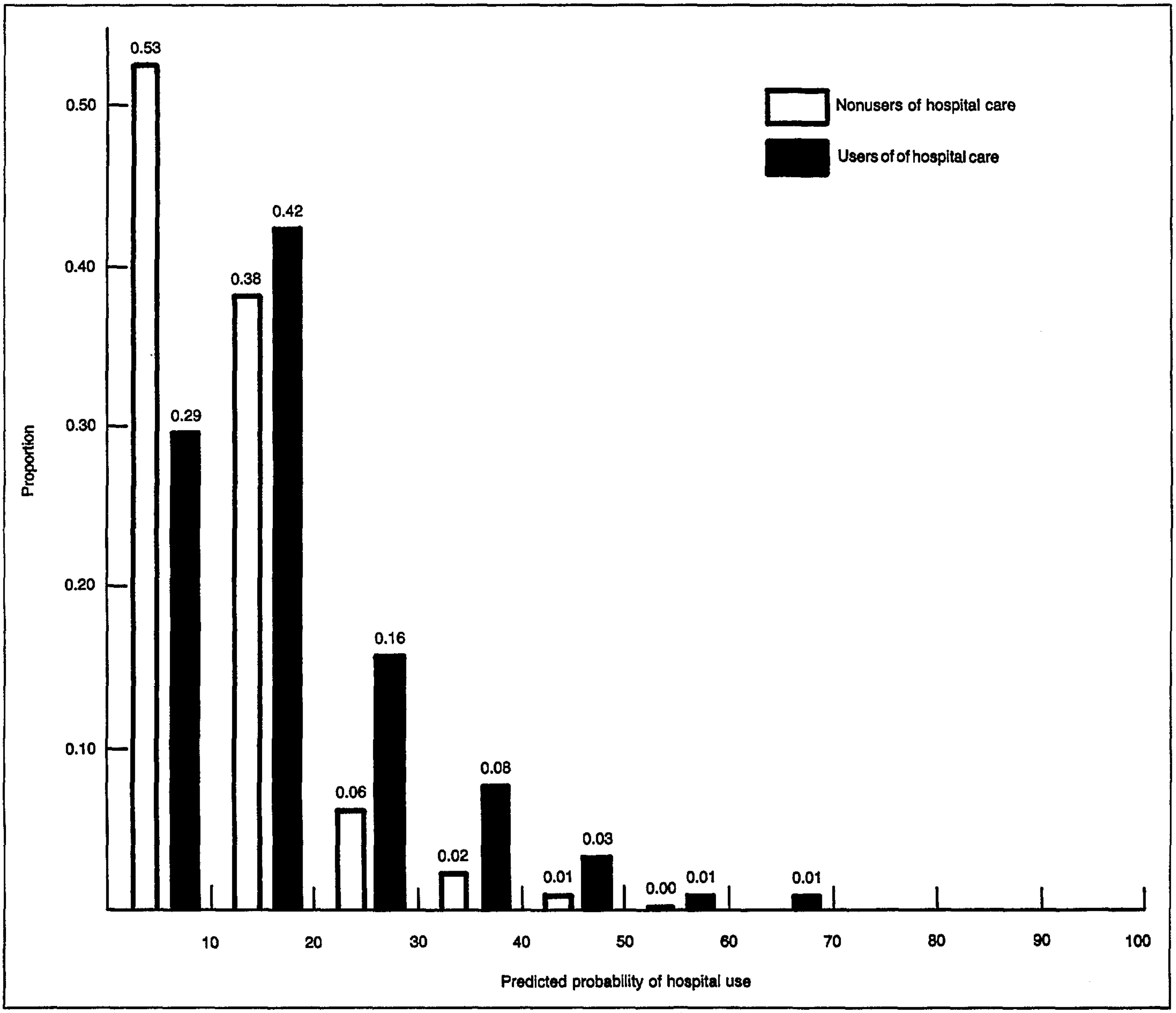


Table D

Predictive accuracy of alternative thresholds for predicting use and nonuse of hospital care: United States, 1980

Threshold of predicted probability	Predictive accuracy	
	Use	Nonuse
0.0	0.115	...
0.1	0.162	0.933
0.2	0.284	0.907
0.3	0.345	0.895
0.4	0.376	0.889
0.5	0.522	0.887
0.6	0.465	0.885
0.7	...	0.885
0.8	...	0.885
0.9	...	0.885
1.0	...	0.885

It is also useful to examine the ability of the model to predict use among all persons classified as users when a given threshold is selected. Table D presents the predictive accuracy of classification schemes corresponding to different thresholds. For example, if a threshold of 0.1 is used, 16.2 percent of those classified as users are actually users; 93.3 percent of those classified as nonusers are actually nonusers. The predictive accuracy for nonuse does not fall below 88.5 percent (when a threshold of 0.6 or greater is used), while the predictive accuracy for use does not fall below 11.5 percent (when a threshold of 0.0 is used, or when everyone is classified as a user). As in Table C, the best predictive accuracy for use occurs at the 0.5 threshold when 52.2 percent of those classified as users are actually users.

The general impression from this examination of the probability of correct prediction and predictive accuracy is consistent with the goodness of fit measure given in Table A (i.e., proportional reduction in error of 5.1 percent). The sensitivities of the tests based on each threshold level examined are low, as are the predictive accuracies. The model offers some improvement in the ability to predict use and nonuse compared with simple 'all or none' classification schemes, but there is still room for considerable improvement in the fit of the model to the data.

The separate model constructed by adding a functional limitation score (treated as a continuous variable) to the model shown in Table B is not presented here. Functional limitation score in that model had the largest partial correlation with hospitalization of any variable in the model. Yet perceived health status indicator variables continue to have significant, though smaller effects than in the model using all the data, as does the indicator variable for low income persons with poor health status. This suggests that functional limitation is associated with hospitalization in some way that is different from reported health status.

Model for Predicting High Use of Hospital Care

Given that factors have been identified which are predictive of hospitalization, it is of interest to examine whether these same factors or a different set are important in determining high use of hospital care. As described previously, a second logistic regression analysis was conducted to investigate factors related to high use among those reporting at least one overnight hospital stay during the year. Although this second model takes its point of departure at hospitalization, the focus of the first model, it is misleading to consider it a second-stage model since it does not incorporate any of the estimates from the first model.

The dependent variable in this second model is again a dichotomous measure, with high-use persons (i.e., those with 17 or more hospital days reported in 1980) given a value equal to one and all other persons, those with more than one but fewer than 17 days, given a value of zero. Those with no hospitalization are excluded from this analysis. The independent variables are listed in Table E and include the same demographic, income, region-of-residence, health-care-coverage, and perceived-health-status variables as were used in the first model. The rationale for including these same variables in both models is that these factors may not only affect whether any hospital care was obtained—the focus of the first model—but also affect how much care was used, and more specifically whether someone was a high user of hospital care or not.

In addition to these basic independent variables, six diagnostic categories were included in the model and are shown in Table F. They include the disease categories that in other studies have been found to be associated with high-cost hospitalizations and which were also found to account for the largest percent of high-use hospitalizations in the descriptive analyses presented earlier. The six categories are neoplasms, diseases of the circulatory system, diseases of the respiratory system, diseases of the digestive system, diseases of the musculoskeletal system and connective tissue, and injury and poisoning. It would have been preferable to identify persons who not only had these diagnoses, but who also had surgery; these diagnoses combined with surgery have been found to be highly resource-intensive. As a preliminary investigation of the joint effects of surgery and each diagnosis, a set of indicator variables was added to the model that was equal to one if the individual had surgery during 1980 and a selected diagnosis during at least one hospital stay and was equal to zero otherwise. These indicators when added to the model in Table F accounted for only another 0.5 percent of the unexplained error. In addition, approximately 80 percent of the persons hospitalized during the year for neoplasms or a cardiac condition also had surgery during the year. Thus, the addition of the combined surgery and diagnosis indicators to the model contributed little to the explanatory power of the model, and the indicators were not included in the final model.

The estimated logistic regression coefficients and odds ratios from the model are presented in Tables E and F. The high-use logistic regression analysis was based on 2,087 persons with one or more hospital days in 1980, of whom 315 or 15.1 percent were hospitalized 17 or more days in 1980. As a result of the smaller sample size, standard errors of coefficients are higher for this model than for the previous

Table E

Estimated logistic regression coefficients and odds ratios for probability of high use of hospital care among users, by selected independent variables: United States, 1980

Independent variables	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	−2.1128	...	...	...	...	...
Sex						
(Male)	...	...	...	...	...	...
Female	−0.5087	0.1579	−3.2216	0.6013	0.4412	0.8194
Race						
White and other	−0.2447	0.2516	−0.9727	0.7829	0.4782	1.2820
(Black)	...	...	...	...	...	...
Perceived health status						
Excellent	−0.5871	0.1592	−3.6889	0.5559	0.4070	0.7595
(Good)	...	...	...	...	...	...
Fair	0.0730	0.2073	0.3521	1.0757	0.7165	1.6151
Poor	0.8624	0.2792	3.0889	2.3688	1.3705	4.0944
Age						
Under 35 years	−0.7582	0.2377	−3.1899	0.4685	0.2940	0.7465
(35–54 years)	...	...	...	...	...	...
55–74 years	0.0986	0.2678	0.3682	1.1036	0.6530	1.8653
75 years and over	0.6290	0.3702	1.6991	1.8757	0.9079	3.8752
Poverty status[1]						
Below 200 percent of poverty level	0.1967	0.1225	1.6059	1.2174	0.9575	1.5477
200–499 percent of poverty level	−0.0277	0.1405	−0.1971	0.9727	0.7385	1.2811
500–699 percent of poverty level	−0.1159	0.1929	−0.6009	0.8906	0.6102	1.2997
(700 percent of poverty level or more)	...	...	...	...	...	...
Health care coverage						
Multiple public	0.5040	0.3285	1.5342	1.6553	0.8695	3.1515
Single public	0.3899	0.3253	1.1985	1.4768	0.7805	2.7942
Private and public	0.4550	0.2732	1.6656	1.5762	0.9227	2.6923
Private only	0.1308	0.2147	0.6093	1.1397	0.7483	1.7359
(None or other)	...	...	...	...	...	...
Region of residence						
(Northeast)	...	...	...	...	...	...
North Central	−0.1840	0.2241	−0.8210	0.8319	0.5362	1.2908
South	−0.4174	0.2162	−1.9308	0.6588	0.4312	1.0063
West	−0.3941	0.2564	−1.5372	0.6743	0.4080	1.1145

[1]Analysis of variance parameterization.

NOTE: Proportional reduction of error = 18.8 percent.

model, which was based on data for the entire sample population.

There are several interesting contrasts with the results from the first model. The income and health care coverage variables which were so prominent in the first model are not significant predictors of high use. This suggests that although those factors may be important determinants of whether someone becomes a user of hospital services (that is, they are indicators of access), they are not determinants of whether someone who had a hospitalization will be a high user. Other factors, more closely related to need, are prominent in the high-use model. Among the diagnostic categories that were hypothesized to entail the need for large amounts of care, five of the six are strongly and significantly related to high use of hospital days. Moreover, if the high-use logistic regression analysis is carried out without the diagnostic indicators, the coefficients for the remaining variables in the model are quite similar to those shown in Table F. In other words, the diagnostic categories are providing different information about risk of high use of hospital care than the other variables in the model. In addition, this model explains 18.8 percent of the error in predicting the probability of high use, and the diagnostic categories alone account for 6.6 percent of the unexplained error. Thus, including diagnostic-category variables directly improves the explanatory power of the model.

On the other hand, the surgery indicator variable was not significant. This finding is not very surprising, because

Table F

Estimated logistic regression coefficients and odds ratios for probability of high use of hospital care among users, by selected diagnostic categories: United States, 1980

Independent variables	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
Neoplasm						
Present	1.7690	0.3278	5.3966	5.8650	3.0849	11.1505
(Absent)	...	...	...	...	...	...
Circulatory system						
Present	1.1153	0.2015	5.5337	3.0505	2.0550	4.5282
(Absent)	...	...	...	...	...	...
Respiratory system						
Present	0.7006	0.2198	3.1879	2.0150	1.3098	3.0998
(Absent)	...	...	...	...	...	...
Digestive system						
Present	0.0263	0.2032	0.1294	1.0266	0.6894	1.5288
(Absent)	...	...	...	...	...	...
Musculoskeletal system						
Present	1.0892	0.3191	3.4128	2.9719	1.5899	5.5552
(Absent)	...	...	...	...	...	...
Injury or poisoning						
Present	1.3743	0.2326	5.9077	3.9523	2.5051	6.2355
(Absent)	...	...	...	...	...	...
Surgery						
(No)	...	...	...	...	...	...
Yes	0.2624	0.1646	1.5945	1.3000	0.9416	1.7949

NOTE: Proportional reduction of error = 18.8 percent.

it was expected that surgery would be an important factor not through its direct effects, but through interaction with the diagnostic categories.

Sex is significantly related to high use of hospital care in the second model. Being female is strongly and negatively associated with high use when the other variables in the model are taken into account, something that is not readily apparent from the univariate analyses presented earlier. Age, on the other hand, is significant only to the extent that those under 35 years of age are significantly less likely to be high users than the reference category of those 35–54 years of age.

Perceived health status is significantly related to high use, even though the five diagnostic categories described earlier are also significantly related to high use. This suggests that the diagnostic categories capture a different dimension than perceived health status.

As in the hospitalization model, the relative importance of the factors in the logistic regression model may be examined by estimating the probability that a hypothetical individual may be a high user of hospital care once hospitalized. Consider, for example, a black female 35–54 years of age in good health with low income (i.e., below 200 percent of poverty level) and no health care coverage living in the Northeast Region, who is hospitalized with a diagnosis other than the six considered in Table F—and who does not have surgery. The predicted probability that this hypothetical female will have 17 or more hospital days in 1980 is computed in three steps as for the previous model:

(1) summation of appropriate coefficients as

$$-2.1128 - 0.5087 + 0.1967 = -2.4248;$$

(2) exponentiation of this logarithm of the odds as $e^{-2.4248} = 0.0885$; and (3) computation of the predicted probabilities as

$$0.0885 / (1 + 0.0885) = 0.0813.$$

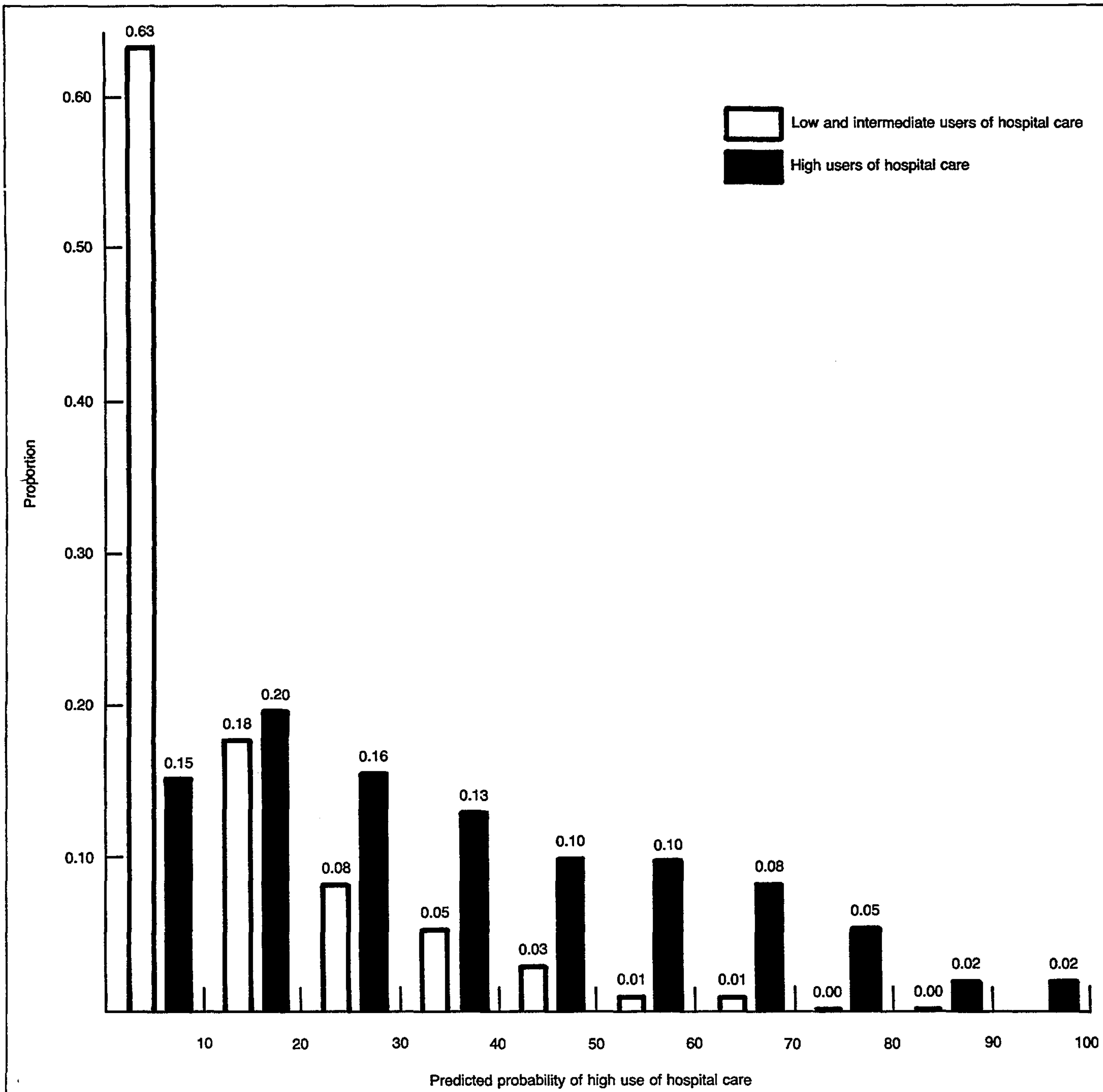
If her health care coverage had been through Medicaid only rather than none, the predicted probability increases to 11.6 percent, a 42-percent increased risk of high use. On the other hand, if she still had no health care coverage, but reported poor instead of good health status, the predicted probability increases to 17.3 percent, a 114-percent increase in risk over the base set of characteristics.

Finally, the importance of the diagnosis is illustrated by considering the same hypothetical female with the original set of characteristics, but with a diagnosis of respiratory disease and surgery during the year while in a hospital. The estimated risk of high use of this individual is 18.8 percent, 2.32 times higher than for the original person.

Again, caution should be exercised in interpreting the predicted probabilities from the model in Tables E and F. The model does not explain all the variation in high use of hospital care, and hence the predicted probabilities are subject to error. The error is also increased by the use of less stable coefficients that were not statistically significant. In addition, these predicted probabilities must be applied to population estimates to assess the importance of the factors considered in the model to high use of inpatient hospital care.

The goodness of fit of the model for predicting high use of hospital care was assessed by examining the distribution of predicted probabilities for high users and all other users (i.e., low and intermediate users). Figure 10 presents the relative frequency distribution of the predicted probabilities from the model for high users and all other users. As before,

Figure 10
Relative frequency distribution of predicted probabilities for high and other users of hospital care: United States, 1980



the model fit is better to the extent that high users are assigned predicted probabilities that are closer to one and other users are assigned values closer to zero.

Although the ability of the model to make such an assignment is clearly less than perfect, there is evidence in Figure 10 that the model is discriminating between these two groups. For example, 97 percent of the low and intermediate users have predicted probabilities less than 0.50; only 74 percent of the high users have such values. Conversely, 26 percent of the high users have predicted probabilities greater than 0.50, compared with only 3 percent of the low and intermediate users. The model is showing some discrimination between high and all other users.

The ability of the model to improve prediction of high use of hospital care can be further understood by examining the probability of correct prediction and predictive accuracy at selected thresholds for the predicted probabilities. Without the model, a simple classification scheme with good predictive power would assign all users to the low or intermediate use group, and since only 15.0 percent of the population were high users, the simple classification scheme would correctly predict low or intermediate use for 85.0 percent of the users. It is of interest, therefore, to examine whether the model can improve the ability to predict correctly the status of users beyond that achieved by the simple classification scheme.

Table G presents the probability of correct classification for high, low or intermediate, and all users at deciles of the predicted probabilities for each user. For example, suppose the scheme is to classify all persons with predicted probabilities greater than 0.1 as high users and all others as low or intermediate users. Among high users, the 0.1 threshold would predict correctly 84.9 percent of the time, whereas for low or intermediate users the probability of a correct prediction is 63.5 percent. The probability of a correct prediction with the 0.1 threshold is 0.667 for all users. As noted previously, the probability of a correct prediction for high users is also the sensitivity of the 0.1 threshold test procedure, and the probability of a correct prediction for the low or intermediate users is the specificity of the test.

Table G

Probability of correct prediction by deciles of predicted probabilities for high users, low and intermediate users, and all users of hospital care: United States, 1980

Threshold of predicted probability	Probability of correct prediction		
	High users	Low and intermediate users	All users
0.0	1.000	0.000	0.150
0.1	0.849	0.635	0.667
0.2	0.652	0.814	0.789
0.3	0.495	0.896	0.836
0.4	0.364	0.946	0.859
0.5	0.265	0.975	0.868
0.6	0.169	0.987	0.864
0.7	0.089	0.996	0.860
0.8	0.037	0.999	0.855
0.9	0.020	1.000	0.853
1.0	0.000	1.000	0.850

Table H

Predictive accuracy of alternative thresholds for predicting high use and low and intermediate use of hospital care: United States, 1980

Threshold of predicted probability	Predictive accuracy	
	High use	Low and intermediate use
0.0	0.150	...
0.1	0.291	0.960
0.2	0.382	0.930
0.3	0.457	0.909
0.4	0.546	0.894
0.5	0.650	0.882
0.6	0.703	0.871
0.7	0.807	0.861
0.8	0.886	0.854
0.9	1.000	0.852
1.0	...	0.850

As the threshold increases, the probability of a correct prediction for high users decreases and that for low or intermediate users increases. For all users, the probability increases to 86.8 percent at a 0.5 threshold and declines to 85.0 percent at higher thresholds. Thus, a threshold of approximately 0.5 will yield the highest probability of a correct prediction for all users, but this probability is dominated by the large number of low or intermediate users relative to the high users. The model for high use predicts high use somewhat better than the model for any use discussed previously.

It is also important to examine the accuracy of the model for predicting high use or low or intermediate use. Table H presents the probability at a given threshold that a user classified by the model as a high or low or intermediate user is actually a high or low or intermediate user. For example, for the threshold of 0.5, 65 percent of those classified as high users under the model are actually high users. Similarly, 88.2 percent of those classified as low or intermediate users at the 0.5 threshold are actually low or intermediate users. As before, the lowest predictive accuracy for high users occurs when every user is classified as a high user (i.e., a threshold of 0.0). For low or intermediate users, the lowest predictive accuracy occurs when all users are classified as low or intermediate users (i.e., a threshold of 1.0).

The results in Tables G and H are improved over those in Tables C and D since the sensitivity of the test procedures at the various thresholds is improved, and the specificity of the test procedures at the various thresholds is improved, and the specificity remains high for most thresholds. But the sensitivity of the various tests remains low and indicates that improved fit of the model is desirable.

As with the first model, a separate model was also constructed that includes the functional limitations variables, and therefore excludes anyone who died or was under 17 years of age.

The logistic regression model with functional limitation added to the model (not shown here) again shows that functional limitation is significantly and positively related to high use. But the health status and age variables are no longer significant in the model with functional limitation present as a predictor. One possible source of this changed relationship may be attributable simply to the dropping of those who are under 17 years of age and those who died. But this finding also suggests that in this population functional limitation is correlated with age and with perceived health status.

Discussion

The analyses in this report once again emphasize the importance of high users of medical care services. A small percent of the users who are high users of a medical service tend to account for a large percent of all such services consumed, for all three medical care services examined. This is most striking for hospital inpatient care. High users of hospital days constituted only 15 percent of those hospitalized and only 1.7 percent of the U.S. civilian noninstitutionalized population, yet they consumed 54 percent of all hospital days and generated 45 percent of all hospital inpatient charges in this population.

High users of each of the three types of medical services share certain characteristics. The univariate analyses show that high users are more likely than low users to be older and poorer, to have lower health status, to have more reported conditions, and to have functional limitations. They are also more likely to have multiple public coverage. This is exemplified by finding that individuals under 65 years of age who are covered by two or more public programs are found to be among high users of hospital care 4.4 times as frequently as they are in the civilian noninstitutionalized population. When all of the variables that appear to be related to high hospital use in univariate analyses are entered simultaneously in a multivariable regression model, significant association is found with respect to only three of the factors: Age, sex, and need as measured by reported health status and the presence of any one of five diagnoses.

Although regression analysis does not address causality and hence does not, in and of itself, permit discussion of the directionality of observed effects, when considered in the context of univariate analyses and the previous literature, the results are not only meaningful and clear but warrant some strong conclusions. These results show that the likelihood of being a high user of hospital inpatient services is increased not by the comprehensiveness of health care coverage but by need for health care, measured both by perceived health status and by the presence of disabling or life-threatening conditions. When individuals covered by Medicaid, two or more public programs, and by a combination of private and public sources are considered, they are found to have an increased probability of being hospitalized at all. However, when the probability of being a high hospital user is investigated, the coverage variables become less important, and in fact fail to attain statistical significance, and need-related factors dominate. Thus it is reasonable to suggest that those who are sicker are more likely to be eligible for Medicaid and are more likely to be covered by two or more public programs. Hence, it is because they are sicker and not because of their coverage that they are more likely to be hospitalized and, once hospitalized, to be high users of hospital inpatient services.

The findings with respect to health status and each of five diagnostic conditions (neoplasms, cardiovascular diseases, respiratory diseases, musculoskeletal diseases and injuries and poisonings) are unambiguous. Poor health status and the presence of any of these five conditions at least doubles the probability that a person will be a high user of hospital inpatient services.

The age effect found might well be an artifact resulting from the way in which the categorical age variables were created. The lower probability of being a higher user of hospital days, once hospitalized, for women may be attributable to the fact that women, particularly in the age range of 17–44 years, are more likely to be hospitalized at all, principally for deliveries. Hence, they constitute a larger percent of the hospitalized population.

The univariate analysis reinforces the dominant role of health status. Although the health status of only 12.9 percent of the civilian noninstitutionalized population as a whole, and only 11.0 percent of those who did not have a hospital inpatient episode in 1980, was reported to be fair or poor, 52.6 percent of those who were high users of hospital days had their health status rated as either fair or poor at the beginning of the survey year. Further, 57.7 percent of high users of hospital inpatient services reported having 6 or more separate medical conditions, compared with 15.2 percent for the entire civilian noninstitutionalized population, and 21.9 percent of those who were hospitalized for only 1 or 2 days during 1980. The findings regarding the relationship between level of disability and high use of hospital inpatient services reinforces this line of reasoning.

In the civilian noninstitutionalized population eligible for the survey for the full year, 82.4 percent had no or minimal functional limitation, and, as one might expect, only 0.6 percent of these persons were high users of hospital inpatient services. At the other end of the disability scale, only 2.1 percent of the reference population had very severe functional limitation (Levels 7 and 8), but 20.2 percent of these persons were high users of hospital days, constituting 21.2 percent of all high users of hospital days. That is, they were found among high hospital users almost 10 times more frequently than in the civilian noninstitutionalized population. Further, the 0.4 percent of the reference population who died during the year comprised 12 percent of all high users of hospital days. It might also be noted that of those with no limitation, only 0.5 percent were high hospital users—a percent that increased to 22.5 for those most severely limited (Level 8) and to 54.2 percent of those who died during the year.

The evidence is persuasive that high hospital inpatient care use is associated with severe illness, functional limitation, and death.

The patterns of use by use level, as well as the differences between the characteristics of persons who are low versus high users of prescribed medications replicates those found for hospital inpatient services.

With respect to ambulatory care, the patterns are somewhat different. High users of ambulatory services are more similar to low users and to nonusers than is the case for hospital days and prescribed medicines with respect to age, family income, education of head of family, perceived health status, and activity limitations. These differences suggest that related, but different, dynamics result in high use for different categories of service. This is confirmed by the finding that only one-third of high users of hospital days were also high users of ambulatory care, while two-thirds of high users of ambulatory visits used no inpatient hospital care whatever, and only 12 percent were high users of hospital days. An important implication of these distinctions is that any gains in reducing high use in one category of service will not necessarily result in comparable gains in another category, and therefore achieving such reductions may require adopting, for each service category, approaches that are tailored to the distinctive characteristics of the category's high users.

On the other hand, nonusers and low users were found to have many similar characteristics, including the key ones of perceived health status and functional limitation. For example, with respect to ambulatory care, 7.4 percent of nonusers and 6.5 percent of low users had their health status reported as fair or poor. Among persons for whom data on functional limitation are available, the civilian noninstitutionalized population 17 years of age or over who survived the year, 88.5 of the nonusers and 88.1 percent of the low users had no reported functional limitation.

When nonuse and low use of hospital inpatient services by insurance status is considered a difference emerges among persons under 65 years of age who had no health care coverage of any kind during the entire year. These persons are found much more frequently in the nonuser group (9.2 percent) than in the low hospital use category (3.8 percent). In fact, the proportion of persons who were not hospitalized during the year is the highest in this group of persons without any health care coverage, 95 percent, as is the proportion who were hospitalized for only one or two days, 1.0 percent. This might well indicate that the lack of health care coverage poses a barrier to the use of inpatient hospital services.

It is worthwhile noting, however, that family income does not differentiate between persons who are nonusers and low users of hospital inpatient services. When arrayed by relative poverty status, between 2.1 percent (200–499 percent of poverty level) and 2.9 percent (below 200 percent poverty level) of persons under 65 years of age are low users of hospital days. At the same relative poverty levels, 89.5 percent and 86.1 percent of persons did not have a hospitalization episode during the year. In fact, the percent of persons with no hospitalization is effectively the same (between 89.4 and 90.6 percent) at all adjusted income levels above 200 percent of the poverty level.

The findings on the characteristics of high users of medical care services are sobering. They indicate that 54.3 percent of hospital days, 32.3 percent of ambulatory visits, and 32.9 percent of prescribed medications were consumed by a very small percent of the U.S. civilian noninstitutionalized population. They also indicate that these high users of medical care services are predominantly sick, functionally limited in their activities, or dying. High use of services, and especially of hospital days, which are the most expensive component of medical care, does not appear to be associated with health care coverage. Although the data used for these analyses did not permit the exploration of the potential effects on level of use of medical resource availabilities and institutional structures, such as the effects of Health Maintenance Organizations, the results indicate that standard cost containment approaches, such as increased patient cost sharing, are not likely to have significant impacts on that large percent of expenditures generated by the small percent of high users.

These findings indicate that high levels of use result from serious illness, severe disability, and death. If this is true, long-range efforts to contain costs will have to attempt to reduce the incidence of the conditions that lead to high use of medical care resources. Alternative means of treatment and new organizational structures will have to be considered in order to reduce the costs of management of conditions that cannot be prevented.

References

Anderson, D. W., and McLaurin, R. L., eds.: Report on the national head and spinal cord injury survey. *J. Neurosurg.* 53: Suppl., Nov. 1980.

Anderson, G. and Knickman, J.: Patterns of expenditures among high utilizers of medical care services: The experience of Medicare beneficiaries from 1974 to 1977. *Med. Care* 22(2): 143–149, Feb. 1984.

Birnbaum, H.: *A National Profile of Catastrophic Illness*. DHEW Pub. No. (PHS) 78–3201. National Center for Health Services Research, Public Health Service. Hyattsville, May 1978a.

Birnbaum, H.: *The Cost of Catastrophic Illness*. Lexington, Mass. D.C. Health and Company, 1978b.

Bloom, B. and Kissick, P.: Home and hospital cost of terminal illness. *Med. Care*, 318(5): 560–564, May 1980.

Budetti, P., McManus, P., Barrand, M., et al.: The costs and effectiveness of neonatal intensive care. *The Implications of Cost-Effectiveness Analysis of Medical Technology*. Background Paper 2, Case Study 10. Office of Technology Assessment, Congress of the United States. Washington, Aug. 1981.

Cancer Care, Inc.: *The Impact, Costs, and Consequences of Catastrophic Illness on Patients and Families*. New York. Cancercare, Inc. and the National Cancer Foundation, 1973.

Casady, R. J.: The survey design and estimation methodology for the NMCUES. *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.

Congressional Budget Office: *Catastrophic Health Insurance*. Washington. U.S. Government Printing Office, 1977.

Congressional Budget Office: *Catastrophic Medical Expenses: Patterns in the Non-elderly, Non-poor Population*. Washington. U.S. Government Printing Office, Dec. 1982.

Cullen, D. J., Ferrara, L. C., Briggs, B. A., et al.: Survival, hospitalization charges and follow-up results in critically ill patients. *N. Engl. J. Med.* 294(18): 982–287, April 29, 1976.

Detsky, A. S., Stricker, S. C., Mulley, A. G., et al.: Prognosis, survival, and the expenditure of hospital resources for patients in an intensive-care unit. *N. Engl. J. Med.* 305(12): 667–672, Sept. 17, 1981.

Eldred, C. A., Woolsey, T. D., Simmons, W. R., et al.: *A Report of the Survey of Intracranial Neoplasms*. Contract No. No–1–NS–4–2336. Rockville, Md. Westat, Inc., 1977.

Forthofer, R. N., Lairson, D. R., and Glasser, J. H.: Catastrophic health insurance and HMO's. *Soc. Sci. Med.* 16(20): 1775–1779, 1982.

Gibbs, J., and Newman, J.: *Study of Health Services Used and Costs Incurred During the Last Six Months of a Terminal Illness*. Chicago. Blue Cross and Blue Shield Association. Nov. 1982.

Kish, L.: *Survey Sampling*. New York. J. W. Wiley and Sons, Inc., 1965.

Kobrinski, E. J., and Matteson, A. L.: Characteristics of high cost treatment in acute care facilities. *Inquiry* 18(2): 179–184, Summer 1981.

Lepkowski, J. M., Landis, J. R., Parsons, P. E., et al.: A statistical methodology for analyzing data from a complex sample survey, the National Medical Care Utilization and Expenditure Survey. *National Medical Care Utilization and Expenditure Survey.* Series A. National Center for Health Statistics, Public Health Service. Hyattsville, Md. To be published.

Lepkowski, J. M., Stehouwer, S., and Landis, J. R.: Strategies for the analysis of imputed data in a sample survey. *Proceedings of the Section on Survey Research Methods, American Statistical Association*, 1984.

Lubitz, J., and Prihoda, R.: The use and costs of Medicare services in the last two years of life. *Health Care Fin. Rev.* 5(3): 117–131, Spring 1984.

McCall, N.; Utilization and costs of Medicare services by beneficiaries in their last year of life. *Med. Care* 22(4): 329–342, Apr. 1984.

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. National Center for Health Statistics, Public Health Service. Hyattsville, Md. To be published.

Schroeder, S. A., Showstack, J. A., and Roberts, H. E.: Frequency and clinical description of high cost patients in 17 acute care hospitals. *N. Engl. J. Med.* 300(23): 1306–1309, June 7, 1979.

Scitovsky, A. A.: The high cost of dying: what do the data show? *Milbank Mem. Fund Q.* 62(4): 591–608, Fall 1984.

Trapnell, G. R.: The rising cost of catastrophic illness. National Center for Health Services Research, DHEW, Dec. 1977.

Webb, S., Berzins, E., Wingardner, T. S., et al.: First year hospitalization costs for the spinal cord injured patient. *Paraplegia*, 15: 311–318, 1977.

Weinfield, F. D., ed.: The national survey of stroke. *Stroke* 12(2): Suppl. No. 1, Mar.–Apr. 1981.

Zook, C. J., and Moore, F. D.: High-cost users of medical care. *N. Engl. J. Med.* 302(18) 996–1002, May 1, 1980.

List of Detailed Tables

Table 1

Numbers and percent distributions and means for service units used and charges for service, by type of service and level of use: United States, 1980

Type of service and level of use	Population in thousands	Percent distribution: All persons	Percent distribution: Users	Service unit used: Number in thousands	Service unit used: Percent distribution	Service unit used: Per capita	Charges for service: Amount in millions	Charges for service: Percent distribution	Charges for service: Per capita	Charges for service: Mean per unit of service
Hospital days										
Total	222,824	100.0	100.0	271,447	100.0	1.2	$87,138	100.0	$391	$321
Level of use										
0	197,283	88.5	...	...	...	...	[1]1,057	1.2	...	...
Low	5,242	2.4	20.5	9,120	3.4	1.7	5,251	6.0	1,002	576
Intermediate	16,462	7.4	64.5	114,770	42.3	7.0	41,473	47.6	2,519	361
High	3,837	1.7	15.0	147,557	54.3	38.5	39,357	45.2	10,257	267
Ambulatory care visits										
Total	222,824	100.0	100.0	1,150,642	100.0	5.2	37,504	100.0	168	33
Level of use										
0	46,716	21.0	...	...	...	...	...	...	...	...
Low	37,921	17.0	21.5	38,665	3.3	1.0	1,344	3.6	35	35
Intermediate	128,163	57.5	72.8	740,724	64.4	5.8	25,394	67.7	198	34
High	10,024	4.5	5.7	371,253	32.3	37.0	10,766	28.7	1,074	29
Prescribed medicine acquisitions										
Total	222,824	100.0	100.0	1,026,767	100.0	4.6	7,831	100.0	35	8
Level of use										
0	83,814	37.6	...	...	...	...	...	...	...	...
Low	30,492	13.7	21.9	30,954	3.0	1.0	204	2.6	7	7
Intermediate	100,345	45.0	72.2	658,061	64.1	6.6	4,947	63.2	49	8
High	8,173	3.7	5.9	337,752	32.9	41.3	2,680	34.2	328	8

[1]Represents charges for persons who were counted as having at least 1 hospital stay but never were in the hospital overnight.

Table 2

Numbers and percent distributions of persons by race, sex, and age, according to level of hospital use: United States, 1980

Race, sex, and age	Total	Level of hospital use			
		0	Low	Intermediate	High
Race		Number in thousands			
All races	222,824	197,283	5,242	16,462	3,837
Black	26,046	22,952	599	1,978	516
White and other	196,779	174,330	4,644	14,484	3,321
Sex and age					
Both sexes, all ages	222,824	197,283	5,242	16,462	3,837
Under 17 years	61,575	57,050	1,681	2,713	131
17-24 years	32,886	29,031	993	2,647	215
25-34 years	35,827	31,493	970	2,991	373
35-44 years	25,489	23,093	476	1,531	389
45-54 years	22,443	19,797	408	1,737	500
55-64 years	21,135	18,404	286	1,797	648
65-74 years	15,165	12,190	240	1,927	808
75 years and over	8,305	6,224	189	1,119	773
Male, all ages	107,481	96,929	2,185	6,404	1,963
Under 17 years	31,585	29,091	913	1,521	60
17-24 years	15,752	14,647	295	688	123
25-34 years	17,506	16,288	245	761	211
35-44 years	12,318	11,319	249	579	170
45-54 years	10,859	9,697	154	711	297
55-64 years	9,970	8,618	133	834	386
65-74 years	6,486	5,012	103	987	383
75 years and over	3,006	2,257	93	323	333
Female, all ages	115,344	100,354	3,058	10,058	1,874
Under 17 years	29,990	27,959	768	1,192	71
17-24 years	17,134	14,384	698	1,959	92
25-34 years	18,321	15,205	724	2,230	162
35-44 years	13,171	11,773	227	952	219
45-54 years	11,584	10,101	255	1,026	203
55-64 years	11,165	9,786	153	964	262
65-74 years	8,679	7,178	137	940	424
75 years and over	5,299	3,968	96	796	440

Table 2 – continued

Numbers and percent distributions of persons by race, sex, and age, according to level of hospital use: United States, 1980

Race, sex, and age	Total	Level of hospital use			
		0	Low	Intermediate	High
Race		Percent distribution			
All races	100.0	100.0	100.0	100.0	100.0
Black	11.7	11.6	11.4	12.0	13.5
White and other	88.3	88.4	88.6	88.0	86.5
Sex and age					
Both sexes, all ages	100.0	100.0	100.0	100.0	100.0
Under 17 years	27.6	28.9	32.1	16.5	3.4
17–24 years	14.8	14.7	18.9	16.1	5.6
25–34 years	16.1	16.0	18.5	18.2	9.7
35–44 years	11.4	11.7	9.1	9.3	10.1
45–54 years	10.1	10.0	7.8	10.6	13.0
55–64 years	9.5	9.3	5.5	10.9	16.9
65–74 years	6.8	6.2	4.6	11.7	21.1
75 years and over	3.7	3.2	3.6	6.8	20.1
Male, all ages	100.0	100.0	100.0	100.0	100.0
Under 17 years	29.4	30.0	41.8	23.8	3.0
17–24 years	14.7	15.1	13.5	10.7	6.3
25–34 years	16.3	16.8	11.2	11.9	10.8
35–44 years	11.5	11.7	11.4	9.0	8.7
45–54 years	10.1	10.0	7.0	11.1	15.1
55–64 years	9.3	8.9	6.1	13.0	19.6
65–74 years	6.0	5.2	4.7	15.4	19.5
75 years and over	2.8	2.3	4.3	5.1	17.0
Female, all ages	100.0	100.0	100.0	100.0	100.0
Under 17 years	26.0	27.9	25.1	11.9	3.8
17–24 years	14.9	14.3	22.8	19.5	4.9
25–34 years	15.9	15.1	23.7	22.2	8.6
35–44 years	11.4	11.7	7.4	9.5	11.7
45–54 years	10.0	10.1	8.3	10.2	10.8
55–64 years	9.7	9.8	5.0	9.6	14.0
65–74 years	7.5	7.2	4.5	9.3	22.6
75 years and over	4.6	4.0	3.2	7.9	23.5

Table 3

Numbers and percent distributions of persons by race, sex, and age, according to level of ambulatory care use: United States, 1980

Race, sex, and age	Total	Level of ambulatory care use: 0	Low	Intermediate	High
Race			Number in thousands		
All races	222,824	46,716	37,921	128,163	10,024
Black	26,046	7,120	4,852	13,196	878
White and other	196,779	39,596	33,069	114,968	9,146
Sex and age					
Both sexes, all ages	222,824	46,716	37,921	128,163	10,024
Under 17 years	61,575	12,741	12,504	34,911	1,420
17-24 years	32,886	7,220	6,418	18,281	967
25-34 years	35,827	7,954	6,373	20,038	1,463
35-44 years	25,489	6,329	4,377	13,325	1,458
45-54 years	22,443	5,342	3,326	12,521	1,253
55-64 years	21,135	3,573	2,781	13,543	1,237
65-74 years	15,165	2,514	1,466	9,819	1,365
75 years and over	8,305	1,042	676	5,725	862
Male, all ages	107,481	27,573	19,710	56,431	3,766
Under 17 years	31,585	6,435	6,459	17,921	769
17-24 years	15,752	4,777	3,625	7,009	342
25-34 years	17,506	5,520	3,555	7,888	543
35-44 years	12,318	3,971	2,048	5,809	490
45-54 years	10,859	3,127	1,730	5,513	489
55-64 years	9,970	1,967	1,354	6,268	381
65-74 years	6,486	1,277	647	4,009	552
75 years and over	3,006	499	292	2,014	201
Female, all ages	115,344	19,143	18,211	71,732	6,258
Under 17 years	29,990	6,305	6,045	16,989	651
17-24 years	17,134	2,443	2,794	11,272	625
25-34 years	18,321	2,434	2,818	12,149	920
35-44 years	13,171	2,359	2,329	7,516	967
45-54 years	11,584	2,215	1,596	7,009	765
55-64 years	11,165	1,607	1,427	7,275	857
65-74 years	8,679	1,238	819	5,809	813
75 years and over	5,299	543	383	3,712	661

Table 3 – continued

Numbers and percent distributions of persons by race, sex, and age, according to level of ambulatory care use: United States, 1980

Race, sex, and age	Total	Level of ambulatory care use: 0	Low	Intermediate	High
Race			Percent distribution		
All races	100.0	100.0	100.0	100.0	100.0
Black	11.7	15.2	12.8	10.3	8.8
White and other	88.3	84.8	87.2	89.7	91.2
Sex and age					
Both sexes, all ages	100.0	100.0	100.0	100.0	100.0
Under 17 years	27.6	27.3	33.0	27.2	14.2
17–24 years	14.8	15.5	16.9	14.3	9.6
25–34 years	16.1	17.0	16.8	15.6	14.6
35–44 years	11.4	13.5	11.5	10.4	14.5
45–54 years	10.1	11.4	8.8	9.8	12.5
55–64 years	9.5	7.6	7.3	10.6	12.3
65–74 years	6.8	5.4	3.9	7.7	13.6
75 years and over	3.7	2.2	1.8	4.5	8.6
Male, all ages	100.0	100.0	100.0	100.0	100.0
Under 17 years	29.4	23.3	32.8	31.8	20.4
17–24 years	14.7	17.3	18.4	12.4	9.1
25–34 years	16.3	20.0	18.0	14.0	14.4
35–44 years	11.5	14.4	10.4	10.3	13.0
45–54 years	10.1	11.4	8.8	9.8	13.0
55–64 years	9.3	7.1	6.9	11.1	10.1
65–74 years	6.0	4.6	3.3	7.1	14.7
75 years and over	2.8	1.8	1.5	3.5	5.3
Female, all ages	100.0	100.0	100.0	100.0	100.0
Under 17 years	26.0	32.9	33.2	23.7	10.4
17–24 years	14.9	12.8	15.3	15.7	10.0
25–34 years	15.9	12.7	15.5	16.9	14.7
35–44 years	11.4	12.3	12.8	10.5	15.5
45–54 years	10.0	11.6	8.8	9.8	12.2
55–64 years	9.7	8.4	7.8	10.1	13.7
65–74 years	7.5	6.5	4.5	8.1	13.0
75 years and over	4.6	2.8	2.1	5.2	10.6

Table 4

Numbers and percent distributions of persons by race, sex, and age, according to level of prescribed medicine use: United States, 1980

Race, sex, and age	Total	Level of prescribed medicine use: 0	Low	Intermediate	High
Race		Number in thousands			
All races	222,824	83,814	30,492	100,345	8,173
Black	26,046	12,172	3.726	9,373	774
White and other	196,779	71,642	26,766	90,972	7,399
Sex and age					
Both sexes, all ages	222,824	83,814	30,492	100,345	8,173
Under 17 years	61,575	27,241	10,635	23,518	182
17-24 years	32,886	13,568	5,469	13.822	26
25-34 years	35,827	13,747	5,458	16,246	375
35-44 years	25,489	10,158	3,526	11,235	569
45-54 years	22,443	8,336	2,168	10,607	1,332
55-64 years	21,135	5,871	1,864	11,325	2.075
65-74 years	15,165	3,596	983	8,584	2,001
75 years and over	8.305	1,296	388	5,008	1,612
Male, all ages	107,481	47,729	15,648	41,131	2,973
Under 17 years	31,585	14,056	5,309	12,105	115
17-24 years	15,752	8,469	2,712	4.545	26
25-34 years	17,506	8,841	2,751	5,745	169
35-44 years	12,318	5,780	1.937	4,438	162
45-54 years	10,859	4,832	1,204	4,289	534
55-64 years	9,970	3,276	1,070	4,918	706
65-74 years	6,486	1,888	532	3,373	692
75 years and over	3,006	586	133	1,718	569
Female, all ages	115,344	36,086	14,844	59,214	5,200
Under 17 years	29,990	13,185	5,326	11,413	66
17-24 years	17,134	5,100	2,756	9,278	-
25-34 years	18,321	4,906	2,707	10,502	206
35-44 years	13,171	4,378	1,589	6,797	407
45-54 years	11,584	3,504	964	6,318	798
55-64 years	11,165	2,595	795	6,406	1,369
65-74 years	8,679	1,708	451	5,211	1,309
75 years and over	5,299	710	256	3.290	1,043

Table 4 – continued

Numbers and percent distributions of persons by race, sex, and age, according to level of prescribed medicine use: United States, 1980

Race, sex, and age	Total	Level of prescribed medicine use: 0	Low	Intermediate	High
Race			Percent distribution		
All races	100.0	100.0	100.0	100.0	100.0
Black	11.7	14.5	12.2	9.3	9.5
White and other	88.3	85.5	87.8	90.7	90.5
Sex and age					
Both sexes, all ages	100.0	100.0	100.0	100.0	100.0
Under 17 years	27.6	32.5	34.9	23.4	2.2
17–24 years	14.8	16.2	17.9	13.8	0.3
25–34 years	16.1	16.4	17.9	16.2	4.6
35–44 years	11.4	12.1	11.6	11.2	7.0
45–54 years	10.1	9.9	7.1	10.6	16.3
55–64 years	9.5	7.0	6.1	11.3	25.4
65–74 years	6.8	4.3	3.2	8.6	24.5
75 years and over	3.7	1.5	1.2	5.0	19.8
Male, all ages	100.0	00.0	100.0	100.0	100.0
Under 17 years	29.4	29.4	33.9	29.4	3.9
17–24 years	14.7	17.7	17.3	11.0	0.9
25–34 years	16.3	18.5	17.6	14.0	5.7
35–44 years	11.5	12.1	12.4	10.8	5.4
45–54 years	10.1	10.1	7.7	10.4	18.0
55–64 years	9.3	6.9	6.8	12.0	23.7
65–74 years	6.0	4.0	3.4	8.2	23.3
75 years and over	2.8	1.2	0.8	4.2	19.1
Female, all ages	100.0	100.0	100.0	100.0	100.0
Under 17 years	26.0	36.5	35.9	19.3	1.3
17–24 years	14.9	14.1	18.6	15.7	–
25–34 years	15.9	13.6	18.2	17.7	4.0
35–44 years	11.4	12.1	10.7	11.5	7.8
45–54 years	10.0	9.7	6.5	10.7	15.3
55–64 years	9.7	7.2	5.4	10.8	26.3
65–74 years	7.5	4.7	3.0	8.8	25.2
75 years and over	4.6	2.0	1.8	5.6	20.0

Table 5

Numbers and percent distributions of persons by family income, poverty status, and education of head of family, according to level of hospital use: United States, 1980

Family income, poverty status, and education of head of family	Total	Level of hospital use: 0	Low	Intermediate	High
Family income			Number in thousands		
Less than $5,000	16,225	13,165	664	1,704	693
$5,000–14,999	58,157	50,116	1,471	5,101	1,470
$15,000–34,999	103,400	92,908	2,123	7,219	1,149
$35,000 or more	45,043	41,094	985	2,438	526
Poverty status					
Below 200 percent of poverty level	70,023	60,257	2,021	5,947	1,798
200–499 percent of poverty level	115,466	103,352	2,371	8,179	1,564
500–699 percent of poverty level	23,872	21,634	549	1,445	243
700 percent of poverty level or more	13,464	12,040	302	891	232
Education of head of family[1]					
None or some elementary	36,837	31,568	747	3,443	1,078
Some high school	35,152	30,745	813	2,787	808
High school graduate	78,063	69,472	1,944	5,581	1,067
Some college	34,849	31,303	726	2,307	513
College graduate	37,784	34,056	1,012	2,345	371
Family income			Percent distribution		
Total	100.0	100.0	100.0	100.0	100.0
Less than $5,000	7.3	6.7	12.7	10.4	18.1
$5,000–14,999	26.1	25.4	28.1	31.0	38.3
$15,000–34,999	46.4	47.1	40.5	43.9	29.9
$35,000 or more	20.2	20.8	18.8	14.8	13.7
Poverty status					
Total	100.0	100.0	100.0	100.0	100.0
Below 200 percent of poverty level	31.4	30.5	38.5	36.1	46.9
200–499 percent of poverty level	51.8	52.4	45.2	49.7	40.8
500–699 percent of poverty level	10.7	11.0	10.5	8.8	6.3
700 percent of poverty level or more	6.0	6.1	5.8	5.4	6.0
Education of head of family[1]					
Total	100.0	100.0	100.0	100.0	100.0
None or some elementary	16.5	16.0	14.3	20.9	28.1
Some high school	15.8	15.6	15.5	16.9	21.1
High school graduate	35.0	35.2	37.1	33.9	27.8
Some college	15.6	15.9	13.9	14.0	13.4
College graduate	17.0	17.3	19.3	14.2	9.7

[1]Includes only families with heads 17 years of age and over.

Table 6

Numbers and percent distributions of persons by family income, poverty status, and education of head of family, according to level of ambulatory care use: United States, 1980

Family income, poverty status, and education of head of family	Total	Level of ambulatory care use: 0	Low	Intermediate	High
Family income			Number in thousands		
Less than $5,000	16,225	2,858	1,984	10,254	1,130
$5,000-14,999	58,157	12,786	9,640	32,795	2,936
$15,000-34,999	103,400	21,719	18,472	59,319	3,890
$35,000 or more	45,043	9,353	7,826	25,795	2,069
Poverty status					
Below 200 percent of poverty level	70,023	16,293	11,448	38,947	3,336
200-499 percent of poverty level	115,466	23,526	20,517	66,902	4,520
500-699 percent of poverty level	23,872	4,654	3,824	14,161	1,233
700 percent of poverty level or more	13,464	2,243	2,133	8,153	935
Education of head of family[1]					
None or some elementary	36,837	9,510	5,442	20,086	1,799
Some high school	35,152	8,820	5,990	19,181	1,160
High school graduate	78,063	15,452	14,335	44,882	3,394
Some college	34,849	6,650	5,536	20,953	1,710
College graduate	37,784	6,257	6,577	22,989	1,961
Family income			Percent distribution		
Total	100.0	100.0	100.0	100.0	100.0
Less than $5,000	7.3	6.1	5.2	8.0	11.3
$5,000-14,999	26.1	27.4	25.4	25.6	29.3
$15,000-34,999	46.4	46.5	48.7	46.3	38.8
$35,000 or more	20.2	20.0	20.6	20.1	20.6
Poverty status					
Total	100.0	100.0	100.0	100.0	100.0
Below 200 percent of poverty level	31.4	34.9	30.2	30.4	33.3
200-499 percent of poverty level	51.8	50.4	54.1	52.2	45.1
500-699 percent of poverty level	10.7	10.0	10.1	11.0	12.3
700 percent of poverty level or more	6.0	4.8	5.6	6.4	9.3
Education of head of family[1]					
Total	100.0	100.0	100.0	100.0	100.0
None or some elementary	16.5	20.4	14.4	15.7	17.9
Some high school	15.8	18.9	15.8	15.0	11.6
High school graduate	35.0	33.1	37.8	35.0	33.9
Some college	15.6	14.2	14.6	16.3	17.1
College graduate	17.0	13.4	17.3	17.9	19.6

[1]Includes families with heads 17 years of age and over.

Table 7

Numbers and percent distributions of persons by family income, poverty status, and education of head of family according to level of prescribed medicine use: United States, 1980

Family income, poverty status, and education of head of family	Total	0	Low	Intermediate	High
		Level of prescribed medicine use			
Family income			Number in thousands		
Less than $5,000	16,225	4,805	1,766	8,116	1,539
$5,000-14,999	58,157	21,331	6,762	26,839	3,225
$15,000-34,999	103,400	40,108	15,598	45,328	2,366
$35,000 or more	45,043	17,571	6,366	20,063	1,043
Poverty status					
Below 200 percent of poverty level	70,023	27,534	8,790	30,031	3,669
200-499 percent of poverty level	115,466	43,448	16,564	51,939	3,514
500-699 percent of poverty level	23,872	8,344	3,392	11,610	525
700 percent of poverty level or more	13,464	4,488	1,746	6,765	465
Education of head of family[1]					
None or some elementary	36,837	14,126	4,044	15,778	2,889
Some high school	35,152	14,174	4,532	14,957	1,488
High school graduate	78,063	29,602	11,002	35,345	2,114
Some college	34,849	12,265	5,272	16,467	846
College graduate	37,784	13,576	5,621	17,751	836
Family income			Percent distribution		
Total	100.0	100.0	100.0	100.0	100.0
Less than $5,000	7.3	5.7	5.8	8.1	18.8
$5,000-14,999	26.1	25.5	22.2	26.7	39.5
$15,000-34,999	46.4	47.9	51.2	45.2	29.0
$35,000 or more	20.2	21.0	20.9	20.0	12.8
Poverty status					
Total	100.0	100.0	100.0	100.0	100.0
Below 200 percent of poverty level	31.4	32.9	28.8	29.9	44.9
200-499 percent of poverty level	51.8	51.8	54.3	51.8	43.0
500-699 percent of poverty level	10.7	10.0	11.1	11.6	6.4
700 percent of poverty level or more	6.0	5.4	5.7	6.7	5.7
Education of head of family[1]					
Total	100.0	100.0	100.0	100.0	100.0
None or some elementary	16.5	16.9	13.3	15.7	35.3
Some high school	15.8	16.9	14.9	14.9	18.2
High school graduate	35.0	35.3	36.1	35.2	25.9
Some college	15.6	14.6	17.3	16.4	10.3
College graduate	17.0	16.2	18.4	17.7	10.2

[1]Includes families with heads 17 years of age and over.

Table 8

Number and percent distribution of persons by region, according to level of hospital use: United States, 1980

Region	Total	Level of hospital use: 0	Low	Intermediate	High
		Number in thousands			
Northeast	46,899	41,788	930	3,259	921
North Central	59,257	51,950	1,447	4,782	1,079
South	69,475	61,059	1,756	5,535	1,126
West	47,194	42,486	1,110	2,886	711
		Percent distribution			
Total	100.0	100.0	100.0	100.0	100.0
Northeast	21.0	21.2	17.7	19.8	24.0
North Central	26.6	26.3	27.6	29.1	28.1
South	31.2	31.0	33.5	33.6	29.3
West	21.2	21.5	21.2	17.5	18.5

Table 9

Number and percent distribution of persons by region, according to level of ambulatory care use: United States, 1980

Region	Total	Level of ambulatory care use: 0	Low	Intermediate	High
		Number in thousands			
Northeast	46,899	9,350	7,793	27,315	2,441
North Central	59,257	11,091	9,969	35,572	2,625
South	69,475	16,255	12,526	38,607	2,087
West	47,194	10,020	7,633	26,669	2,871
		Percent distribution			
Total	100.0	100.0	100.0	100.0	100.0
Northeast	21.0	20.0	20.5	21.3	24.4
North Central	26.6	23.7	26.3	27.8	26.2
South	31.2	34.8	33.0	30.1	20.8
West	21.2	21.4	20.1	20.8	28.6

Table 10

Number and percent distribution of persons by region, according to level of prescribed medicine use: United States, 1980

Region	Total	Level of prescribed medicine use: 0	Low	Intermediate	High
		Number in thousands			
Northeast	46,899	18,410	6,022	20,918	1,548
North Central	59,257	21,713	8,591	26,836	2,118
South	69,475	25,206	9,090	31,892	3,287
West	47,194	18,485	6,789	20,699	1,221
		Percent distribution			
Total	100.0	100.0	100.0	100.0	100.0
Northeast	21.0	22.0	19.8	20.8	18.9
North Central	26.6	25.9	28.2	26.7	25.9
South	31.2	30.1	29.8	31.8	40.2
West	21.2	22.1	22.3	20.6	14.9

Table 11

Number and percent distribution of persons under 65 years of age by type of health care coverage, according to level of hospital use: United States, 1980

Health care coverage	All persons	Level of hospital use: 0	Low	Intermediate	High
Coverage all year		Number in thousands			
Single source:					
Private insurance	129,515	117,224	2,970	8,158	1,164
Medicaid	9,029	7,618	314	946	151
Other public programs	3,348	2,918	132	199	98
More than 1 source:					
Private and public	15,912	13,607	482	1,431	392
2 or more public programs	4,341	3,162	215	745	219
Other					
Covered for part of year only	19,869	17,856	519	1,324	170
No coverage all year	17,342	16,482	182	614	63
		Percent distribution			
Total	100.0	100.0	100.0	100.0	100.0
Coverage all year					
Single source:					
Private insurance	65.0	65.5	61.7	60.8	51.6
Medicaid	4.5	4.3	6.5	7.1	6.7
Other public programs	1.7	1.6	2.7	1.5	4.4
More than 1 source:					
Private and public	8.0	7.6	10.0	10.7	17.4
2 or more public programs	2.2	1.8	4.5	5.5	9.7
Other					
Covered for part of year only	10.0	10.0	10.8	9.9	7.5
No coverage all year	8.7	9.2	3.8	4.6	2.8

Table 12

Number and percent distribution of persons under 65 years of age by type of health care coverage, according to level of ambulatory care use: United States, 1980

Health care coverage	All persons	Level of ambulatory care use: 0	Low	Intermediate	High
Coverage all year	Number in thousands				
Single source:					
Private insurance	129,515	26,027	23,455	75,083	4,951
Medicaid	9,029	1,863	1,600	5,118	448
Other public programs	3,348	835	773	1,581	159
More than 1 source:					
Private and public	15,912	2,368	2,136	10,280	1,128
2 or more public programs	4,341	329	504	3,146	362
Other					
Covered for part of year only	19,869	4,944	3,584	10,806	535
No coverage all year	17,342	6,794	3,728	6,605	215
	Percent distribution				
Total	100.0	100.0	100.0	100.0	100.0
Coverage all year					
Single source:					
Private insurance	65.0	60.3	65.6	66.7	63.5
Medicaid	4.5	4.3	4.5	4.5	5.7
Other public programs	1.7	1.9	2.2	1.4	2.0
More than 1 source:					
Private and public	8.0	5.5	6.0	9.1	14.5
2 or more public programs	2.2	0.8	1.4	2.8	4.6
Other					
Covered for part of year only	10.0	11.5	10.0	9.6	6.9
No coverage all year	8.7	15.7	10.4	5.9	2.8

Table 13

Number and percent distribution of persons under 65 years of age by health care coverage, according to level of prescribed medicine use: United States, 1980

Health care coverage	All persons	Level of prescribed medicine use: 0	Low	Intermediate	High
	Number in thousands				
Coverage all year					
Single source:					
Private insurance	129,515	49,056	19,670	58,394	2,395
Medicaid	9,029	3,638	1,433	3,669	289
Other public programs	3,348	1,488	333	1,309	218
More than 1 source:					
Private and public	15,912	5,077	2,105	7,818	912
2 or more public programs	4,341	964	627	2,338	411
Other					
Covered for part of year only	19,869	8,636	2,997	7,999	238
No coverage all year	17,342	10,063	1,956	5,226	97
	Percent distribution				
Total	100.0	100.0	100.0	100.0	100.0
Coverage all year					
Single source:					
Private insurance	65.0	62.2	67.5	67.3	52.5
Medicaid	4.5	4.6	4.9	4.2	6.3
Other public programs	1.7	1.9	1.1	1.5	4.8
More than 1 source:					
Private and public	8.0	6.4	7.2	9.0	20.0
2 or more public programs	2.2	1.2	2.2	2.7	9.0
Other					
Covered for part of year only	10.0	10.9	10.3	9.2	5.2
No coverage all year	8.7	12.8	6.7	6.0	2.1

Table 14

Number and percent distribution of persons 65 years of age and over by type of Medicare coverage, according to level of hospital use: United States, 1980

Type of Medicare coverage	All persons	Level of hospital use: 0	Low	Intermediate	High
		Number in thousands			
Covered by Medicare					
Medicare only	3,836	3,193	63	403	177
Medicare and private coverage	15,681	12,123	310	2,172	1,076
Medicare and other nonprivate coverage	2,647	1,866	41	423	318
Not covered by Medicare					
Other coverage	1,033	973	15	35	10
No coverage at all	272	260	–	12	0
		Percent distribution			
Total	100.0	100.0	100.0	100.0	100.0
Covered by Medicare					
Medicare only	16.3	17.3	14.6	13.2	11.2
Medicare and private coverage	66.8	65.8	72.3	71.3	68.1
Medicare and other nonprivate coverage	11.3	10.1	9.5	13.9	20.1
Not covered by Medicare					
Other coverage	4.4	5.3	3.6	1.2	0.6
No coverage at all	1.2	1.4	–	0.4	0.0

Table 15

Number and percent distribution of persons 65 years of age and over by type of Medicare coverage, according to level of ambulatory care use: United States, 1980

Type of Medicare coverage	All persons	Level of ambulatory care use: 0	Low	Intermediate	High
		Number in thousands			
Covered by Medicare					
Medicare only	3,836	997	361	2,265	212
Medicare and private coverage	15,681	1,853	1,381	10,802	1,645
Medicare and other nonprivate coverage	2,647	198	231	1,857	361
Not covered by Medicare					
Other coverage	1,033	322	156	546	10
No coverage at all	272	187	12	74	–
		Percent distribution			
Total	100.0	100.0	100.0	100.0	100.0
Covered by Medicare					
Medicare only	16.3	28.0	16.9	14.6	9.5
Medicare and private coverage	66.8	52.1	64.5	69.5	73.9
Medicare and other nonprivate coverage	11.3	5.6	10.8	11.9	16.2
Not covered by Medicare					
Other coverage	4.4	9.0	7.3	3.5	0.4
No coverage at all	1.2	5.2	0.6	0.5	–

Table 16

Number and percent distribution of persons 65 years of age and over by type of Medicare coverage, according to level of prescribed medicine use: United States, 1980

Type of Medicare coverage	All persons	Level of prescribed medicine use			
		0	Low	Intermediate	High
	Number in thousands				
Covered by Medicare					
Medicare only	3,836	1,267	261	1,896	412
Medicare and private coverage	15,681	2,847	897	9,341	2,597
Medicare and other nonprivate coverage	2,647	291	175	1,663	518
Not covered by Medicare					
Other coverage	1,033	316	25	617	75
No coverage at all	272	172	13	76	11
	Percent distribution				
Total	100.0	100.0	100.0	100.0	100.0
Covered by Medicare					
Medicare only	16.3	25.9	19.1	13.9	11.4
Medicare and private coverage	66.8	58.2	65.4	68.7	71.9
Medicare and other nonprivate coverage	11.3	5.9	12.8	12.2	14.3
Not covered by Medicare					
Other coverage	4.4	.6.5	1.8	4.5	2.1
No coverage at all	1.2	3.5	1.0	0.6	0.3

Table 17

Number and percent distribution of persons by perceived health status, according to level of hospital use: United States, 1980

Perceived health status	All persons	Level of hospital use			
		0	Low	Intermediate	High
	Number in thousands				
Excellent	111,641	102,687	2,503	5,828	623
Good	82,293	72,953	2,011	6,131	1,197
Fair	20,834	16,300	617	3,016	900
Poor	8,057	5,342	111	1,487	1,116
	Percent distribution				
Total	100.0	100.0	100.0	100.0	100.0
Excellent	50.1	52.1	47.7	35.4	16.2
Good	36.9	37.0	38.4	37.2	31.2
Fair	9.3	8.3	11.8	18.3	23.5
Poor	3.6	2.7	2.1	9.0	29.1

Table 18

Number and percent distribution of persons by perceived health status, according to level of ambulatory care use: United States, 1980

Perceived health status	All persons	Level of ambulatory care use			
		0	Low	Intermediate	High
		Number in thousands			
Excellent	111,641	26,724	22,302	60,174	2,440
Good	82,293	16,555	13,176	48,768	3,794
Fair	20,834	2,788	1,963	13,796	2,286
Poor	8,057	648	479	5,425	1,505
		Percent distribution			
Total	100.0	100.0	100.0	100.0	100.0
Excellent	50.1	57.2	58.8	47.0	24.3
Good	36.9	35.4	34.7	38.1	37.8
Fair	9.3	6.0	5.2	10.8	22.8
Poor	3.6	1.4	1.3	4.2	15.0

Table 19

Number and percent distribution of persons by perceived health status, according to level of prescribed medicine use: United States, 1980

Perceived health status	All persons	Level of prescribed medicine use			
		0	Low	Intermediate	High
		Number in thousands			
Excellent	111,641	49,779	17,335	43,823	703
Good	82,293	29,116	11,094	39,832	2,250
Fair	20,834	3,911	1,701	12,560	2,662
Poor	8,057	1,008	362	4,129	2,558
		Percent distribution			
Total	100.0	100.0	100.0	100.0	100.0
Excellent	50.1	59.4	56.9	43.7	8.6
Good	36.9	34.7	36.4	39.7	27.5
Fair	9.3	4.7	5.6	12.5	32.6
Poor	3.6	1.2	1.2	4.1	31.3

Table 20

Number and percent distribution of persons 17 years of age and over and living at the end of the survey period by functional limitation score, according to level of hospital use: United States, 1980

Functional limitation score	All persons	Level of hospital use: 0	Low	Intermediate	High
		Number in thousands			
Total	160,437	140,051	3,524	13,615	3,247
Level 1 – no limitation	122,298	111,182	2,626	7,842	648
Level 2 – minimal limitation	9,945	8,392	252	1,094	207
Level 3	6,900	5,550	143	871	337
Level 4	7,100	5,255	279	1,142	425
Level 5	6,604	4,995	134	1,069	407
Level 6	4,178	2,722	57	864	535
Level 7	1,721	1,039	13	360	309
Level 8 – most severe limitation	1,691	917	20	373	381
		Percent distribution			
Total	100.0	100.0	100.0	100.0	100.0
Level 1 – no limitation	76.2	79.4	74.5	57.6	20.0
Level 2 – minimal limitation	6.2	6.0	7.1	8.0	6.4
Level 3	4.3	4.0	4.1	6.4	10.4
Level 4	4.4	3.8	7.9	8.4	13.1
Level 5	4.1	3.6	3.8	7.9	12.5
Level 6	2.6	1.9	1.6	6.3	16.5
Level 7	1.1	0.7	0.4	2.6	9.5
Level 8 – most severe limitation	1.1	0.7	0.6	2.7	11.7

Table 21

Number and percent distribution of persons 17 years of age and over and living at the end of the survey period by functional limitation score, according to level of ambulatory care use: United States, 1980

Functional limitation score	All persons	Level of ambulatory care use			
		0	Low	Intermediate	High
	Number in thousands				
Total	160,437	33,889	25,388	92,734	8,426
Level 1 – no limitation	122,298	29,992	22,358	66,531	3,417
Level 2 – minimal limitation	9,945	1,065	1,022	7,032	826
Level 3	6,900	924	460	4,467	1,049
Level 4	7,100	720	661	4,939	780
Level 5	6,604	572	482	4,704	846
Level 6	4,178	313	170	2,928	767
Level 7	1,721	184	97	1,095	345
Level 8 – most severe limitation	1,691	118	137	1,038	398
	Percent distribution				
Total	100.0	100.0	100.0	100.0	100.0
Level 1 – no limitation	76.2	88.5	88.1	71.7	40.6
Level 2 – minimal limitation	6.2	3.1	4.0	7.6	9.8
Level 3	4.3	2.7	1.8	4.8	12.5
Level 4	4.4	2.1	2.6	5.3	9.3
Level 5	4.1	1.7	1.9	5.1	10.0
Level 6	2.6	0.9	0.7	3.2	9.1
Level 7	1.1	0.5	0.4	1.2	4.1
Level 8 – most severe limitation	1.1	0.3	0.5	1.1	4.7

Table 22

Number and percent distribution of persons 17 years of age and over and living at the end of the survey period by functional limitation score, according to level of prescribed medicine use: United States, 1980

Functional limitation score	All persons	Level of prescribed medicine use			
		0	Low	Intermediate	High
	Number in thousands				
Total	160,437	56,465	19,801	76,389	7,782
Level 1 - no limitation	122,298	50,301	17,295	53,389	1,313
Level 2 - minimal limitation	9,945	2,244	885	6,201	615
Level 3	6,900	1,274	468	4,302	856
Level 4	7,100	1,102	511	4,372	1,115
Level 5	6,604	796	349	3,969	1,490
Level 6	4,178	371	127	2,372	1,308
Level 7	1,721	165	89	894	574
Level 8 - most severe limitation	1,691	214	76	890	511
	Percent distribution				
Total	100.0	100.0	100.0	100.0	100.0
Level 1 - no limitation	76.2	89.1	87.3	69.9	16.9
Level 2 - minimal limitation	6.2	4.0	4.5	8.1	7.9
Level 3	4.3	2.3	2.4	5.6	11.0
Level 4	4.4	2.0	2.6	5.7	14.3
Level 5	4.1	1.4	1.8	5.2	19.1
Level 6	2.6	0.7	0.6	3.1	16.8
Level 7	1.1	0.3	0.4	1.2	7.4
Level 8 - most severe limitation	1.1	0.4	0.4	1.2	6.6

Table 23

Number and percent distribution of persons by number of unique conditions occurring during the year, according to level of hospital use: United States, 1980

Number of unique conditions	All persons	Level of hospital use: 0	Low	Intermediate	High
	Number in thousands				
0	32,385	31,834	216	335	–
1	42,428	40,267	580	1,444	137
2	40,341	37,338	773	1,907	324
3	33,010	29,150	914	2,610	336
4	23,485	20,022	750	2,289	424
5	17,409	14,398	533	2,074	405
6	10,952	8,381	550	1,541	480
7	8,003	6,042	248	1,331	382
8	5,281	3,885	259	868	269
9	3,127	2,140	203	582	202
10	2,250	1,452	77	472	249
11–24	4,153	2,373	140	1,010	629
	Percent distribution				
Total	100.0	100.0	100.0	100.0	100.0
0	14.5	16.1	4.1	2.0	–
1	19.0	20.4	11.1	8.8	3.6
2	18.1	18.9	14.7	11.6	8.4
3	14.8	14.8	17.4	15.9	8.8
4	10.5	10.1	14.3	13.9	11.0
5	7.8	7.3	10.2	12.6	10.6
6	4.9	4.2	10.5	9.4	12.5
7	3.6	3.1	4.7	8.1	10.0
8	2.4	2.0	4.9	5.3	7.0
9	1.4	1.1	3.9	3.5	5.3
10	1.0	0.7	1.5	2.9	6.5
11–24	1.9	1.2	2.7	6.1	16.4

Table 24

Number and percent distribution of persons by number of unique conditions occurring during the year, according to level of ambulatory care use: United States, 1980

Number of unique conditions	All persons	Level of ambulatory care use: 0	Low	Intermediate	High
		Number in thousands			
0	32,385	23,711	6,094	2,532	48
1	42,428	12,682	14,042	15,393	311
2	40,341	5,907	9,258	24,387	789
3	33,010	2,607	4,792	24,698	913
4	23,485	998	1,981	19,307	1,199
5	17,409	367	837	14,746	1,459
6	10,952	224	619	9,261	848
7	8,003	129	135	6,849	891
8	5,281	28	90	4,336	827
9	3,127	35	73	2,380	638
10	2,250	12	-	1,709	530
11-24	4,153	17	-	2,564	1,572
		Percent distribution			
Total	100.0	100.0	100.0	100.0	100.0
0	14.5	50.8	16.1	2.0	0.5
1	19.0	27.1	37.0	12.0	3.1
2	18.1	12.6	24.4	19.0	7.9
3	14.8	5.6	12.6	19.3	9.1
4	10.5	2.1	5.2	15.1	12.0
5	7.8	0.8	2.2	11.5	14.6
6	4.9	0.5	1.6	7.2	8.5
7	3.6	0.3	0.4	5.3	8.9
8	2.4	0.1	0.2	3.4	8.2
9	1.4	0.1	0.2	1.9	6.4
10	1.0	0.0	-	1.3	5.3
11-24	1.9	0.0	-	2.0	15.6

Table 25

Number and percent distribution of persons by number of unique conditions occurring during the year, according to level of prescribed medicine use: United States, 1980

Number of unique conditions	All persons	Level of prescribed medicine use: 0	Low	Intermediate	High
	Number in thousands				
0	32,385	30,760	873	751	–
1	42,428	24,445	8,204	9,737	41
2	40,341	14,255	9,199	16,458	429
3	33,010	7.729	5,841	18,856	584
4	23,485	3,552	3,042	16,214	678
5	17,409	1,624	1,640	13,231	915
6	10,952	855	806	8,470	821
7	8,003	289	439	6,268	1,007
8	5,281	95	218	4,034	934
9	3,127	95	92	2,179	761
10	2,250	53	61	1,575	561
11–24	4,153	64	76	2,572	1,442
	Percent distribution				
Total	100.0	100.0	100.0	100.0	100.0
0	14.5	36.7	2.9	0.7	–
1	19.0	29.2	26.9	9.7	0.5
2	18.1	17.0	30.2	16.4	5.3
3	14.8	9.2	19.2	18.8	7.1
4	10.5	4.2	10.0	16.2	8.3
5	7.8	1.9	5.4	13.2	11.2
6	4.9	1.0	2.6	8.4	10.0
7	3.6	0.3	1.4	6.2	12.3
8	2.4	0.1	0.7	4.0	11.4
9	1.4	0.1	0.3	2.2	9.3
10	1.0	0.1	0.2	1.6	6.9
11–24	1.9	0.1	0.2	2.6	17.6

Table 26

Number and percent distribution of persons with specified diagnoses associated with hospital admissions by diagnostic category, according to level of hospital use: United States, 1980

Diagnostic category	Number of hospitalized persons in thousands[1]	Level of hospital use: All users	Low users	Intermediate users	High users
		Percent distribution			
Total	...	100.0	100.0	100.0	100.0
Infectious and parasitic diseases	731	2.9	2.3	3.2	2.3
Neoplasms	1,638	6.4	3.3	5.0	16.7
Endocrine, nutritional and metabolic diseases, and immunity disorders	1,026	4.0	1.0	4.5	6.1
Diseases of the blood and blood-forming organs	306	1.2	0.5	1.0	3.0
Mental disorders	695	2.7	1.8	2.0	7.1
Diseases of the nervous system and sense organs	1,558	6.7	6.9	5.2	8.7
Diseases of the circulatory system	3,506	13.7	5.6	11.9	32.8
Diseases of the respiratory system	3,113	12.2	15.0	10.4	16.1
Diseases of the digestive system	3,366	13.2	8.3	14.8	12.8
Diseases of the genitourinary system	2,977	11.7	15.0	11.3	8.5
Complications of pregnancy, childbirth, and the puerperium	1,901	7.4	6.9	8.7	2.6
Diseases of the skin and subcutaneous tissue	406	1.6	1.0	1.5	2.7
Diseases of the musculoskeletal system and connective tissue	1,952	7.6	3.5	7.2	15.1
Congenital anomalies	263	1.0	2.3	0.6	0.9
Certain conditions originating in the perinatal period	87	0.3	–	0.4	0.7
Symptoms, signs and ill-defined conditions	1,834	7.2	4.8	6.7	12.6
Injury and poisoning	3,123	12.2	13.5	9.6	21.8
No condition or condition unknown	3,699	14.5	14.7	17.1	2.8

[1] Unduplicated count of persons within diagnostic categories; a person who had multiple hospital admission for the same condition is counted only once. However, since each hospital admission may have been associated with up to 3 separate diagnostic categories, the same person may be counted in more than 1 diagnostic category. Diagnoses were examined only for persons having at least 1 overnight hospital stay.

Table 27

Number and percent distribution of persons by whether surgery was performed during the stay, according to level of hospital use: United States, 1980

Presence of surgery	All persons	Level of hospital use: Low	Intermediate	High
		Number in thousands		
Surgery	11,904	2,566	7,389	1,949
No surgery	13,637	2,676	9,073	1,888
		Percent distribution		
Total	100.0	100.0	100.0	100.0
Surgery	46.6	49.0	44.9	50.8
No surgery	53.4	51.0	55.1	49.2

Table 28

Number and percent distribution of persons by level of ambulatory care use, according to level of hospital use: United States, 1980

Level of ambulatory care use	All persons	Level of hospital use: 0	Low	Intermediate	High
		Number in thousands			
Total	222,824	197,283	5,242	16,462	3,837
0	46,716	46,204	70	324	119
Low	37,921	36.824	293	668	136
Intermediate	128,163	107,790	4,541	13,488	2,344
High	10,024	6,465	339	1,982	1,238
		Percent distribution			
Total	100.0	100.0	100.0	100.0	100.0
0	21.0	23.4	1.3	2.0	3.1
Low	17.0	18.7	5.6	4.1	3.6
Intermediate	57.5	54.6	86.6	81.9	61.1
High	4.5	3.3	6.5	12.0	32.3

Number and percent distribution of persons with hospital use by number of hospital admissions, according to level of hospital use: United States, 1980

Number of hospital admissions	All persons	Level of hospital use: Low	Intermediate	High
	Number in thousands			
1	19,629	5,182	13,228	1,219
2	4,274	60	2,800	1,414
3	986	–	327	659
4	351	–	93	258
5 or more	303	–	15	288
	Percent distribution			
Total	100.0	100.0	100.0	100.0
1	76.8	98.9	80.4	31.8
2	16.7	1.1	17.0	36.8
3	3.9	–	2.0	17.2
4	1.4	–	0.6	6.7
5 or more	1.2	–	0.1	7.4

Table 30

Number and percent distributions of persons by survival status, according to level of hospital use: United states, 1980

Survival status	All persons	Level of hospital use: 0	Low	Intermediate	High
	Number in thousands				
Total	222,824	197,283	5,242	16,462	3,837
Survivors	221,969	197,079	5,204	16,312	3,373
Deaths	856	204	38	150	464
	Percent distributions				
Total	100.0	88.5	2.4	7.4	1.7
Survivors	100.0	88.8	2.3	7.3	1.5
Deaths	100.0	23.8	4.4	17.5	54.2
Total	100.0	100.0	100.0	100.0	100.0
Survivors	99.6	99.9	99.3	99.1	87.9
Deaths	0.4	0.1	0.7	0.9	12.1

Table 31

Number and percent distributions of persons by survival status, according to level of ambulatory care use: United States, 1980

Survival status	All persons	Level of ambulatory care use: 0	Low	Intermediate	High
		Number in thousands			
Total	222,824	46,716	37,921	128,163	10,024
Survivors	221,969	46,601	37,892	127,638	9,838
Deaths	856	115	29	525	187
		Percent distributions			
Total	100.0	21.0	17.0	57.5	4.5
Survivors	100.0	21.0	17.1	57.5	4.4
Deaths	100.0	13.4	3.4	61.4	21.8
Total	100.0	100.0	100.0	100.0	100.0
Survivors	99.6	99.8	99.9	99.6	98.1
Deaths	0.4	0.2	0.1	0.4	1.9

Table 32

Number and percent distributions of persons by survival status, according to level of prescribed medicine use: United States, 1980

Survival status	All persons	Level of prescribed medicine use: 0	Low	Intermediate	High
		Number in thousands			
Total	222,824	83,414	30,492	100,345	8,173
Survivors	221,969	83,670	30,436	99,900	7,964
Deaths	856	144	56	446	209
		Percent distributions			
Total	100.0	37.6	13.7	45.0	3.7
Survivors	100.0	37.7	13.7	45.0	3.6
Deaths	100.0	16.9	6.6	52.1	24.5
Total	100.0	100.0	100.0	100.0	100.0
Survivors	99.6	99.8	99.8	99.6	97.4
Deaths	0.4	0.2	0.2	0.4	2.6

Appendixes

List of Appendix Figures

List of Appendix Tables

Appendix I Sample Design, Data Collection, and Processing

Introduction

The National Medical Care Utilization and Expenditure Survey (NMCUES) was designed to collect data about the U.S. civilian noninstitutionalized population during 1980. The complexity of the survey requires the analyst to be familiar with a range of design features during the analysis, both to determine appropriate analytic methods and to investigate the impact that the design may have on a particular analysis. Several topics are addressed in this appendix: The overall design of NMCUES, the survey background, sampling methods, data collection methods, weighting, and compensation procedures for missing data. These descriptions essentially present NMCUES data as they are available to the user of the public use data tape. This appendix draws heavily from a paper in the Proceedings of the 19th National Meeting of the Public Health Conference on Records and Statistics (Casady, 1983).

Survey Background

During the course of NMCUES, information was obtained on health, access to and use of medical services, associated charges and sources of payment, and health care coverage. The survey was cosponsored by the National Center for Health Statistics (NCHS) and the Health Care Financing Administration (HCFA). Data collection was provided under contract by the Research Triangle Institute (RTI) and its subcontractors, National Opinion Research Center (NORC) and SysteMetrics, Inc.

The basic survey plan for NMCUES drew heavily on two previous national surveys: The National Health Interview Survey (NHIS), which is conducted by NCHS, and the National Medical Care Expenditure Survey (NMCES), which was cosponsored by the National Center for Health Services Research (NCHSR) and NCHS.

NHIS is a continuing, multipurpose health survey first conducted in 1957. The primary purpose of NHIS is to collect information on illness, disability, and the use of medical care. Although some information on medical expenditures and insurance payments has been collected in NHIS, the cross-sectional nature of the NHIS survey design is not well suited for providing annual data on expenditures and payments.

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.

NMCUES is similar to NMCES in survey design and question wording, so that analysis of change during the 3 years between 1977 and 1980 is possible. Both NMCUES and NMCES are similar to NHIS in terms of question wording in areas common to the three surveys. Together they provide extensive information on illness, disability, use of medical care, costs of medical care, sources of payment for medical care, and health care coverage at two points in time.

Sample Design of NMCUES

General plan—The NMCUES sample of housing units and group quarters, hereafter jointly referred to as dwelling units, is a concatenation of two independently selected national samples, one provided by RTI and the other by NORC. The sample designs used by RTI and NORC are quite similar with respect to principal design features; both can be characterized as stratified, multistage area probability designs. The principal differences between the two designs are the type of stratification variables and the specific definitions of sampling units at each stage.

Target population—All persons living in a sample dwelling unit at the time of the first interview became part of the national sample. Unmarried students 17–22 years of age who lived away from home were included in the sample when their parent or guardian was included in the sample. In addition, 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 sample dwelling units and were living in the sample dwelling before their death. All of these persons were considered "key" persons, and data were collected for them for the full 12 months of 1980 or for the portion of time they were part of the U.S. civilian noninstitutionalized population. 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, and data were collected for them 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; but 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 are used in an analysis of families because they contribute to the family's utilization of and expenditures for health care during the time they are part of the family. Because family analysis is not part of this investigation, it 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 who lived in the same dwelling unit. The combined NMCUES sample consisted of approximately 7,200 reporting units, of which 6,600 agreed to participate in the survey. In total, complete data were obtained on 17,123 key persons. The RTI sample yielded approximately 8,300 respondents and the NORC sample 8,800.

Research Triangle Institute Sample Design

Primary sampling units (PSU's)—A PSU was defined as a county, a group of contiguous counties, or parts of counties with a combined minimum 1970 population size 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. The PSU's were classified as one of two types. The 16 largest Standard Metropolitan Statistical Areas (SMSA's) were designated as self-representing PSU's, and the remaining 1,670 PSU's in the primary sampling frame were designated as nonself-representing PSU's.

Stratification of PSU's—PSU's were grouped into strata whose members tend to be relatively alike within strata and relatively unlike between strata. PSU's derived from the 16 largest SMSA's were of sufficient 1970 population size to be treated as primary strata. The 1,659 nonself-representing PSU's from the continental United States were stratified into 42 approximately equal-sized, primary strata. Each of these primary strata had a 1970 population size of about 3.3 million. One supplementary primary stratum of 11 PSU's, with a 1970 population size of about 1 million, was added to the RTI primary frame to include Alaska and Hawaii.

First-stage selection of PSU's—The total RTI primary sample consisted of 59 PSU's of which 16 were self-representing. The nonself-representing PSU's were obtained by selecting 1 PSU from each of the 43 nonself-representing primary strata. These PSU's were selected with probability proportional to 1970 population size.

Secondary stratification—In each of the 59 sample PSU's, the entire PSU was divided into nonoverlapping smaller area units called secondary sampling units (SSU's). Each SSU consisted of one or more 1970 census-defined enumeration districts (ED's) or block groups (BG's). Within each PSU the SSU's were ordered and then partitioned to form approximately equal-sized secondary strata. Two secondary strata were formed in the nonself-representing PSU drawn from Alaska and Hawaii, and four secondary strata were formed in each of the remaining 42 nonself-representing PSU's. Thus, the nonself-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.

Second-stage selection of SSU's—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 the SSUs' total noninstitutionalized population in 1970. The total number of sample SSU's was $2 \times 170 + 144 = 484$.

Third-stage selection of areas and segments—Each SSU was divided into smaller nonoverlapping geographic areas, and one area within the SSU was selected with probability proportional to the 1970 total number of housing units. Next, one or more nonoverlapping segments of at least 60 housing units (HU's) were formed in the selected area. 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 household sample. Thus, the total third-stage sample was reduced to 404 segments.

Fourth-stage selection of housing units—All of the dwelling units within the segment were listed, and a systematic sample of dwelling units was selected. The procedures used to determine the sampling rate for segments guaranteed that all dwelling units had an approximately equal probability of selection. All of the reporting units within the selected dwelling units were included in the sample.

National Opinion Research Center Sample Design

Primary sampling units (PSU's)—The land area of the 50 States and Washington, D.C. was divided into nonoverlapping PSU's. A PSU consisted of SMSA's, parts of SMSA's, counties, parts of counties, or independent cities. Grouping of counties into a single PSU occurred when individual counties had a 1970 population of less than 10,000.

Zoning of PSU's—The PSU's were classified into two groups according to metropolitan status (SMSA or not SMSA). These two groups were individually ordered and then partitioned into zones with a 1970 census population size of 1 million persons.

First-stage zone selection of PSU's—A single PSU was selected within each zone with a probability proportional to its 1970 population. It should be noted that this procedure allows a PSU to be selected more than one time. For instance, an SMSA PSU with a population of 3 million may be selected at least twice and possibly as many as four times. The full general-purpose sample contained 204 PSU's. These 204 PSU's were systematically allocated to 4 subsamples of 51 PSU's. The final set of 76 sample PSU's was chosen by randomly selecting 2 complete subsamples of 51 PSU's; 1 subsample was included in its entirety, and 25 of the PSU's in the other subsample were selected systematically for inclusion in NMCUES.

Second-stage zone selection of SSU's—Each of the PSU's selected in the first stage was partitioned into a nonoverlapping set of SSU's defined by BG's, ED's, or a combination of the two types of census units. SSU's were selected from the ordered list of these SSU's. The cumulative number of households in the second-stage frame for each PSU was divided into 18 zones of equal width. One SSU had the opportunity to be selected more than once, as was the case in the PSU selection. If a PSU had been hit more than once in the first stage, then the second-stage selection process was repeated as many times as there were the first-stage hits. Some 405 SSU's were identified by selecting 5 SSU's from each of the 51 PSU's in the subsample that were included in their entirety, and 6 SSU's from each of the 25 PSU's in the group for which one-half of the PSU's were included.

Third-stage selection of segments—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.

Fourth-stage selection of housing units—Sample selection at this level was essentially the same as for the RTI design.

Collection of Data

Field operations for NMCUES were performed by RTI and NORC under specifications established by the cosponsoring agencies. Persons in the sample dwelling units were interviewed at approximately 3-month intervals beginning in February 1980 and ending in March 1981. The Core Questionnaire was administered during each of the five interview rounds to collect data on health, health care, health care charges, sources of payment, and health care coverage. A summary of responses was used to update information reported in previous rounds. Supplements to the Core Questionnaire were used during the first, third, and fifth interview rounds to collect data that did not change during the year or that were needed only once. Approximately 80 percent of the third- and fourth-round interviews were conducted by telephone; all remaining interviews were conducted in person. The respondent for the interview was required to be a household member 17 years of age or over. A nonhousehold proxy respondent was permitted only if all eligible household members were unable to respond because of health, language, or mental condition.

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 RU's (i.e., nonresponse), and failure to cover some portions of the population because the sampling frame did not include them (i.e., undercoverage).

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 weight 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 when multiple selection of the sample unit was possible.

Two-sample adjustment factor—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 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 RU level to compensate for RU-level nonresponse and undercoverage. Because every RU within a dwelling unit is included in the sample, the adjusted basic sample design weight assigned to an RU is simply the adjusted basic sample design weight for the dwelling unit in which the RU is located. An RU was classified as responding if members of the RU initially agreed to participate in NMCUES and as nonresponding otherwise.

Initially, 96 RU weight adjustment cells were formed by cross-classifying the following variables: Race of RU head (2 levels), type of RU head (3 levels), age of RU head (4 levels), and size of RU (4 levels). These cells were then collapsed to 63 cells so that each cell contained at least 20 responding RU's. Within each cell an adjustment factor was computed so that a sum of adjusted basic sample design weights would equal the March 1980 *Current Population Survey* estimate for the same population. The nonresponse and undercoverage adjusted weight was computed for each RU as the product of the adjusted basic sample design weight and the nonresponse and undercoverage adjustment factor for the cell containing the RU.

Poststratification adjustment—Once the nonresponse and undercoverage adjusted RU weights were computed, a poststratification adjusted weight was computed at the person level. Because each person within an RU is included in the sample, the nonresponse and undercoverage adjusted weight for a sample person is the nonresponse and undercoverage adjusted weight for the RU 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 (2 levels) and sex (2 levels). One poststratum (black males 75 years of age and over) had less than 20 respondents, so it was combined with an adjacent poststratum (black males 65–74 years of age), resulting in 59 poststrata.

Estimates based on population projections from the 1980 census were obtained from the Bureau of the Census for the U.S. civilian noninstitutionalized population by age, race, and sex poststrata for February 1, May 1, August 1, and November 1, 1980. The mean of these mid-quarter population estimates for each of the poststrata was computed and used as the 1980 average target population for calculating the poststrata adjustment factors.

Survey-based estimates of the average poststrata population were developed using the nonresponse and undercoverage adjusted weights. First, a survey-based estimate of the target population of each poststratum for each quarter was computed by summing the nonresponse and undercoverage adjusted weights for respondents eligible for the survey on the mid-quarter date. Then the survey-based estimate of the 1980 average population was computed as the mean of the 4 mid-quarter estimates. Finally, the poststratification adjustment factor was computed in each poststratum as the ratio of the 1980 average target population (obtained from 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 and undercoverage adjusted weight and the poststratification adjustment factor for the poststratum containing the respondent.

Thus, the weighting procedure is composed of three steps: Development of base sample design weights for each RU, adjustment for RU-level nonresponse and undercoverage, and adjustment for person-level nonresponse and undercoverage. A further adjustment for the number of days a person was an eligible member of the U.S. civilian noninstitutionalized population was made, but this adjustment only affects certain types of estimates from NMCUES and is discussed in the appendix on estimation (Appendix III).

Survey Nonresponse

Nonresponse in panel surveys such as NMCUES occurs when sample individuals refuse to participate in the survey (total nonresponse), when initially participating individuals drop out of the survey (attrition nonresponse), or when data for specific items on the questionnaire are not collected (item nonresponse). Response rates for RU's and person in NMCUES were high, with approximately 90 percent of the sample RU's agreeing to participate in the survey and approximately 94 percent of the individuals in the participating RU's supplying complete information. Even though the overall response rates are high, survey-based estimates of means and proportions may be biased if nonrespondents tend to have different health care experiences than respondents or if there is a substantial response rate differential across subgroups of the target population. Furthermore, annual totals tend to be underestimated unless allowance is made for the loss of data attributable to nonresponse.

Two methods commonly used to compensate for survey nonresponse are data imputation and adjustment of sampling weights. For NMCUES, data imputation was used to compensate for attrition and item nonresponse, and weight adjustment was used to compensate for total nonresponse. The calculation of the weight adjustment factors was discussed previously in the section on sampling weights.

Attrition Imputation

A special form of the 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 (i.e., in scope), and the imputed ineligible days were those days for which the donor was ineligible (i.e., out of scope). The donor's medical care experiences, such as medical provider visits, dental visits, and hospital stays, during the imputed eligible days 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 a total nonrespondent, and the data for the person were removed from the data file.

Item Nonresponse and Imputation

Among persons who are classified as respondents, there is still the possibility that they may fail to provide information for some or many items in the questionnaire. In NMCUES, item nonresponse was particularly a problem for health care charges, income, and other sensitive topics. The extent of missing data varied by question, and imputation for all items in the data file would have been expensive. Imputations were made for missing data on key demographic, economic, and charge items across five of the six data files in the public use data tape. Table I illustrates the extent of the item nonresponse problem for selected survey measures that received imputation in four data files used in this report.

Demographic items tend to require the least amount of imputation, some at insignificant levels such as for age, sex, and education. Income items had higher levels of nonresponse. For total personal income, which is a cumulation of the earned income and 11 sources of unearned income, imputation was required for at least one component for nearly one-third of the persons. The bed-disability days, work-loss days, and cut-down days have levels of imputation that are intermediate to the demographic and income items.

The highest levels of imputation occurred for the important charge items on the various visit, hospital stay, and medical expenses files. Total charges for medical visits, hospital stays, and prescribed medicines and other medical expenses were imputed for 25.9 percent, 36.3 percent, and 19.4 percent of the events, respectively. Among the source of payment

Table I

Percent of data imputed for selected survey items in 4 of the NMCUES public use data files: United States, 1980

Tape location	Description	Percent imputed
	Person file (n = 17,123)	
P54	Age	0.1
P57	Race	[1]20.0
P59	Sex	0.1
P62	Highest grade attended	0.1
P67	Perceived health status	0.8
P592	Functional limitation score	3.2
P125	Number of bed-disability days	7.9
P128	Number of work-loss days	8.9
P135	Number of cut-down days	8.2
P399	Wages, salary, business income	9.7
P434	Pension income	3.5
P445	Interest income	21.6
P462	Total personal income	[2]30.4
	Medical visit file (n = 86,594)	
M117	Total charge	25.9
M123	First source of payment	1.8
M125	First source of payment amount	11.6
	Hospital stay file (n = 2,946)	
H252	Nights hospitalized	3.1
H124	Total charge	36.3
H130	First source of payment	2.2
H132	First source of payment amount	17.6
	Medical expenses file (n = 58,544)	
E117	Total charge	19.4
E123	First source of payment	2.8
E125	First source of payment amount	10.0

[1]Race for children under 14 years of age imputed from race of head of reporting unit.
[2]Cumulative across 12 types of income.

data, the imputation rates for the source of payment were small, but the rates for the amount paid by the first source of payment were generally subject to high rates of imputation. Nights hospitalized on the hospital stay file were imputed at a rate comparable to the first source of payment.

The methods used to impute for missing items were diverse and tailored to the measure requiring imputation. Three types of imputation predominate: Editing or logical imputations, a sequential hot deck, and a weighted sequential hot deck. The edit or logical imputations were used to eliminate missing data that could reasonably be determined from other data items that provided overlapping information for the given item. The sequential hot deck was used primarily for small numbers of imputations for the demographic items; the weighted sequential hot deck was used more extensively and for virtually all other items for which imputations were made.

The edit or logical imputation is a process in which the value of a missing item is deduced from other available information in the data file. For example, race was not recorded for children under 14 years of age during the survey. Instead, a logical imputation was made during data processing that assigned the race of the head of the reporting unit to the child. Similarly, extensive editing was performed for the charge data before any imputations were made. If first source of payment was available, only one source of payment was given; and if total charge was missing, the value of the first source of payment amount was assigned to the total charge item.

In the sequential hot deck procedure, the data are grouped within imputation classes formed by variables thought to be correlated with the item to be imputed. An additional sorting within imputation classes by other variables also thought to be correlated with the imputed item is also typically used. An initial value, such as the mean of the nonmissing cases for the item, is assigned as a "cold deck" value. The first record in the file is then examined. If it is missing, the "cold deck" value replaces the missing data code; if real, the real value replaces the "cold deck" value and becomes a "hot deck" value. Then the next record is examined. Again, if missing, the "hot deck" value is used to replace missing data, and, if real, the "hot deck" value is replaced by that real value. The process continues sequentially through the sorted file. The weighted hot deck, a modification of the sequential hot deck, uses the weights to determine which real values are used to impute for a particular record needing an imputation.

The imputation process will be described for two items to illustrate the nature of imputation for NMCUES. For Hispanic origin, two different imputation procedures were used: logical and sequential hot deck. Because Hispanic origin was not recorded during the interview for children under 17 years of age, a logical imputation was made by assigning to the child the Hispanic origin of the wife of the head of the reporting unit, if present, and the origin of the head of the reporting unit otherwise. For the remaining cases that were not assigned a value by this procedure, the data were grouped into classes by observed race of the head of the reporting unit; within classes, the data were sorted by reporting unit identification number, primary sampling unit, and segment. An unweighted sequential hot deck was used to impute values of Hispanic origin for the remaining cases with missing values.

The imputations for medical visit total charge were made after extensive editing had been performed to eliminate as many inconsistencies as possible between sources of payment data and total charge. The medical visit records were then separated into three types: emergency room, hospital outpatient department, and doctor visits. Within each type, the records were classed and sorted by several measures, which differed across visit types, prior to a weighted hot deck imputation. For example, the records for doctor visits were classified by reason for visit, type of doctor seen, whether work was done by a physician, and age of the individual. Within the groups formed by these classing variables, the records were then sorted by type of health care coverage and month of visit. The weighted hot deck procedure was then used to impute for missing total charge, sources of payment, and sources of payment amounts for the classified and sorted data file.

Because imputations were made for missing items 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 would 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 in Lepkowski, Stehouwer, and Landis (1984).

Appendix II
Data Modifications to Public Use Files

Overview of Data Changes

During the preparation of this report, a number of problems were discovered in the NMCUES public use files that required modification of the data. Eight sets of problems were identified:

(1) Sampling weights for 68 newborns (i.e., persons born in 1980) were in error.
(2) Six respondents had extremely high hospital stay charges.
(3) Forty-seven respondents had health care coverage categories inconsistent with source of payment for some medical events.
(4) For 173 respondents, fewer bed-disability days were reported than hospital nights. (Length-of-stay data were recorded in terms of the number of nights—as opposed to days—spent in the hospital.)
(5) Four respondents had extremely long lengths of stay in the hospital as a result of incorrect hospital admission dates.
(6) Four respondents had poverty status categories that were inconsistent with their poverty status level.
(7) Nine respondents were coded as deliveries in the hospital file but had inconsistent values for other hospital stay data.
(8) One respondent had duplicate hospital stay records.

Details of the changes made to correct these problems may be obtained from NCHS. General information on the problems and changes is outlined below. Detailed descriptions of the specific changes are provided in the NMCUES report series by Lepkowski, et al. (to be published).

1) Records for 68 newborns were incorrectly coded as eligible for the entire survey period (all 366 days) despite a birth after January 1, 1980. These errors were corrected by changing the eligible time-adjustment factor and the person time-adjusted weight for each of the 68 records.

2) After careful examination, the University of Michigan and NCHS determined that six hospital stay records with charges of at least $90,000 were incorrect and should be changed. These six records and related information in the person file (e.g., hospital stay charges, total charges) were changed to conform with records in the Medicare best estimate file or with other information about each of the 6 respondents' hospitalizations contained in the hospital stay file.

3) Discrepancies between source of payment and health care coverage were noted in the course of analysis. 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 the respondents was reclassified strictly according to source of payment data. 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. Where reassignment based on imputed data for source of payment would conflict with real data for health care coverage, the real data were used in preference to the imputed data.

4) For 173 cases, the value for hospital nights was greater than the value for bed-disability days. 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 adjusted to equal hospital nights for these 173 cases, a procedure used in the Health Interview Survey processing. However, it does not fully compensate for the errors in recording or computing bed-disability days. It is likely that after the edit, bed-disability days are still underestimated for these 173 cases. The edit was performed without regard to the imputation status of either bed-disability days or hospital nights.

5) Four cases with discrepancies between bed-disability days and hospital days also had improperly coded hospital admission dates causing excessively long lengths of stay. In these cases, the admission dates and hospital nights were corrected and the bed-disability days edit was not necessary.

6) Comparison of the continuous and the categorical poverty status variables on the public use file identified four respondents whose categorical poverty status was inconsistent with their continuous poverty status value. The categorical variable was changed for these four respondents to correspond to their poverty status on the continuous variable.

7) A variety of problems were discovered on nine records coded as deliveries in the hospital stay file.

(a) Two deliveries were attributed to male respondents. Examination of the data files suggested that the sex variable was incorrectly coded in these two cases. The sex variable was, therefore, recoded to female. A third male delivery 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 respondents.

(b) Four hospitalizations for newborns were incorrectly coded as deliveries. These were recoded 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 respondents.
(c) One delivery was attributed to a 74-year-old woman. Following an NCHS recommendation, the response was recoded to reflect signs, symptoms, and ill-defined conditions as the admitting condition.

8) Two sets of duplicate records (four records in total) in the hospital stay 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 for reasons of attrition. No changes were made in the records for the respondent receiving the attrition-imputed records.

Appendix III Analytical Strategies

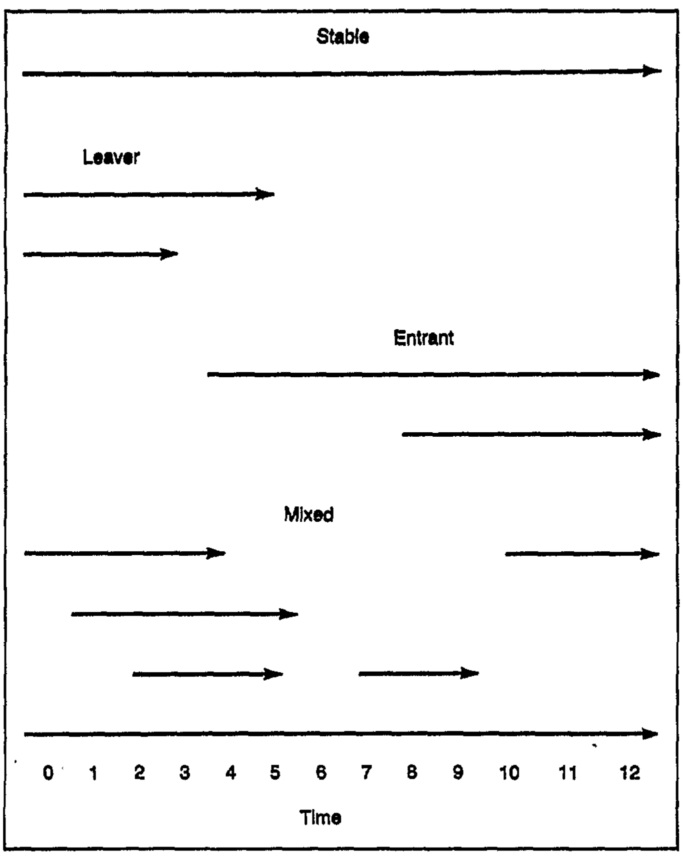
Notion of an Average Population

The NMCUES was a panel survey in which members of the population were followed during the panel period (i.e., 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 NMCUES data, it is useful to review the nature of a dynamic population over time.

Figure I 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 residences during the panel period and may be quite difficult to trace. Leavers are persons who are eligible at the beginning of a time period, but then become ineligible at some later time. Leaving may occur through events such as death, institutionalization, or moving outside the geographic boundary of the population. At the same time, new members may enter the population (i.e., entrants) through births or 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 of the elements in the dynamic population considered over the entire time period, NMCUES followup rules were carefully specified to include entrants, leavers, and mixed population elements properly.

Figure I

Dynamic population for 12 time period panel survey



As an illustration, consider a member of the Armed Forces 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 to provide information about the civilian population, medical care use and charges during the first 5 months of 1980 for this person 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 only "at risk" of incurring health care use or charges for 7 of the 12 months. This person thus contributes only 7⁄12 or 0.58 years of eligibility or "person year" to the study. This quantity is referred to as the "time-adjustment factor" in the documentation and throughout these appendixes.

For those 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 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. On the other hand, there were also several ways that persons who were eligible members of the population could become ineligible. Death obviously removes a person from further followup, as would institutionalization, joining the Armed Forces, or moving to residence outside the United States. For every person selected for NMCUES, information was collected to monitor the exact number of days they were 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, total charges for health care) do not need to account for time adjustments. Estimates for person characteristics (e.g., total population, proportion of the population in a given subgroup) must be based on person years to obtain estimates that correspond to those for health care estimates. 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 for health care during 1980 must use the total charges for health care as a numerator, without time adjustment, but 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 (see Table II). 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 (i.e., deaths, institutionalization, and so on). 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 20–29 years of age on January 1 are a cohort for which few additions are expected due to returns from institutions, the Armed Forces, or living abroad. And few removals are expected due to death, institutionalization, joining the Armed Forces, or moving abroad. On the other hand, males 80 years of age or 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 during the year.

Role of Weights and Imputation

Estimated means and sampling errors from NMCUES for bed-disability days, work-loss days, work-loss days in bed, cut-down days, and restricted-activity days are presented in Table III. For each survey measure, separate estimates were computed using all data (i.e., both real and imputed) and using only the real data. The unweighted and weighted mean, unweighted and weighted simple random sampling standard error of the mean, and the weighted complex standard error which accounts for the stratified, multistage nature of the design are presented.

For each measure, the weighted means computed using all the data and using only the real data are quite similar. This similarity is not unexpected given that the weighted hot deck imputation procedure is designed to preserve the weighted mean for overall sample estimates. The simple random sampling standard errors, however, are smaller when all data are used simply because the simple random sampling variance is inversely related to the sample size. For the complex standard error, three of the five measures have smaller standard errors when all data are used, and the other two measures show the opposite relationship. Weighting and imputation for the disability measures have little or no effect on estimated means or their standard errors for the total population because the amount of missing data for these measures is small (approximately 7 or 8 percent).

For other measures that have larger amounts of missing data, imputation has larger effects. Consider the means and standard errors for total charge for a hospital outpatient department visit shown in Table IV. There were 9,529 hospital outpatient department visits (real visit records plus those generated from the attrition imputation process), and 4,841 of these have a total charge that was imputed from one of the other hospital outpatient department visit records. Thus, more than one-half of the total charges were missing for this particular medical event. Despite the large amount of missing data, the weighted means using all the data and using only real values are quite similar; weighting does not

Table II

Effect of person-year adjustment on counts and sampling weights, by 4 population groups: United States, 1980

Population group	Sample size	Person years	Sum of sampling weights: Basic weight in thousands	Sum of sampling weights: Adjusted weight in thousands
Total population	17,123	16,862.84	226,368	222,824
Females, 25–29 years of age	702	699.39	9,529	9,494
Males, 80 years of age and over	113	104.05	1,384	1,274
All persons born during 1980	251	121.02	3,560	1,713

Table III

Sample size, means, and standard errors for 5 disability measures, by all and real data subgroups: United States, 1980

Disability measure and data type	Sample size	Unweighted estimates: Mean	Unweighted estimates: Simple random sampling standard error	Weighted estimates: Mean	Weighted estimates: Simple random sampling standard error	Weighted estimates: Complex standard error
Bed-disability days						
All data	17,123	5.303	0.1279	5.268	0.1269	0.1540
Real data	15,777	5.253	0.1326	5.228	0.1319	0.1599
Work-loss days						
All data	13,069	3.614	0.1221	3.696	0.1220	0.1629
Real data	11,537	3.510	0.1284	3.574	0.1277	0.1716
Work-loss days in bed						
All data	13,069	1.516	0.0508	1.568	0.0518	0.0592
Real data	10,970	1.530	0.0556	1.578	0.0568	0.0652
Cut-down days						
All data	17,123	6.831	0.1681	6.881	0.1697	0.3343
Real data	15,724	6.609	0.1721	6.639	0.1735	0.3322
Restricted-activity days						
All data	17,213	13.746	0.2559	13.805	0.2573	0.4716
Real data	14,049	13.036	0.2732	13.064	0.2742	0.4658

Table IV

Sample size, means, standard errors, and element variance for total charge for a hospital outpatient department visit, by data type: United States, 1980

Data type	Sample size	Unweighted estimates: Mean	Unweighted estimates: Simple random sampling standard error	Weighted estimates: Mean	Weighted estimates: Simple random sampling standard error	Complex standard error	Element variance ($\times 10^{-3}$)
All data	9,529	51.86	1.030	51.61	1.018	1.914	9.87
Real data only	4,688	52.28	1.436	52.27	1.430	2.936	9.59
Imputed data	4,841	51.45	1.476	50.98	1.447	1.60	10.14
Real data							
Not donor	929	47.83	2.108	48.53	2.117	3.935	4.17
Donor once	2,789	55.85	2.016	55.76	1.982	3.386	11.00
Donor twice	841	48.61	3.525	49.37	3.579	4.879	10.78
Donor 3–5 times	120	29.45	7.340	28.97	7.987	11.64	7.66

affect the estimated means. However, sampling errors are changed substantially when imputed values are added to real values to form an estimate. The weighted and unweighted simple random sampling standard errors are markedly smaller for all data than for the real data.

To investigate whether this decrease in sampling error is due to changes in sample size, changes in the element variance, or both, the element variances were computed by multiplying the weighted simple random sampling variances by the sample sizes. Inspection of Table IV suggests that the element variances are quite similar using all data and real data; the differences in standard error when all data and only real data are used can mostly be attributed to the loss in sample size when going from all data to real data.

Not all of the real data were used as donors for imputation, and some of the real values were used as donors several times. Table IV also suggests that those real values not used as donors have a lower mean total charge than those used as donors, but values used as donors more than twice tend to have even smaller mean total charges. The means for donors used once, twice, or more frequently are a function of the use of imputation classes within which the mean total charge and the amount of missing data varied.

The difference in complex standard errors between all data and the real data in Table IV illustrates the large effects of imputation. However, neither the complex standard error computed using all the data nor that computed using only the real data is the correct standard error of the weighted mean estimated 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 primary sampling units and strata used in the sample selection process and in the variance estimation procedure. The assumption that the observations were selected independently between primary sampling units and strata, which is needed to justify the variance estimation procedure, is incorrect. Hence, the complex standard error for all data shown in Table IV fails to account for two sources of variability present in estimates based on all data: 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 sampling error of the weighted mean for all the data is probably larger than that shown for the mean estimated using all the data; it may even be larger than the sampling error computed using only the real data.

As a final illustration of the effects that imputation can have on survey results, Figure II presents estimated mean charges per hospital outpatient department visit for four family income groups computed using all the data and using only the real data. For the real data, the mean charge per visit

Figure II

Estimated mean charges per hospital outpatient department visit, by 4 family income classes for all and real data: United States, 1980

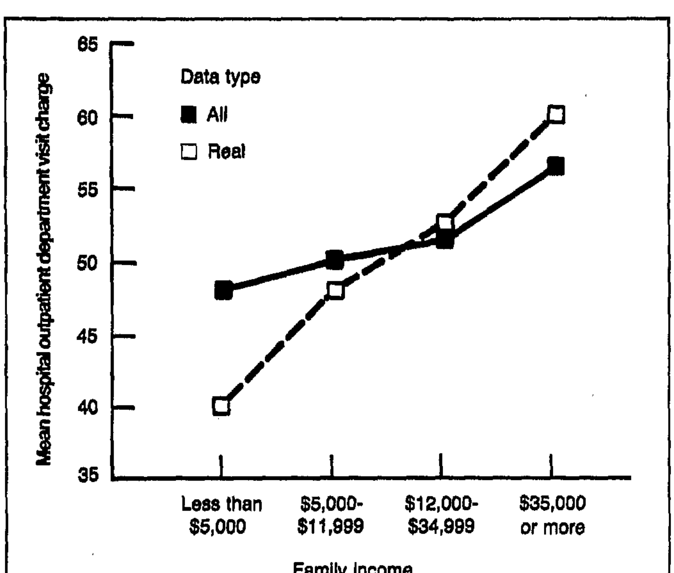
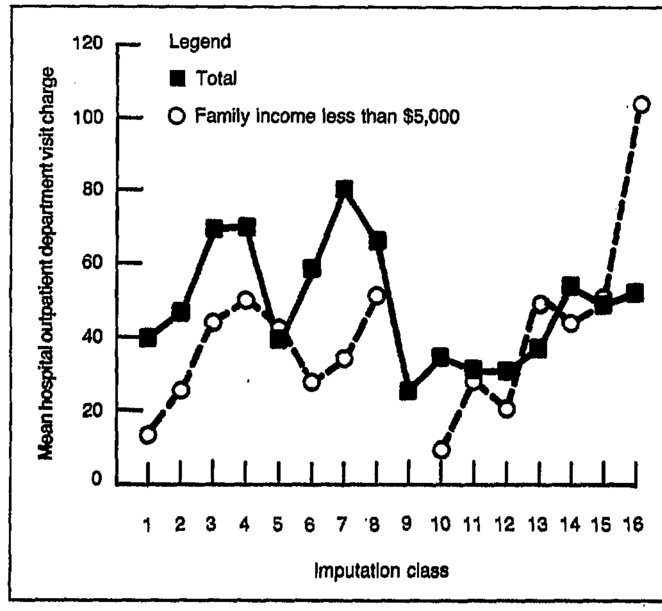


Figure III

Estimated mean charge per hospital outpatient department visit, by 16 imputation classes for all persons and for persons in families with income less than $5,000: United States, 1980



increases in a linear fashion as the family income increases. However, when all the data are used to estimate the mean charge per visit, the mean charge does not increase as rapidly with increasing family income. The strong relationship between family income and mean charge per hospital outpatient department visit in the real data has been attenuated by the imputed values.

The reason for this attenuation is shown in Figure III. Sixteen imputation classes were formed for the imputation of total charges for hospital outpatient department visits. Figure III shows mean charge for real data for the total sample and the subgroup with family incomes less than $5,000 in 1980. The low income group has lower mean charges than the total sample. Because family income was not one of the variables used to form imputation classes, low family income persons within an imputation class with missing hospital 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 occurs in almost every imputation class. When the real and imputed data are combined for persons with family incomes less than $5,000, 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 hospital 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 hospital outpatient department visits and family income is a regression toward the mean charge for real data for low and high income subgroups.

The results in Tables III and IV and Figure II demonstrate the effect that imputation can have on estimated means, on estimated sampling errors, and on relationships between variables. Several strategies for handling imputation in estimation are suggested by these findings. It is beyond the scope of this discussion to evaluate various strategies and indicate the reasons why one was chosen for this report. The strategy used in preparing estimates in this report was, despite the sizeable effects due to imputation noted here, to use all the data in all estimates. This strategy means that estimated means and totals presented in the report have been adjusted for item nonresponse, but sampling errors and relationships among some variables may be adversely affected by the imputation process. The reader should keep in mind that sampling errors for estimates that are subject to large amounts of item nonresponse may be substantially underestimated, and the strength of relationships between a variable receiving imputed values and a variable that was not used to form imputation classes may be attenuated by the imputation process.

Estimation Procedures

Sample estimators from NMCUES data, regardless of whether they are totals, means, medians, proportions, or standard errors, must account for the complexity of the sample survey design. Totals, means, or 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 which were made to compensate for missing items. The weighting adjustment factors, the imputations, and the estimated sampling errors for estimates with imputed values affect the results shown in this report.

A variety of estimators were used for the descriptive analyses. To illustrate the role of time adjustments, we consider six specific estimates that were used in the analysis:

1. An estimated total charge for a selected subgroup (e.g., high-volume users).
2. An estimated total population.
3. The mean charge per visit.
4. The mean charge per person.
5. The proportion of charges that fall in a certain range of charges.
6. The proportion of persons whose charges are less than or equal to a fixed level.

To define these estimators, suppose we introduce the following notation for these quantities for the *i*th person:

y_i = total charges for health care in 1980,
x_i = total number of medical visits for 1980,
w_i = nonresponse and undercoverage adjusted person weight,
t_i = time-adjustment factor (i.e., the proportion of days in 1980 that the person was an eligible member of the population),

$$d_i = \begin{cases} 1, \text{ if total charges are less than or equal to a fixed value,} \\ 0, \text{ otherwise,} \end{cases}$$

$$e_i = \begin{cases} 1, \text{ if the total charge is between two fixed values,} \\ 0, \text{ otherwise, and} \end{cases}$$

$$\delta_i = \begin{cases} 1, \text{ if the } i\text{th person is a member of a designated subgroup of the population,} \\ 0, \text{ otherwise.} \end{cases}$$

Estimating total charges or any quantity from NMCUES which 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. In particular,

$$\hat{y} = \sum w_i y_i \delta_i$$

is the estimated total charge for a particular service for a selected subgroup. On the other hand, for estimates of total population, a time-adjusted estimator is required such as

$$\hat{y}' = \sum w_i t_i \delta_i.$$

Thus, $\hat{y}'$ denotes an estimate of the 1980 average subgroup population, while $\hat{y}$ denotes the 1980 charges by a subgroup of the noninstitutionalized civilian population.

Estimated means may or may not need to include a time-adjustment factor in the denominator. For example, to estimate the mean charge *per visit* during 1980, no time adjustment is needed. Hence,

$$\bar{y} = \sum w_i y_i / \sum w_i x_i$$

can be used to estimate mean charge per visit. However, to estimate mean charge *per person*, a time adjustment is required in the denominator, because the denominator is actually an estimate of the total average population in 1980. In particular, the estimator has the form

$$\bar{y}' = \sum w_i y_i / \sum w_i t_i.$$

Estimates of mean charges for subgroups have a similar form with the indicator variable δ_i included in the numerator and denominator for the appropriate subgroup of interest.

Estimated proportions are, of course, means with an indicator variable in the numerator and a count variable in the denominator. Proportions may also have time adjustments not only in the denominator, 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

$$p' = \sum w_i d_i t_i / \sum w_i t_i$$

was used. Appropriate indicator variables were added to the numerator and denominator to make estimates for selected subgroups.

On the other hand, the estimated proportion of total charges between two fixed levels of charges does not require time adjustments in the numerator or the denominator. In particular,

$$p = \sum w_i y_i e_i / \sum w_i y_i$$

is the estimated proportion of all charges for persons that occurred between two levels of charges.

The Logistic Regression Model

One of the statistical methods used in this report is logistic multiple regression. The methodology is used to identify characteristics of individuals in the U.S. civilian noninstitutionalized population in 1980 that are predictive of use of hospital services or of high use of hospital services. Logistic regression is closely related to the standard regression methods but was used here because of the nature of the dependent variable in the models, a measure with only two possible outcomes (i.e., yes or no).

In the standard regression situation, the dependent variable Y is intervally scaled (i.e., continuous in distribution). The regression methodology is used to examine independent variables $X_1, X_2, \ldots, X_p$ which are statistically important predictors of the dependent variable Y. The ordinary regression model predicts Y as a linear combination of the $X_i's$, that is,

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p + \varepsilon$$

where ε is a random variable with mean zero and equal variance for all sample observations. The parameters $\beta_0, \beta_1, \ldots, \beta_p$ are often referred to as the slopes or (partial) regression coefficients and are estimated in the regression analysis.

Logistic regression methods are applied when the dependent variable is not continuous but is dichotomous (i.e., Y is only one of two possible values). Use of the standard regression methodology when the dependent variable is dichotomous is inappropriate for a variety of reasons: Besides violating a number of important assumptions of the regression model, predicted values for Y can fall outside the range of Y. The logistic regression approach addresses these deficiencies by examining not the dependent variable Y but the logit of the probability that the dependent variable assumes one of the two possible values. Suppose that Y can assume only two values, 1 or 2: The logit is defined to be the logarithm of the ratio of the probabilities that Y is 1 or 2:

$$\text{Logit}(Y) = \ln \left[P\{Y=1\}/P\{Y=2\}\right]$$

where ln[•] denotes the natural logarithm of the argument [•]. The logistic regression model then predicts the logit of Y as a linear combination of the $X_i's$:

$$\text{Logit}(Y) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p.$$

The β_i' are again slope or (partial) regression coefficients, but in this case denote the partial linear regression of the logit of Y on X_i given that the other $X's$ are in the model. Several methods are used to obtain estimates of the coefficients including an iterative method to derive a maximum likelihood estimate.

An alternative interpretation of these logistic regression coefficients can be made by observing that the logit of Y is actually the logarithm of the odds that an individual will be classified as $Y = 1$ rather than $Y = 2$. Suppose, for example, that $Y = 1$ corresponds to the event that an individual is hospitalized in 1980 and $Y = 2$ to the event that the individual is not hospitalized. Then $P\{Y = 1\} / P\{Y = 2\}$ denotes the odds that an individual will be hospitalized during the year and the logit is simply the logarithm of the odds ratio.

The logistic regression coefficients can be translated into statements about the ratios of odds for two different individuals with different characteristics X_i. Consider the following three indicator variables:

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

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

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

These three indicators combined use the individuals 35–54 years of age as a reference or comparison group against which the odds of being hospitalized, for example, 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, in Table B the logistic regression coefficient corresponding to X_1 is 0.1510. Persons under 35 years of age have somewhat higher odds of being hospitalized than persons 35–54 years of age (the reference group) when all the other variables in the model in Table B are controlled at an average value. Similarly, persons 55–74 years of age and 75 years of age and over have higher odds of being hospitalized than persons 35–54 years of age.

The coefficients can be used to predict the probability of hospitalization or high use of hospital care 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 to obtain a predicted logarithm of the odds ratio.
3. Convert the predicted logarithm of the odds ratio to obtain the predicted probability.

For example, suppose that it is desired to estimate the predicted probability of hospitalization in 1980 for a hypothetical individual using the estimated coefficients in Table B. Following the characteristics in Table B, suppose the person is a male, white or other race, reported good health status, 35–54 years of age, income that is 200 to 400 percent of poverty level, with private health insurance, residing in the northeast region, and with a usual source of care.

The logarithm of the odds ratio is estimated by summing the appropriate coefficients from the table. For characteristics that have two levels, such as sex, either a coefficient will be included in the sum or nothing will be added. For sex, for instance, males are the reference group, and the hypothetical person would not have a coefficient added to the logarithm of the odds ratio sum. On the other hand, for race the reference group is the black race, and the hypothetical individual would have the coefficient $\hat{\beta} = 0.0948$ would be added to the sum.

For characteristics with several levels, the coefficient for the reference group in the table (enclosed in parentheses) is zero, while the coefficients for the other groups of the characteristic are nonzero. Thus, for the hypothetical individual reporting good health status, the reference group is good health status and the health status coefficient for the person is zero. On the other hand, for type of health care coverage, the reference group is none or some mixture of part-year coverages. Because the person has private coverage, the private insurance coefficient $\hat{\beta} = 0.3468$ would be added to the sum.

The coefficients for poverty level must be handled in a slightly different way. The coefficient for the group with income that is 700 percent or more of the poverty level is determined by adding the negative value of the sum of the other coefficients to the logarithm of the odds ratio. For the hypothetical person with income 200–400 percent of the poverty level, the contribution of poverty level to the logarithm of the odds ratio for that person would be simply the value of the coefficient for that group, i.e., $\hat{\beta} = -0.0728$. If the hypothetical person had an income that was 700 percent or more of the poverty level the contribution of poverty level would be $-(0.1483 + -0.0728 + -0.1191) = -0.0436$.

Once the coefficients of each separate characteristic are determined and combined, the estimated logarithm of the odds ratio is calculated by adding the sum of coefficients to the constant term. This coefficient represents the effect for a person who has all the reference group characteristics. Thus, the logarithm of the odds ratio for the hypothetical individual is computed as the sum of the coefficients for the constant term, for the white and other race category, for the category with income 200–400 percent of the poverty level, for private insurance coverage, and for persons with a usual source of care:

$$(-2.9031) + (0.0984) + (-0.0728) + (0.3468) + (-0.0608) = -2.5915.$$

The last step to obtain a predicted probability of hospitalization for the hypothetical individual is to take the natural exponent of the logarithm of the odds ratio to obtain the estimated odds ratio:

$$[\hat{p}/(1-\hat{p})] = \exp(-2.5915) = 0.0749.$$

Finally, the estimated odds ratio is converted to an estimated probability as

$$\hat{p} = \frac{0.0749}{1 + 0.0749} = 0.0736.$$

That is, the hypothetical individual has predicted probability of hospitalization in 1980 of 7.4 percent.

The logistic regression coeffcients for these indicator variables can also be interpreted in terms of a ratio of odds. For example, using the three indicator variables for age, the function e^{β_1} is the ratio of the odds of hospitalization for an individual under 35 years of age to the corresponding odds that an individual 35–54 years of age is hospitalized. In Table B, this odds ratio is 1.1630 suggesting that persons under 35 years of age have approximately 16 percent higher odds of hospitalization than persons 35–54 years of age.

The logistic regression coefficients presented in this report were computed using a program for logistic regression analysis in the OSIRIS statistical software system (Computer Support Group, 1982) called DREG (Dichotomous REGression). The program uses an iterative maximum likelihood method to estimate the logistic regression coefficients and has a feature that 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 selections.

At present, there is no accessible software for estimating the standard errors of logistic regression coefficients under a complex sample design. An *ad hoc* procedure was used to adjust the estimated standard errors from the DREG program to account for the complex NMCUES sample design. For each logistic regression model, the identical 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). For each coefficient, the ratio of the actual standard error to the corresponding standard error computed under the assumptions of independent sample selection is computed. Assuming that the same ratio for logistic regression coefficients is similar, 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 B, the estimated odds ratio for persons under 35 years of age (relative to persons 35–54 years of age) is 1.1630 with 95-percent confidence limits from 1.0321 to 1.3105. Thus, with 95-percent confidence one can conclude that the odds of being hospitalized for an individual under 35 years of age are greater than for an individual 35–54 years of age.

Finally, 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 logistic regression methodology. For each individual in the sample, the probability that that individual was hospitalized, for example, can be estimated from the logistic regression coefficients by substituting the individual's values for each X_i into the model, computing the log odds of hospitalization for that individual, and converting the log odds into the corresponding probability. For a particular individual, suppose that they were hospitalized, i.e., $Y = 1$. The probability that this individual was hospitalized under the logistic regression model can then be computed. Similarly, for every other individual in the sample, the probability that they were hospitalized or not hospitalized, as the case may be for each person, can be computed. Probabilities close to one denote that the predictive power of the model is good for that individual, while 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 predictive powers for all sample individuals can be computed. In this case, a simple average is not appropriate since the logistic regression model operates on a logarithmic scale rather than a linear one as in the standard regression method. Let $\hat{P}(Y_i = y_i)$ denote the predicted probability from the logistic model that for the ith individual y_i is the observed value. The predictive power of the model over all individuals can be computed as the geometric mean of these probabilities,

$$\hat{\pi} = \left[\pi \hat{P}(Y_i = y_i) \right]^{1/n},$$

where n denotes the sample size. If all individuals' observed values are predicted well from the model, the probabilities $\hat{P}(Y_i = y_i)$ will be close to one and $\hat{\pi}$ will be close to one. Since $\hat{\pi}$ is a measure of how well the model fits the data, then $\hat{\pi}_e = 1 - \hat{\pi}$ 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 by an intercept-only model. The importance of the X_i's to 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 $X_1, X_2, \ldots, X_p$ 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. Such a measure was used in this report to examine alternative models and assess whether the predictors were adding predictive power to the model.

Appendix IV Sampling Errors

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

Preparation of sampling errors for every estimate in this report is a sizeable task. A more difficult task, though, is to find a way to present sampling error estimates that would not greatly increase the length of the report or would not make it difficult to distinguish the estimates from their standard errors in a single table. Rather than compute and display standard errors for every estimate in this report, standard errors were computed for a subset of estimates. A set of functions was fitted to these estimated standard errors to determine whether a model could be identified that would allow computation of standard error using the function that would be reasonably close to the estimated standard error.

This appendix is designed to provide the reader of the report with these summary formulae derived from the estimated standard errors, which can be used to approximate the standard error for any given estimate in the report. The formulae have been designed to allow the reader to compute an estimated standard error using an electronic calculator with basic arithmetic operators and a square root function. The computed estimate will be an average or smoothed estimate of the actual standard error of the estimate.

The formulae for standard error estimates are presented for three types of estimates found in the report:

1. Totals or aggregates (e.g., total charges for ambulatory visits made during 1980; total persón years for males).
2. Means (e.g., mean charge per ambulatory visit).
3. Proportions and percents (e.g., percent of persons having no hospitalizations during 1980).

The reader also may be interested in making comparisons between point estimates from two different subgroups of the population. Formulae are also given for computing standard errors for two types of comparisons that are made in this report:

a. Comparisons of two mutually exclusive subgroups (e.g., comparing the percent of females who were low users of hospital services and the percent of females who were high users, where low-user and high-user subgroups have no members in common).
b. Comparisons between a subgroup and a larger group in which the subgroup is contained (e.g., comparing the mean charge per hospital day for high users of hospital services and for all users of hospital services).

The standard error of a difference is based on the standard error of the totals, means, proportions, or percents of interest and ignores certain covariances between estimates that typically are small relative to the standard errors of the estimates themselves.

The standard errors calculated from the formulae in this appendix can be used to form intervals for which confidence statements can be made for estimates from all possible samples drawn in exactly the same way as NMCUES. The confidence level is determined by multiplying the estimated standard error by a constant derived from the standardized normal probability distribution. In particular, for the estimate $\hat{\theta}$ with estimated standard error $S_{\hat{\theta}}$, the upper limit for a $(1 - \alpha) \times 100$-percent confidence interval can be formed by adding $z_{\alpha/2}$ times $S_{\hat{\theta}}$ to $\hat{\theta}$; the lower limit is formed by subtracting $z_{\alpha/2}$ times $S_{\hat{\theta}}$ 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 $\alpha \doteq 0.05$ can be formed with $z_{0.025} = 1.96$; a 99-percent confidence interval (i.e., $\alpha = 0.01$) uses $z_{0.005} = 2.346$. Illustrations of these calculations are provided in the discussion section for each formula.

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

Totals

Let $\hat{y}$ 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

$$S_{\hat{y}} = \left[a\hat{y} + b\hat{y}^2\right]^{1/2},$$

where a and b are constants chosen from Table V for the particular estimate of interest. This formula was derived from a study of the relationship between the estimated total $\hat{y}$ and its standard error $S_{\hat{y}}^2$ in which a parabolic or quadratic relationship was observed.

As an illustration of the use of this formula, suppose that the standard error of the estimated number of persons having no ambulatory visits during 1980 is needed. From Table 1, $\hat{y}$ = 46,716,000, the estimated total number of person years accumulated in 1980 by persons having no ambulatory visits during 1980. From Table V, we obtain the coefficients a = 25,011 and b = 0.00048043 to use in the formula to calculate the standard error of $\hat{y}$. The estimated standard error is then computed as

$$S_{\hat{y}} = \left[(25{,}011)(46{,}716{,}000) + (0.00048043)(46{,}716{,}000)^2\right]^{1/2}$$
$$= \left[(1.1684 \times 10^{12}) + (1.0484 \times 10^{12})\right]^{1/2}$$
$$= 1{,}488{,}925.$$

This estimated standard error for the total $\hat{y}$ can be used to create confidence intervals for the number of persons having no ambulatory visits. 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 exactly in the same way as the NMCUES, the estimated number of persons having no ambulatory visits during 1980 will be between 45,227,075 and 48,204,925. 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 the NMCUES, the estimated number of persons having no ambulatory visits would be between 43,797,707 and 49,634,293.

Table V

Coefficients for standard error formula for estimated aggregates or totals

Estimator	Coefficient a	Coefficient b
Person years	2.5011×10^4	4.8043×10^{-4}
Charges	1.0986×10^8	4.5524×10^{-4}
Visits or acquisitions	4.6408×10^2	5.7634×10^{-1}

Means

A number of means for different types of measures are presented in this report. Despite the variety of measures presented, a single formula is recommended for calculating an estimated standard error for a mean. The formula given here is based on the assumption that the standard error of the mean is determined by two quantities, the population variance and the effect of the sample design on the variances. The population variance for weighted survey data with weights w_i is estimated as

$$\hat{s}^2 = \frac{\sum w_i(y_i - \bar{y})^2}{(n-1)\sum w_i},$$

where n is the size of the sample, y_i denotes the value of the characteristic Y for the ith sample person, and $\bar{y}$ is the weighted sample mean. The effect of the sample design on the variance of a sample mean is called the design effect or 'deff' (Kish, 1965), and is often expressed as

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

where a is the number of clusters in the sample design and roh is a measure of within cluster similarity among observations from the same cluster.

The formula recommended for estimating the standard error of a mean in this report is a function of both the estimated population variance and the design effect. In particular, the estimated standard error for a mean $\bar{y}$ can be calculated as

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

where $\hat{n}$ is the estimated population total for the subgroup under consideration and 1,795,637 represents the number of clusters (a = 138) times the average basic person weight. Consequently, $\hat{n}$/1,795,637 is an estimator for n/a in the expression for deff. The values of roh and $\hat{s}^2$ for a variety of means appearing in this report can be obtained from Table VI. The table provides, for example, values of roh and $\hat{s}^2$ for mean charges and mean utilization measures of various types.

As an illustration, suppose that the standard error of the mean number of hospital days for high users of hospital services is needed. From Table 1, $\bar{y}$ = 38.5, and from Table VI under the entry for 'Mean visits per person, Hospital days' the values roh = 0.013098 and $\hat{s}^2$ = 8.5018 × 10^5 are obtained. There were an estimated $\hat{n}$ = 3,837,000 persons who were high users of hospital services. Substituting these values into the expression for $S_{\bar{y}}$,

$$S_{\bar{y}} = \left[\left[1 + (\frac{3{,}837{,}000}{1{,}795{,}637} - 1)(0.013098)\right] \cdot \frac{8.5018 \times 10^5}{3{,}837{,}000}\right]^{1/2}$$
$$= \left[[1 + (2.1368 - 1)(0.013098)]\,(0.22157)\right]^{1/2}$$
$$= [(1.0149)\,(0.22157)]^{1/2}$$
$$= 0.47420.$$

That is, the standard error of the mean number of hospital days for persons with high use of hospital services is 0.5.

Approximate confidence intervals may 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 the estimated mean number of hospital days for persons with

Table VI

Values of *roh* and $\hat{s}^2$ for standard error formula for estimated means

Estimator	*roh*	$\hat{s}^2$
Mean charge per visit		
Ambulatory visits	0.018777	2.4613×10^7
Hospital days	0.018777	1.9938×10^9
Prescribed medications	0.018777	3.8099×10^5
Mean visits per person		
Ambulatory visits	0.048246	1.6398×10^6
Hospital days	0.013098	8.5018×10^5
Prescribed medications	0.048246	1.6651×10^6
Mean charge per person		
Ambulatory visits	0.029644	2.5650×10^9
Hospital stays	0.029644	6.1652×10^{10}
Prescribed medications	0.029644	2.4323×10^8

high use of hospital services, 1.96 times the estimated standard error is added to and subtracted from the estimated mean $\hat{y} = 38.5$. In this case, the 95-percent interval ranges from 37.6 to 39.4.

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 effectively becomes equal to one. Thus, when $\hat{n} \leq 1{,}795{,}000$, the design effect portion of the standard error formula is not necessary, and the estimated standard error can be calculated simply as

$$S_{\hat{y}} = \left[\hat{s}^2/\hat{n}\right]^{1/2}$$

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

Proportions and Percents

The standard error of a proportion is computed using a formula similar to that recommended for the standard error of a mean. Let $\hat{p}$ denote the estimated proportion for which a standard error is needed. The standard error for $\hat{p}$ is calculated as

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

where $\hat{n}$ is the estimated sample size on which the proportion is based, *roh* is a value selected from Table VII, and the constant 13,012 is the average time-adjusted weight for all persons in the sample. For proportions, the population variance can be estimated simply as

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

and hence, can be estimated directly from the sample proportions themselves (i.e., no value of $\hat{s}^2$ is needed in Table VII). The design effect, the ratio of the actual sampling variance for the estimated proportion to the variance that would be achieved for a simple random sample of the same size, is calculated for proportions in the same way it was calculated for means.

As an illustration of the use of the formula for $S_{\hat{p}}$, consider obtaining the standard error for the proportion of persons ($\hat{p} = 0.572$) who had no ambulatory visits during 1980 and rate their health as excellent (see Table 18). To calculate the standard error for percents, the same formula may be used as for proportions, after the percent has been divided by 100. There are an estimated $\hat{n} = 46{,}716{,}000$ persons having no ambulatory visits (see Table 18), and $roh = 0.0057805$ is obtained from Table VII. Substituting these values into the formula for $S_{\hat{p}}$,

$$S_{\hat{p}} = \left[\left[1 + \left(\frac{46{,}716{,}000}{1{,}795{,}637} - 1\right)(0.0057805) \right] \frac{13{,}012 \cdot (0.572)(1-0.572)}{46{,}716{,}000} \right]^{1/2}$$

$$= \left[\left[1 + (25.016)(0.0057805) \right] \frac{3{,}185.5}{46{,}716{,}000} \right]^{1/2}$$

$$= \left[(1.1446)\,(6.8189 \times 10^{-5}) \right]^{1/2}$$

$$= 0.0088345.$$

Because $S_{\hat{p}} = 0.0088345$ is the estimated standard error for the proportion $\hat{p} = 0.572$, simply multiply $S_{\hat{p}}$ by 100 for a standard error of 0.88345 for the percent 57.2.

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. In this case, the 95-percent interval ranges from 55.5 to 58.9 percent of those persons having no ambulatory visits during 1980 rating their health as excellent.

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

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

as described for the standard error of a mean in the previous section. For example, 88.1 percent of persons 65 years of age or over having high use of hospital services during 1980 received Medicare supplemented by other health care

Table VII

Values of *roh* for standard error formula for estimated proportions

Estimator	*roh*
Person years	
Proportion of users of hospital days	0.0057805
Proportion of users of ambulatory services	0.0057805
Proportion of users of prescribed medicines	0.0057805
Demographic subgroups (e.g., age, race, sex)	0.069992

coverage (see Table 14). For the $\hat{n}$ = 1,580,000 estimated persons in this subgroup (see Table 14), the standard error of the proportion associated with this percent is estimated as

$$\left[\frac{13{,}012\cdot(0.881)(1-0.881)}{1{,}580{,}000}\right]^{1/2} = 0.029384\ .$$

A 95 percent confidence interval for the estimated percent is calculated by multiplying this estimated standard error by $100\cdot(1.96) = 196$ and adding the result to and subtracting the result from the percent. Thus, the 95 percent interval ranges from 82.3 to 93.9 percent.

Mutually Exclusive Subgroup Differences

Many comparisons between the same estimate for two different subgroups in the population are made in this report. Let $\hat{d} = \hat{\theta}_1 - \hat{\theta}_2$ denote the difference between two subgroup estimates, where $\hat{\theta}_1$ and $\hat{\theta}_2$ are the estimates for the two subgroups. For example, suppose that the proportion of persons with high use of hospital services having family incomes less than \$15,000 is to be compared with the proportion of persons with low use of hospital services having family incomes less than \$15,000 (see Figure 5). Then, $\hat{\theta}_1 = \hat{p}_1 = 0.564$ for high users, $\hat{\theta}_2 = \hat{p}_2 = 0.408$ for low users, and $\hat{d} = \hat{p}_1 - \hat{p}_2 = 0.156$. The standard error of this difference is computed as

$$S_{\hat{d}} = \left[S^2_{\hat{\theta}_1} + S^2_{\hat{\theta}_2}\right]^{1/2},$$

where $S^2_{\hat{\theta}_1}$ and $S^2_{\hat{\theta}_2}$ are the estimated sampling variances for $\hat{\theta}_1$ and $\hat{\theta}_2$, respectively. (This formula ignores the nonzero covariance between $\hat{\theta}_1$ and $\hat{\theta}_2$ that arises in complex samples such as the NMCUES. This covariance is typically positive and small relative to the variances themselves. Ignoring the covariance will result in standard errors for differences that are on average somewhat larger than the actual standard errors.)

From Table 5, $\hat{n}_1$ = 3,837,000 and $\hat{n}_2$ = 5,242,000, and from Table VII, *roh* = 0.0057805. Hence,

$$S_{\hat{p}_1} = \left[\left[1 + \left(\frac{3{,}837{,}000}{1{,}795{,}637} - 1\right)(0.0057805)\right] \cdot \frac{13{,}012\cdot(0.564)(1-0.564)}{3{,}837{,}000}\right]^{1/2}$$

$$= 0.028972\ .$$

$$S_{\hat{p}_2} = \left[\left[1 + \left(\frac{5{,}242{,}000}{1{,}795{,}637} - 1\right)(0.0057805)\right] \cdot \frac{13{,}012\cdot(0.408)(1-0.408)}{5{,}242{,}000}\right]^{1/2}$$

$$= 0.024621\ .$$

Hence, the standard error of the difference is computed as

$$S_{\hat{d}} = \left[(0.028972)^2 + (0.024621)^2\right]^{1/2} = 0.038021\ .$$

This standard error can be used to form an approximate confidence interval for the difference in the same manner described previously for estimates of totals, means, proportions, and percents. In this instance, the 95 percent confidence interval is from 0.081 to 0.231. Since this interval does not include the value zero, one could conclude with 95-percent confidence that the proportion of persons having family incomes less than \$15,000 differ for the two use groups. In other words, the chances are only 5 out of 100 that the difference over a large number of identical surveys will be equal to zero.

Subgroup to Total Group Differences

Another type of comparison made in this report is between an estimate for a subgroup and the same estimate for a group which contains the subgroup. 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. The standard error of the difference is computed as

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

where $S_{\hat{\theta}_1}$ denotes the standard error of the estimator $\hat{\theta}_1$, 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., $S^2_{\hat{\theta}_1}$). This assumption results in an estimated standard error for the difference that is on average somewhat larger than the actual standard error.)

For example, suppose that the standard error of the difference between the proportion of high users of ambulatory care living in the South and the proportion of all persons living in the South is needed. From Table 9, $\hat{\theta}_1 = \hat{p}_1 = 0.208$, $\hat{\theta}_T = \hat{p}_T = 0.312$, $\hat{n}_1 = 10{,}024{,}000$, and $\hat{n}_T = 222{,}824{,}000$. Using the formula for estimating the standard error of the proportion and the value from Table VII (i.e., *roh* = 0.0057805),

$$S_{\hat{p}_1} = \left[\left[1 + \left(\frac{10{,}024{,}000}{1{,}795{,}637} - 1\right)(0.0057805)\right] \cdot \frac{13{,}012\cdot(0.208)(1-0.208)}{10{,}024{,}000}\right]^{1/2}$$

$$= 0.014816\ .$$

Hence, the standard error of the difference, $\hat{d} = 0.208 - 0.312 = -0.104$, is computed as

$$S_{\hat{d}} = 0.014816\left[1 - (10{,}024{,}000\ /\ 222{,}824{,}000)\right]^{1/2} = 0.014479\ .$$

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 -0.132 to -0.076. One would conclude with 95-percent confidence, that fewer high users of ambulatory care live in the South, because this confidence interval does not include zero.

Appendix V
Definition of Terms

Age—The age of the person as of January 1, 1980. Babies born during the survey period were included in the youngest age category.

Ambulatory care visit—A direct personal exchange between an ambulatory patient and a health care provider. The visit may have taken place in the provider's office, hospital outpatient department, emergency room, clinic, health center, or the patient's home. Services may have been rendered by a physician, chiropractor, podiatrist, optometrist, psychologist, social worker, nurse, or other ancillary personnel.

Average length of stay—The average length of stay is the total number of hospital days accumulated at time of discharge by patients discharged during the year divided by the number of patients discharged.

Bed-disability day—A bed-disability day is one in which a person stays in bed for more than half of the daylight hours because of a specific illness or injury. All hospital days for inpatients are considered to be bed-disability days even if patient was not actually in the bed at the hospital.

Condition—Any entry on the questionnaire that describes a departure from a state of physical or mental well-being. It is any illness, injury, complaint, impairment, or problem perceived by the respondent as inhibiting usual activities or requiring medical treatment. Pregnancy, vasectomy, and tubal ligation were not considered to be conditions; however, related medical care was recorded as if they were conditions. Neoplasms were classified without regard to site. Conditions, except impairments, are classified by type according to the *Ninth Revision of the International Classification of Diseases* (World Health Organization, 1977) as modified by the National Health Interview Survey Medical Coding Manual (NCHS, 1979); these modifications make the code more suitable for a household interview survey. Impairments are chronic or permanent defects, usually static in nature, that result from disease, injury, or congenital malformation. They represent decrease or loss of ability to perform various functions, particularly those of the musculoskeletal system and the sense organs. Impairments are classified by using a supplementary code specified in the coding manual. In the supplementary code, impairments are grouped according to type of functional impairment and etiology.

Disability—Disability is the general term used to describe any temporary or long-term reduction of a person's activity as a result of an acute or chronic condition.

Disability day—Short-term disability days are classified according to whether they are days of restricted activity, bed-disability days, hospital days, or work-loss days. All hospital days are by definition days of bed disability; all days of bed disability are by definition days of restricted activity. The converse form of these statements is, of course, not true. Days lost from work applies only to the working population, but these too are days of restricted activity. Hence, restricted-activity days is the most inclusive term used to describe disability days.

Education of head of family—The years of school completed by the head of family, when the family head was 17 years of age and over. Only years completed in regular schools, where persons are given a formal education, were included. A "regular" school is one that advances a person toward an elementary or high school diploma or a college, university, or professional school degree. Thus, education in vocation, trade, or business schools outside the regular school system was not counted in determining the highest grade of school completed.

Family—A group of people living together related to each other by blood, marriage, adoption, or foster care status. An unmarried student 17–22 years of age living away from home was also considered part of the family even though his or her residence was in a different location during the school year.

Family head—At the time of the first interview, the respondent for the family was asked to designate a "family head." If no head was designated or this information was missing, a family head was imputed.

Family income in 1980—Each member of a family is classified according to the total income of the family of which he or she is a member. Because some persons changed families during the year, their family income is defined as the income of the family they were in the longest. If a family did not exist for the entire year, the family income is adjusted to an annual basis by dividing actual income by the proportion of the year the family existed. Unrelated persons are classified according to their own income. For each person, 12 categories of income were collected, including income from employment for persons 14 years of age and over and income from various government programs, pensions, alimony or child support, interest, and net rental income. Where information was missing it was imputed. For persons who were members of more than one family, their total income was allocated to each family in proportion to the amount of time they were in that family.

Health care coverage—Twelve mutually exclusive categories of health care coverage were developed. Because of the importance and extent of Medicare coverage for persons 65 years of age and over, the population was first divided into those under 65 years of age and those 65 years of age and over. For persons under 65 years of age, coverage is divided into four mutually exclusive categories: Coverage all year from a single source, coverage all year from a mixture of sources, coverage only part of the year, and no health care coverage. For those under 65 years of age and covered all year from a single source, three subcategories of coverage were designated: Private insurance only, such as commercial carrier or Blue Cross; Medicaid only; and other public programs including Medicare, CHAMPUS/CHAMPVA, Indian Health Service, and other programs covering the cost of health care. Persons in the part-year-coverage category had health care coverage under either a private insurance policy or a public program, but the coverage did not extend throughout the year.

People 65 years of age and over are partitioned into two major coverage categories: Those covered by Medicare and not covered by Medicare. The former group is subdivided into persons having only Medicare coverage, those who have supplemented their Medicare with private policies, and those who are covered not only by Medicare but also by Medicaid, the Indian Health Service, or other public program. The second subgroup, those not having Medicare, is divided into persons who have some other type of health care coverage, whether private or public, and those who have no coverage at all.

For the multivariate analyses, categories that do not distinguish between age groups were formed, as follows:

- Multiple Public: Coverage by more than one public program during the year; in most, but not all cases denotes *simultaneous* coverage by two programs, such as coverage by both Medicare amd Medicaid at the same time.
- Single Public: Coverage by only one public program, either Medicare, Medicaid, Indian Health Service, CHAMPUS, or other government program.
- Private and Public: Coverage during the year by at least one private insurance plan and at least one public program; in most cases both types of coverage overlay for at least part of the year.
- Private Only: Coverage during the entire year by private insurance only.
- None or other: Includes those with no health care coverage during the year; those with coverage for part of the year by only one source, and, for those 65 years of age and over, coverage other than Medicare.

Hospital admission—The formal acceptance by a hospital of a patient who is provided room, board, and regular nursing care in a unit of the hospital. Included as a hospital admission is a patient admitted to the hospital and discharged on the same day. Also included is a hospital stay resulting from an emergency department visit.

Hospital days—The total number of inpatient days accumulated at time of discharge by patients discharged from short-stay hospitals during a year constitute hospital days. A stay of less than 1 day (patient admission and discharge on the same day) is counted as 0 days in the summation of hospital days. For patients admitted and discharged on different days, the number of days of care is computed by counting all days from (and including) the date of admission to (but not including) the date of discharge.

Hospital outpatient department visit—A face-to-face encounter between an ambulatory patient and a medical person. The patient comes to a hospital-based ambulatory care facility to receive services and departs on the same day. If more than one department or clinic is visited on a single trip, each department or clinic visted is counted as a separate visit.

Household—Occupants of a housing unit or group quarters that was included in the sample. This could have been one person, a family of related people, a number of unrelated people, or a combination of related and unrelated people.

Housing unit—A group of rooms or a single room occupied or intended for occupancy as separate living quarters: that is, (1) the occupants did not live and eat with any other persons in the structure, and (2) there was either direct access from the outside or through a common hall, or there were complete kitchen facilities for the use of the occupants only.

Key person—A key person was (1) an occupant of a national household sample housing unit or group quarters at the time of the first interview; (2) a person related to and living with a State Medicaid household case member at the time of the first interview; (3) an unmarried student 17–22 years of age living away from home and related to a person in one of the first two groups; (4) a related person who had lived with a person in the first two groups between January 1, 1980, and the round 1 interview, but was deceased or had been institutionalized; (5) a baby born to a key person during 1980; or (6) a person who was living outside the United States, was in the Armed Forces, or was in an institution at the time of the round 1 interview but who had joined a related key person.

Limitation of activity—A functional limitation score was developed for classifying limitation of activity. It ranges from 0, indicating no limitation of activity, to 8, meaning severe activity limitation, and 9, indicating death during the survey period. The functional limitation score was developed from responses to a battery of questions designed to assess ability to perform various common functions such as walking, driving a car, and climbing stairs. For NMCUES, these questions were asked of persons 17 years of age and over.

Mean charge per unit of service—The arithmetic mean calculated from charges reported by the household respondent without consideration for the amount actually paid or the source of payment. Zero charges were assigned to service units that the household reported as free from the provider in response to three separate questions.

Nonkey person—A person related to a key person who joined him or her after the round 1 interview but was part of the civilian noninstitutionalized population of the United States at the date of the first interview.

Patient—A person formally admitted to the inpatient service of a short-stay hospital for observation, care, diagnosis, or treatment. In this report the number of patients refers to the number of discharges during the year, including any multiple discharges of the same individual from one or more short-stay hospitals. The terms "patient" and "inpatient" are used synonymously.

Per capita charges—Calculated by dividing the total charges by the number of people in the reference population.

Perceived health status—The family respondent's judgment of the health of the person compared with others the same age, as reported at the time of the first interview. The categories were excellent, good, fair, or poor.

Poverty status—The poverty status in 1980 was calculated by dividing the person's family income in 1980 by the appropriate 1980 nonfarm poverty level threshold and converting it to percent. These thresholds, as used by the U.S. Bureau of Census, are determined by the age and sex of the family head and the average number of persons in the family.

Prescribed medicine acquisitions—The number of times a person had a prescription filled, regardless of whether it was an initial filling or a refill of a prescription.

Race—The race of people 17 years of age and over reported by the family respondent; the race of those under 17 years of age derived from the race of other family members. If the head of the family was male and had a wife who was living in the household, her race was assigned to any children under 17 years of age. In all other cases, the race of the head of the family (male or female) was assigned to any children under 17 years of age. Race is classified as "white," "black," or "other." The "other" race category includes American Indian, Alaskan Native, Asian, Pacific Islander. The category "white and other" includes the categories "white" and "other."

Region—NORTHEAST: Maine, New Hampshire, Vermont, Massachusetts, Rhode Island, Connecticut, New York, New Jersey, Pennsylvania; NORTH CENTRAL: Michigan, Wisconsin, Ohio, Indiana, Illinois, Minnesota, Iowa, Missouri, North Dakota, South Dakota, Nebraska, Kansas; SOUTH: Delaware, Maryland, District of Columbia, Virginia, West Virginia, North Carolina, South Carolina, Georgia, Florida, Kentucky, Tennessee, Alabama, Mississippi, Arkansas, Louisiana, Oklahoma, Texas; WEST: Montana, Idaho, Wyoming, Colorado, New Mexico, Arizona, Utah, Nevada, Washington, Oregon, California, Alaska, Hawaii.

Reporting unit—The basic unit for reporting data in the household components of NMCUES. A reporting unit consisted of all related people residing in the same housing unit or group quarters. One person could give information for all members of the reporting unit.

Restricted-activity day—A restricted-activity day is one on which a person cuts down on his usual activities for the whole of that day because of an illness or an injury. The term "usual activities" for any day means the things that the person would ordinarily do on that day. A day spent in bed or a day home from work because of illness or injury is, of course, a restricted-activity day.

Round—A round was the administrative term used to designate all interviews that occurred within a given period of time and that used the same instruments and procedures.

Work-loss day—A work-loss day is a day on which a person did not work at his or her job or business because of a specific illness or injury. The number of days lost from work is determined only for persons 17 years of age and over who reported that at any time during the survey period they either worked at or had a job or business.