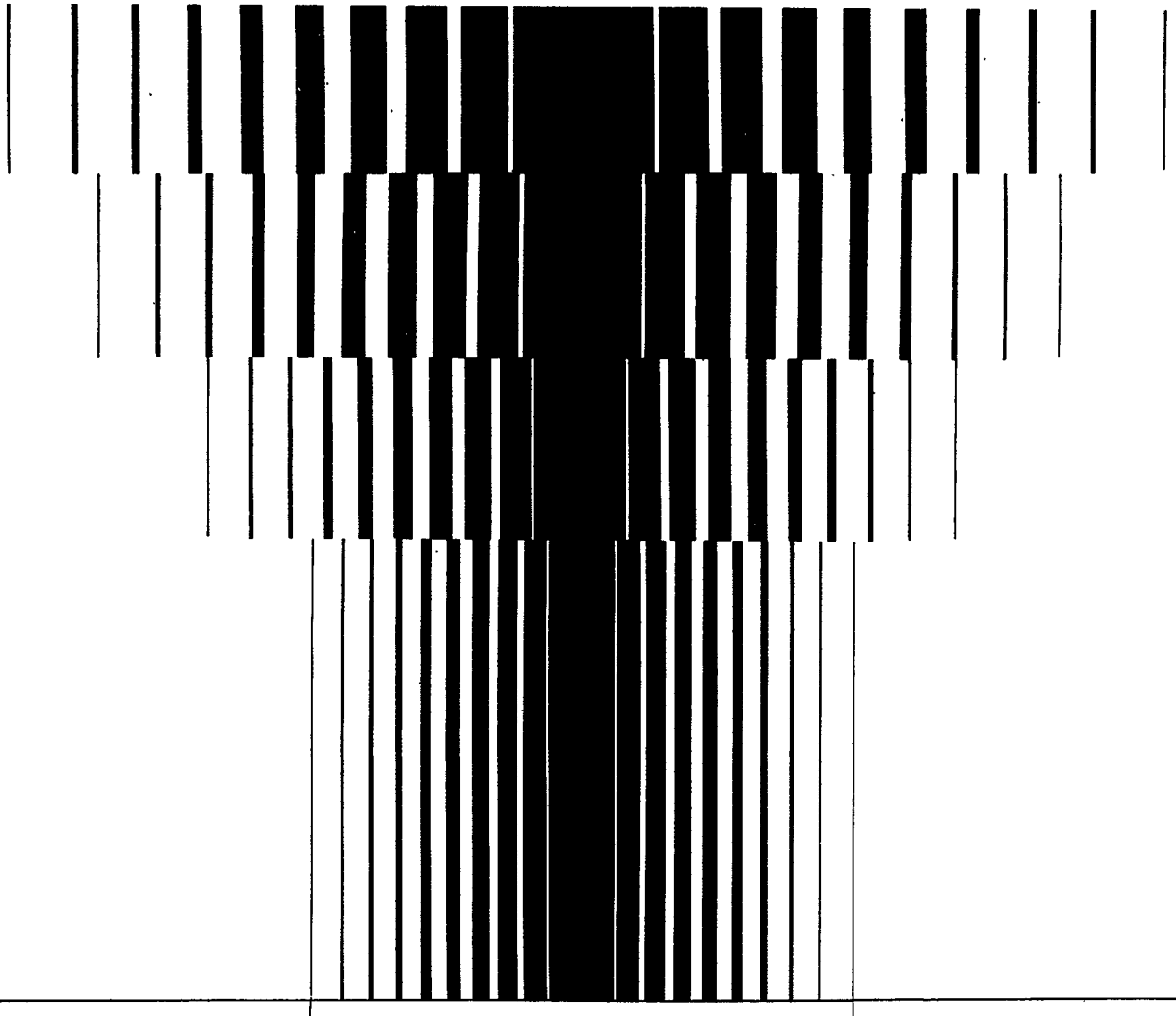
National Medical Care Utilization and Expenditure Survey

Determinants of Financially Burdensome Family Health Expenses United States, 1980

Series C, Analytical Report No. 6

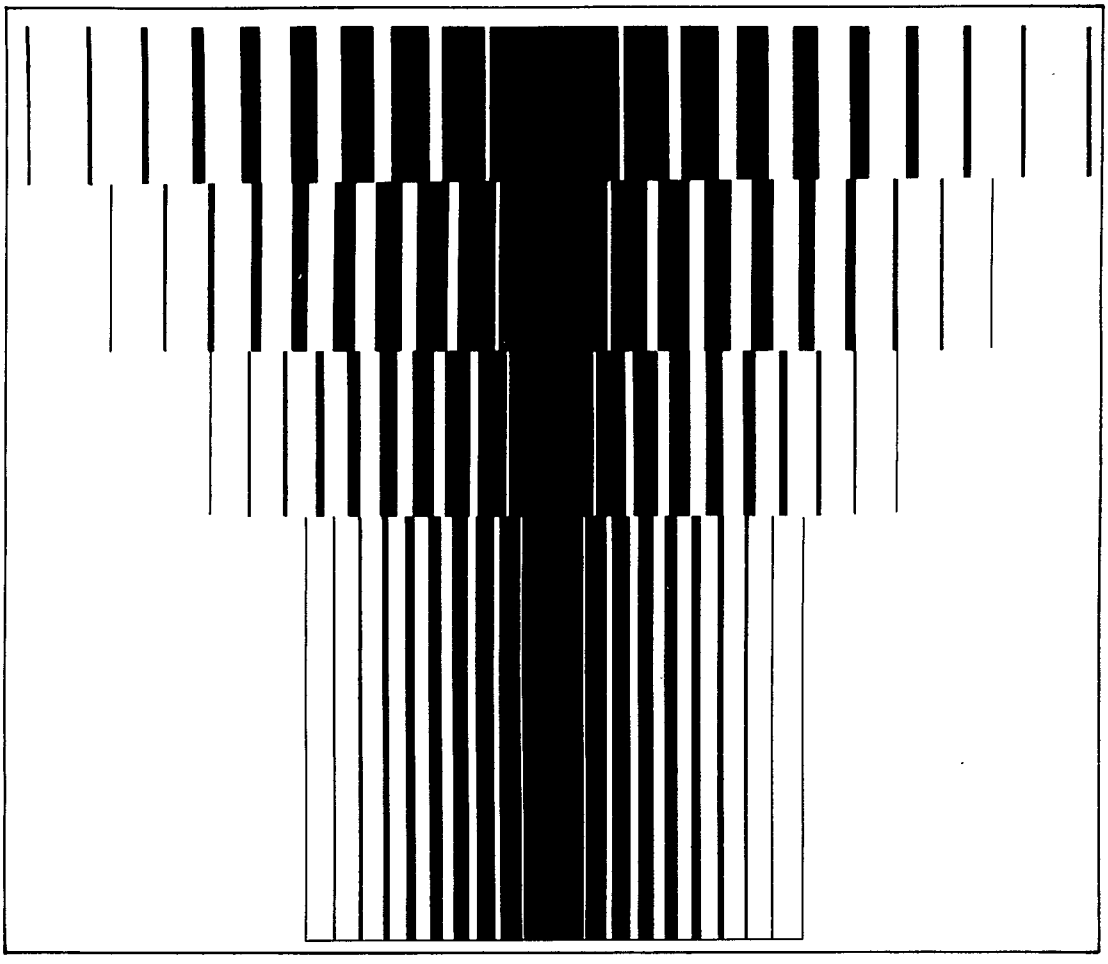


Health Care Financing
Administration
Office of Research
and Demonstrations

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

Determinants of Financially Burdensome Family Health Expenses United States, 1980

Series C, Analytical Report No. 6



U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES

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

April 1988

National Medical Care Utilization and Expenditure Survey

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

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

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

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

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

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

Suggested Citation

Dicker, M., and Sunshine, J. H.: Determinants of financially burdensome family health expenses, United States, 1980. *National Medical Care Utilization and Expenditure Survey.* Series C, Analytical Report No. 6. DHHS Pub. No. 88–20406. National Center for Health Statistics, Public Health Service. Washington. U.S. Government Printing Office, Apr. 1988.

Library of Congress Cataloging-in-Publication Data

Dicker, Marvin.
Determinants of financially burdensome family health expenses.
(National Medical Care Utilization and Expenditure Survey. Series C, Analytical report ; no. 6) (DHHS publication ; no. 88–20406)
Bibliography: p.
1. Family—Medical care—United States—Costs—Statistics. 2. Family—Medical care—United States—Costs—Statistical methods. I. Sunshine, Jonathan H. II. National Center for Health Statistics (U.S.) III. Title. IV. Series: National Medical Care Utilization and Expenditure Survey (Series). Series C, Analytical report ; no. 6. V. Series: DHHS publication ; no. 88–20406. [DNLM: 1. Expenditures, Health—United States. 2. Financing, Personal—United States. W2 A N224n no. 6]
RA407.3.D47 1988 338.4'33621 87–600309
ISBN 0–8406–0382–7

Contents

List of Text Tables

Symbols

-	No families with these characteristics in sample
*	Potential reliability problem; statistic is based on sample size of fewer than 50 or has relative standard error greater than 30 percent
. . .	Category not applicable

Determinants of Financially Burdensome Family Health Expenses: United States, 1980

By Marvin Dicker, Ph.D.,
National Center for Health Statistics,
and Jonathan H. Sunshine, Ph.D.,
Applied Management Sciences, Inc.

Executive Summary

This report focuses on two questions of current interest to policymakers. First, "What percent of U.S. families experience financially burdensome health expenses?" and, second, "What are the determinants of financially burdensome health expenses among U.S. families?" The first question is addressed by examining how the distribution in the United States of families with financially burdensome health expenses is affected by six different possible measures of financial burden. The second question is addressed by using multiple regression techniques on one of the measures selected as a preferred measure.

The data used are from the family data files of the 1980 National Medical Care Utilization and Expenditure Survey (NMCUES). This report presents data on approximately 5,000 multiple-person families interviewed in this longitudinal survey. It provides a separate analysis for each of three socioeconomic family populations that have consistently been of interest to policymakers. These are (1) older families (defined for this report as all U.S. multiple-person families with a member 65 years of age or over); (2) younger, lower-income families (defined as all U.S. multiple-person families below 200 percent of the poverty level in 1980 and with all members under 65 years of age); and (3) younger, better-off families (defined as all U.S. multiple-person families at 200 percent of the poverty level or higher in 1980 and with all members under 65 years of age).

NOTE: The authors are grateful for the support received during all stages of the preparation of this report from our colleagues at both the National Center for Health Statistics and Applied Management Sciences, Inc. At the National Center for Health Statistics, Thomas Hodgson consulted and advised on health economics and econometrics, and Cecelia Snowden consulted and advised on research methods and multiple regression techniques. Margaret Cooke, Pennifer Erickson, and Robert A. Wright reviewed several drafts of the manuscript and also made many useful suggestions. Editors in the Publications Branch provided valuable assistance during all stages of this report.

At Applied Management Sciences, Inc., Colleen Goodman, JoAnn Kuchak, and Alfred J. Meltzer provided executive management, skillfully making the firm's resources available to meet the changing needs of the project. Daniel Cohn provided programming skills as the staff member principally responsible for data processing. Dr. James Bethel acted as statistical consultant for the project.

Overall Findings: The Six Measures

Two general conceptual approaches have been used in the literature to assess financially burdensome health expenses. The first approach measures financial burden by the size of a family's health bill in dollars. The second approach focuses on a family's ability to pay its health bill, and it measures financial burden as a ratio of health expenses to family income. There is no agreement on which of the two approaches is preferable and also no agreement on which of several operational measures in each category is the most appropriate. In order to shed light on this controversy, this report compares six potentially useful operational measures of financially burdensome health expenses. Three are dollar measures and three are ratio measures.

The three dollar measures are (1) total charges for health care (irrespective of who pays the bill or whether or not the bill is paid), (2) out-of-pocket expenses for health care services (family-paid premiums for health insurance are not included), and (3) total out-of-pocket expenses for health (the previous measure plus out-of-pocket premiums).

The three ratio measures use the three dollar measures to construct measures involving a ratio of expenses to total family income. This gives (1) the ratio of total charges for health care to family income, (2) the ratio of out-of-pocket expenses for health care services to family income, and (3) the ratio of total out-of-pocket expenses for health to family income.

Given these measures, the question still remains as to what level of expense, or ratio of expense to income, constitutes a financially burdensome expense. The usual practice in the literature has been to use several different thresholds arbitrarily selected from the upper part of the particular expense distribution under examination, and this practice is followed in this report.

The overall finding for the six measures was that different results were found for the different measures even when the same threshold was used. For example, when looking at dollar measures, the following was found for a threshold of $3,000 or more in 1980. For total health care charges, 18 percent of all U.S. families had dollar charges this high. For out-of-pocket expenses for health care services (excluding family-paid premiums), 2 percent of all U.S. families had dollar

expenses this high. And, finally, for total out-of-pocket expenses for health (including premiums), 3 percent of all U.S. families had dollar expenses this high. As this example illustrates, the percent of families found at any given dollar threshold was generally highest when health costs were measured by total charges, lowest when health costs were measured by out-of-pocket expenses for health care services, and in between when health costs were measured by total out-of-pocket expenses.

Different results for the same threshold were also found when ratio measures were compared. For example, the following was found for a financial burden threshold of health expenses of 10 percent or more of income in 1980. For total health care charges as a percent of family income, 29 percent of all U.S. families had bills of 10 percent or more of their income. For out-of-pocket expenses for health care services as a percent of family income, 6 percent of all U.S. families had expenses of 10 percent or more. And, finally, for total out-of-pocket expenses for health (including insurance premiums) as a percent of family income, 12 percent of all U.S. families had expenses of 10 percent or more. Also, the overall pattern between the different ratio measures was similar to the overall pattern found for the dollar measures. That is, the percent of families reaching any given ratio threshold was highest for total charges for health care as a percent of income, lowest for out-of-pocket expenses for health care services as a percent of income, and in between for total out-of-pocket expenses for health as a percent of income.

When total out-of-pocket expenses for health as a percent of family income was used as a measure of financial burden in 1980, as indicated above, twice as many U.S. families were found to have spent 10 percent or more of their income on health than when the more common measure of out-of-pocket expenses for health care services as a percent of income was used. (As previously pointed out, the difference between these two measures is that the first measure includes premiums for health insurance and public health care coverage programs paid by the family, whereas the second does not.) The inclusion of family-paid premiums in out-of-pocket health expenses had the consistent effect of approximately doubling the proportion of U.S. families at a given expense-to-income threshold in 1980. Thus, the inclusion of family-paid premiums is an important factor in assessing the out-of-pocket health expenses and the financial burden of health that U.S. families bear.

The above patterns associated with the different types of measures were also generally found when each of the three socioeconomic family populations was examined separately. However, the three socioeconomic populations also differed from each other in the proportion of families found at different financial burden thresholds. For dollar measures, the general finding was that a larger proportion of older families than of younger families was found at each dollar threshold, but that the two younger family categories did not differ from each other in the proportion of their populations found at the different threshold levels. For example, 4 percent of older families had out-of-pocket expenses for health of $3,000 or more while only 2 percent of families in each of the younger family categories had expenses this high. For ratio measures, both older families and younger, lower-income families had a higher proportion of their populations at a given ratio threshold level than younger, better-off families. For example, 27 percent of older families and 20 percent of younger, lower-income families had total out-of-pocket expenses for health of 10 percent or more of income compared with only 4 percent of younger, better-off families. In short, patterns of relative financial burden differed with family socioeconomic status, family age, and type of measure used. For dollar measures, only the age of family members was associated with differences in the distribution of financial burden in the U.S. population. For ratio measures, both the age of family members and family income were associated with such differences.

Overall Findings: The Regression Analysis

Of the six measures discussed above, the ratio of total out-of-pocket expenses for health to family income was chosen as the preferred measure for investigating the causes of financially burdensome health expenses among U.S. families. This measure seemed to be the best indicator of the total financial burden related to health that a family had, and it was labeled the financial burden index.

Multiple regression analysis was used to investigate the effect on this index of family social and demographic characteristics, family health and illness characteristics, family income and health insurance characteristics, and family geographic and urbanization characteristics. Regressions were run separately for each of the three socioeconomic family populations with the financial burden index as the dependent variable and approximately 45 of the above family characteristics as independent variables. Because of the large number of independent variables involved, a multiple-step regression process was used, which will not be described here. The statistical significance of the findings was assessed by using a SAS multiple regression program (SURREGR) that takes into account the complex sample design of NMCUES and by using a F test that was analogous to a multiple *t*-test.

When the relationship between all 45 family characteristics and the financial burden index was examined using multiple regression techniques, the overall finding was that in 1980 the major determinants of financially burdensome family health expenses (as measured by the index) were family income and the completeness and type of health insurance coverage (or public health care coverage program) the family had. This was the

finding in each of the three socioeconomic populations. By contrast, health status variables such as the general health status of family members, special family health events (such as death or institutionalization of a family member), and the specific illnesses that family members had were much less important as overall predictors of a family's health-related financial burden as measured by the index. This comparatively negative finding for health status variables was especially surprising given their importance in predicting total family health charges in a separate multiple regression analysis using the same data set, the same three socioeconomic populations, and the same 45 family characteristics as independent variables.

Several health status variables, however, did show up as population-specific predictors of the financial burden index. Among older families, heart and circulatory diseases (as a general category) was a statistically significant predictor; among younger, lower-income families, family work-loss days was a predictor; and, finally, among younger, better-off families, both hospitalization and family illness days in bed were predictors.

In addition to these health-status variables, other population-specific predictors of the financial burden index were race of family head (significant for younger, better-off families); residence in the South (significant for both older families and younger, better-off families); age and education of family head (significant for both younger family populations); and, finally, family head-spouse structure (significant for both older families and younger, lower-income families).

Introduction

Overview

A consistent goal of contemporary U.S. health policy has been adequate health care at an affordable out-of-pocket cost. In pursuing this goal, the most recent emphasis has been on relieving financially burdensome health care expenses among U.S. families (Catastrophic Illness Expenses, 1986). However, there has been a lack of adequate data on which to base national policy initiatives. This report attempts to supply some of the needed data. It is the fifth in a series of reports from the 1980 longitudinal National Medical Care Utilization and Expenditure Survey (NMCUES) that have presented data on health care expenses and the use of health care by U.S. families (Dicker, 1983; Dicker and Sunshine, 1987; Sunshine and Dicker, 1987a; Sunshine and Dicker, 1987b).

Although many aspects of the U.S. health care system can be studied using the individual as the unit of analysis, the family is the more appropriate unit for studying financially burdensome expenses. This is because the financial burden of health care, and decisions concerning the use and financing of health care, usually are family responsibilities rather than individual responsibilities.

Given that the family should be the unit of analysis, the NMCUES is a particularly good data source, as it was originally conceptualized with family analysis in mind. It has a distinct and carefully conceptualized method of longitudinal family construction, a collection of specially developed longitudinal and cross-sectional family variables, and a family public use data tape that will allow other researchers to reexamine and carry further the exploratory research presented here (Public Use Data Tape Documentation: Family Data, 1986).

At the present time there is no agreement among researchers as to what constitute financially burdensome health expenses for a family. Therefore, the first part of this report uses frequency tables to explore multiple conceptualizations and multiple operational definitions of such expenses. On the basis of this examination, an index of financially burdensome family health expenses (or, for short, the financial burden index) was developed. It is defined as the ratio of annualized total out-of-pocket family expenses for health (for all family members *and* including all family-paid premiums) to annualized total family income (from all sources and all family members). The concept of total out-of-pocket expenses in the above definition differs from the concept of out-of-pocket expenses for health care services in that total out-of-pocket expenses include all family-paid premiums for health care coverage.

The second part of this report uses the financial burden index as the dependent variable in a series of multiple regression analyses designed to test hypotheses about the causes of financially burdensome health expenses among families. In testing these hypotheses, a large number of sociodemographic, health-related, socioeconomic, and geographic explanatory variables are examined.

Because the focus of this report is on supplying data for use in making health policy, three populations of particular interest to policymakers are examined. These are (1) older families (defined as all U.S. multiple-person families with a member 65 years of age or over); (2) younger, lower-income families (defined as all U.S. multiple-person families with all members under 65 years of age and with characteristics (income, family size, and so forth) that placed them below 200 percent of the poverty level in 1980); and (3) younger, better-off families (defined as all U.S. multiple-person families with all members under 65 years of age and with characteristics (income, family size, and so forth) that placed them at 200 percent of the poverty level or higher in 1980). Data were tabulated separately on all three family populations and a separate regression analysis was done for each population. The underlying assumption of this approach was that both the causes of and remedies for financially burdensome health care expenses may be population-specific.

This report, therefore, attempts to answer three questions: What is the proper measure of financially burdensome health care expenses, what percent of U.S. families had such expenses in 1980, and, what are the determinants of such expenses among different policy-relevant populations of multiple-person families? It differs from other reports on these subjects in its focus on multiple-person families, its systematic examination of different measures of financial burden, its examination of a larger and more varied collection of variables, its consistent

controlling for age and family socioeconomic status through the use of separate populations for analysis, and its use of multiple regression analysis.

Background

Wyszewianski (1986a) makes a distinction between the size of a health care expense and a family's ability to pay the bill. This distinction underlies the major distinction among studies dealing with financially burdensome or catastrophic health care expenses. Studies that focus on the size of health care expenses usually measure financially burdensome expenses by the total charges (in dollars) for health care services. If these charges are large (above a certain threshold) they are considered to be catastrophic charges. Wyszewianski suggests that this approach reflects the most common concept of catastrophic expenses as involving illnesses with very large expenses. By contrast, studies that focus on a family's ability to pay the bill usually measure financially burdensome expenses by the ratio of out-of-pocket health care expenses to family economic resources (usually income). In these types of studies, families with relatively moderate or low total charges may be found to have burdensome health care expenses if the families are both poor and have inadequate or no health care coverage. What follows is a discussion of selected studies from both orientations to illustrate some of the issues involved.

Most studies on the size of health care expenses focus on the individual rather than on the family as the unit of analysis. However, one study that focuses on the family and is illustrative of this type of research is the Congressional Budget Office study, *Catastrophic Medical Expenses: Patterns in the Non-Elderly, Non-Poor Population* (Koretz, 1982). This study used data collected between 1974 and 1978 from a sample of non-poor families enrolled in the Federal Blue Cross or Blue Shield health benefits plan. The analysis combined one-person and multiple-person families into one analytic aggregate. Although the data were from the years 1974–78, the analysis was done in terms of 1982 constant dollars. The expense variable of interest was total annual health care expenses (charges) as reported to Blue Cross, regardless of who paid for them. Four "catastrophic" thresholds were used to designate families with "high cost" or "catastrophic illnesses." Families at or above a threshold were considered to be families with catastrophic expenses. These thresholds were total annual health expenses of $3,000, $5,000, $10,000, and $20,000. Two findings from the report are as follows:

1. Among the population sampled, families exceeding any of the catastrophic thresholds in a single year are relatively rare, but they account for a sizable proportion of total health expenses. For example, only 5 percent of the examined families exceeded $5,000 in total expenses in a given year, but those families accounted for half of all expenses for all families.
2. Although only a small proportion of families have catastrophic expenses in a given year, a much larger proportion have high expenses at least once over several years. For example, over a 3-year period, 27 percent of the families exceeded the $3,000 catastrophic threshold at least once, while in a single year only 11 percent did so.

The above findings are interesting and important, but they deal with the size of health expenses rather than with a family's ability to pay. Studies of the size of health expenses, therefore, are of limited use for studying the financial burden of health expenses on a family, though they are often very useful for studying the financial burden of health expenses on a society.

Some other limitations of the Koretz study should be noted. Because it combines one-person and multiple-person families into one category, it is not a study of families as this concept is generally understood by the public and many social scientists, for whom the term "family" usually refers to a multiple-person social unit. In the Koretz study, multiple-person families cannot be distinguished from one-person families. As a consequence, the percent of multiple-person families that exceeds a given catastrophic threshold cannot be ascertained. Second, as its title indicates, the Koretz study is not a study of the full family population of the United States. Finally, the study makes a valuable contribution in its use of multi-year data, but it lacks a rigorous definition of a longitudinal family. Because family groupings change over time, there is a need for clear and consistent rules identifying which changes constitute the beginning or end of a family and which are regarded as changes in composition within the same family. (See Appendix II for more information on this point.)

Recently there have been several studies that have focused on a family's ability to pay health care bills. They may all be considered as part of the same research project as they all use the same data base and have been written by researchers from the University of Michigan School of Public Health (Berki, et al., 1985; Berki, 1986; Wyszewianski, 1986b). These studies used data from the longitudinal 1977 National Medical Care Expenditure Survey (NMCES). This survey used a representative sample of the civilian noninstitutionalized population of the United States. The variable of interest in these analyses was the percent of family income consumed by out-of-pocket expenses for health care services. This measure of health expenses did not include out-of-pocket premium payments families made for health care coverage. The analyses from the Michigan group combined one-person families and multiple-person families into one analytic aggregate.

As in the Koretz report, a series of catastrophic thresholds was specified. But unlike the Koretz report, in which dollar thresholds were used, these reports used

percent-of-income thresholds. The thresholds were out-of-pocket health care expenses of 5 percent of income, 10 percent of income, and 20 percent of income. Families at or above these thresholds were considered to be families with "highly burdensome or financially catastrophic health care expenditures" (Berki, et al., 1985, p. 4). Some findings and conclusions from these reports are as follows:

1. Of all U.S. families, 20 percent had out-of-pocket expenses of 5 percent or more of income in 1977, and 4 percent had expenses of 20 percent or more (Berki, et al., 1985).
2. Families with high out-of-pocket expenses relative to income accounted for a disproportionate amount of the total expenses for all families. For example, the 20 percent of families with expenditures of 5 percent or more of income accounted for 42 percent of total expenditures for all families (Berki, et al., 1985).
3. Two distinct types of families have out-of-pocket expenses that are relatively high in relation to income. These are families with a high-cost illness where the costs are so large that insurance does not keep the family from being burdened, and families with relatively small expenses but inadequate or no health care coverage and low income (Berki, et al., 1985).
4. Most families with high out-of-pocket health services expenditures relative to their income resemble the medically indigent and the uninsured (Wyszewianski, 1986b).

Although from the above it seems that some of the families with high out-of-pocket expenses relative to income overlap with some of the families with high total expenses (in dollars), the focus on a family's ability to pay the bill also directs the researcher's attention to poor families with low expenses and to the role of health insurance in reducing out-of-pocket costs.

Some of the limitations of the Michigan group's research are similar to those previously discussed for the Koretz study. The Michigan studies always use a unit of analysis composed of both one-person and multiple-person families, and they do not present a rigorous definition of a longitudinal family. The consequences of these limitations are the same as in the Koretz study.

A third and novel approach to measuring financially burdensome or catastrophic health care expenses was used in the report of the Secretary of the Department of Health and Human Services to the President, *Catastrophic Illness Expenses* (1986). This approach combined the two orientations discussed above. The report defined catastrophic expenses as out-of-pocket expenses that exceeded an absolute dollar threshold (of either $2,200 or $4,400) *plus* a percent-of-income threshold (of either 5 percent or 10 percent). Note that in this measurement approach the dollar thresholds refer to out-of-pocket spending rather than to total charges as in the Koretz study. The approach used in the Secretary's report sets a dollar floor (determined by the dollar threshold) below which out-of-pocket health expenses are not considered catastrophic, although they may still be experienced by the family as financially burdensome.

The above discussion indicates that not only is there no agreement in the literature on what should be the threshold indicating financially burdensome or catastrophic expenses, but there is also no agreement on what measure of expenses the term "catastrophic" should refer to. In the above studies, it has referred to total charges for health care, out-of-pocket expenses for health care services, and a ratio of out-of-pocket expenses for care to family income. Better conceptualization is necessary, and it is carried out in this report in the chapter titled "Measuring Financially Burdensome Health Expenses."

Methods and Limitations

Source of the Data

As previously pointed out, the data presented in this report are from the National Medical Care Utilization and Expenditure Survey (NMCUES). In NMCUES, information was collected on health problems, health care received, expenditures for care, health insurance, and related topics. Data were obtained throughout calendar year 1980 from a sample of the U.S. civilian noninstitutionalized population. NMCUES included both a national household sample encompassing approximately 6,800 families and four state Medicaid samples. All information in this report is based on the national household sample. Detailed technical information on the sample, on estimation procedures, and on measurement procedures can be found in Appendix II.

The NMCUES differs from most surveys of health in that it was a panel (or longitudinal) survey. Altogether, there were either four or five interviews, approximately 3 months apart, that were conducted with each family in the sample from early 1980 to early 1981. In each interview, information on all family members was gathered, usually from a single family respondent.

Definition of the Family

Because NMCUES is a longitudinal survey, covering an entire year, the important concept of *longitudinal family* was developed to deal with the facts that the composition of a family can change over time and that families come into and go out of existence over time. The concept of longitudinal family used in this report is presented in detail in Appendix II. Simplified, it is as follows:

At a point in time, a family is defined as a group of persons sharing a common household and related by blood, marriage, adoption, or a formal foster care

relationship. An unmarried student 17–22 years of age living away from home is also considered a part of a family.

When an initially sampled family had a change in membership during 1980, the prechange and postchange groups were considered the same family if and only if the "majority" of members of the prechange group became members of the postchange group, and the "majority" of members of the postchange group had previously been members of the prechange group. For the purpose of counting a "majority," persons moving into or out of the sample universe—namely, the universe of civilian noninstitutionalized persons residing in the United States—were omitted from the count. For example, persons who were born, or who had died, or who had moved into or out of an institution, or into or out of the military were omitted from the count.

Standardization for Part-Year Families

One problem with analyzing data from a longitudinal survey is that some families enter and leave the survey universe during the time covered by the survey. This has two consequences. First, the number of different families in the longitudinal universe is larger than the number of families that would be found in a cross-sectional survey. Second, a number of families (about 12 percent in NMCUES) did not exist for the full survey year (Dicker and Casady, 1984).

If each family that ever existed during the year were treated equally as one unit, the count of families, which would be equal to the gross total number of distinct families that ever existed during the year, would be larger than the average number of families that existed at a single point in time (the average cross-sectional estimate). Also, if each family that ever existed during the year were treated as one unit, measures of health spending and use of health care by families would not be comparable, as some counts of family spending and use of care would be for a whole year and some for less than a whole year.

Consequently, the following standardizing procedures were chosen. The population of families was time-adjusted so that, for example, a family that had existed for only half a year was counted as only half a unit. Therefore, in this report the total number of families in any category represents the total number of family years for that category. (Alternatively, this can be considered the average daily number of families in that category during the year 1980.) Moreover, the counts for any health behavior event were adjusted to represent annual rates for that event. For example, a family in the survey for half the year with $150 in total charges is represented as a half family year unit with total charges at an annual rate of $300 per year. Because these concepts are awkward to use in writing, families will be generally discussed in the following text as if they represented one unit each, and the expenses will be discussed as if they were actual expenses rather than annual rates. It should be noted, however, that the term, "family," as used in the text, means *family years* and that all health expenses are rates per family year.

The Definition of a Health Expense

Annualized family health expenses are the health expenses used in this report. Depending on the analysis, the expenses can be either total charges for care or out-of-pocket expenses. Ideally, these measures would include expenses for all types of health care. However, the actual analysis is limited by the type of expense data collected in NMCUES, which did not cover all types of health expenses. The data used in this report include expenses for the following types of health services: inpatient hospital care, inpatient physician care, outpatient hospital and emergency room care, ambulatory physician care, dental care, acquisitions of prescription medicines, care from other independent medical providers (such as chiropractors, speech therapists, faith healers, and psychologists), and the acquisition of health care supplies and services (such as eyeglasses, orthopedic items, hearing aids, ambulance services, and diabetic items). In previous reports (Sunshine and Dicker, 1987a; Sunshine and Dicker, 1987b), the expense measure used in this report has been labeled expenditures for "all health care combined." However, this measure does not include expenses for nonprescription medicines, nursing homes, and other types of long-term care institutions.

When relevant in this report, family-paid premiums for private health insurance or public health care coverage programs were also added to the above list of expenses. When premiums are included, the inclusion is noted in the text.

Adjustments to the Sample

As previously pointed out, this report covers only multiple-person families (defined as families with an average family size of 1.5 members or more during the survey year). This is the type of family that the general public and most social scientists mean when they use the concept "family." (See the discussion in Dicker and Casady, 1985). Moreover, as the review of the literature indicates, this social unit has not been treated separately by most previous researchers examining financially burdensome health expenses. Also, to have included one-person families in the analysis would have meant having a separate analysis for that type of social unit. This would have excessively increased both the size of the report and the amount of time needed to complete it. Thus, one-person families (defined as families with an average size of less than 1.5 members during the survey year) were excluded from the analysis in this report.

The NMCUES sample consisted of 4,888 responding multiple-person families. Of these, 43 (or 0.9 percent)

were families with military heads. Because NMCUES was a survey of the noninstitutionalized civilian population, another family member (usually the spouse) was imputed as the head of family in these families. This imputation produced many anomalies in the data (see Public Use Data Tape Documentation: Family Data, 1986, pp. 22–23). Consequently, it was decided to exclude these families from the analysis. This gave a basic sample of 4,845 multiple-person families that were unequivocally representative of the civilian family population of the United States.

Another problem particular to this report resulted from the use of ratio measures composed of health expenses in the numerator and family income in the denominator. Some families reported either a zero income or a very low income (defined as an income less than $1,000 or under 20 percent of the poverty level). As the above type of ratio measure would be undefined for families with a zero income, and as a very low reported income is probably not a good measure of the actual financial resources available to the families reporting such incomes, some adjustment was necessary. Two types of adjustments have been used in the literature. The first imputes a minimum income for such families (Duan, et al., 1982). The second leaves such families out of the analysis (Berki et al., 1985). Each has its positive and negative elements. This report follows Berki and leaves these families out of the analysis when ratio measures are used. Therefore, when such measures are used, 21 multiple-person families with zero or very low incomes (0.4 percent of the total) are excluded, giving a basic sample of 4,824 civilian families. It is believed that this exclusion does not fundamentally distort the analysis presented here.

The Two-Part Model

Another adjustment to the sample is the result of using a two-part model recommended by Duan, et al. (1982) for analyzing the determinants of health spending (or of use of health services). The first part of this model identifies what distinguishes populations with health spending (or use of health services) from populations without health spending (or use of health services). The second part of the model identifies the determinants of the amount of spending or use in populations with spending or use. This model has been used in previous NMCUES reports on family health care spending and use of health services. (See Dicker and Sunshine, 1987; Sunshine and Dicker, 1987a; Sunshine and Dicker, 1987b.)

The appropriate model to use depends on the research question under examination. In the descriptive tables in this report, Tables A through F, the research question is, "What is the proportion of the total U.S. family population (and the proportion of the three socioeconomic subpopulations) that have financially burdensome health expenses when different measures of such expenses are used?" All families are included in this analysis, and the two-part model was not used. In contrast, in the multiple regression analyses, the research question was "What are the causes of financially burdensome health expenses among families with expenses?" Here the two-part model is appropriate. Following Duan, et al. (pp. 20–24), the population for this second research question was limited to families with positive (nonzero) health expenses. This eliminated the need to impute dollar amounts for families with zero expenses in order to avoid having to calculate the logarithm of zero, which is undefined. Of the 4,824 families remaining in the sample after the adjustments discussed previously had been made, 181 families had zero total out-of-pocket expenses. (Total out-of-pocket expenses, as previously stated, includes family paid premiums.) These 181 families (about 3.7 percent of the adjusted sample) were excluded, leaving a total of 4,643 multiple-person, civilian families for the regression part of the study.

The exclusion of families with zero total out-of-pocket health expenses from the regression analyses differentially affected the three socioeconomic populations. Whereas only 1 percent of older families and only 2 percent of younger, better-off families were excluded, 11 percent of younger, lower-income families were excluded. Therefore, the reader is cautioned to interpret the regression results as being statistically valid only for family populations with nonzero total out-of-pocket health expenses. As a result, the question actually investigated using multiple regression was "What are the causes of financially burdensome health expenses among families with nonzero total out-of-pocket health expenses?"

Sampling Error

Because the statistics shown in this report are based on a sample of families rather than on information from all families, they are subject to sampling error. The standard error is a statistic that measures such errors. Standard errors for most estimates in this report have been computed and are presented along with the estimates.

Non-Sampling Error

In addition, estimates presented in this report are subject to nonsampling errors such as biased interviewing and reporting, misrecording of responses, undercoverage, and nonresponse. Extensive efforts were made to minimize these errors in the data collection and data processing for the survey (see Bonham, 1983).

In terms of nonsampling error, it should be noted that data in this report are derived from information furnished by a survey of households—that is, "consumers" of health care. Data reported by providers of care, for example, in surveys of physicians, hospitals, and nursing homes, are generally different from those re-

ported by households (Sunshine, 1982). Anderson and Thorne (1985) specifically compared use of health care and expenditures on health care, as reported by families in NMCUES, with estimates underlying the national health accounts, which are generally provider-based. They reported good agreement on total U.S. use of health care and on out-of-pocket expenditures for health care services after coverage differences—such as the omission of military and institutionalized persons from NMCUES—are taken into account. However, they found an approximate 10-percent difference between the national health accounts and NMCUES in total charges for health care services. It is likely that total charges, as estimated in this report, underestimate the true amount. Appendix II provides further information on this problem. (Also see the discussion in Sunshine and Dicker, 1987b, on the imputation of total charges in the section on nonsampling error.)

Statistical Significance and Hypothesis Testing

All frequency tables in this report show not only an estimate of the percent of families with various high levels of expenses, but also an estimate of the standard error of these percents. Where the text indicates that two estimates differ, the difference has been tested by a multiple *t*-test using the Bonferroni inequality (see Levy and Lemeshow, 1980, p. 296) and found significant at the 0.05 level. Standard errors were computed by the SESUDAAN computer software package (Shah, 1981), which takes into account the effect of the NMCUES complex sample design upon the standard errors of statistics estimated from its data.

This report also uses multiple regression analysis to examine the relationship between the index of financially burdensome family health care expenses and approximately 50 independent variables that characterize families. Even after stepwise regression produced smaller preferred models, the preferred models still had a large number of variables, and an adjustment was made in estimating significance at the .05 level that was similar to the adjustment made with a multiple *t*-test using the Bonferroni inequality. Using this adjustment, significance at the 0.05 level of probability was the equivalent of significance at the 0.0029 or lower level of probability as estimated by the SURREGR program (Holt and Shah, 1982). For more details, see Appendix I.

Measuring Financially Burdensome Health Expenses

In the "Background" section, two basic types of measures of financially burdensome health care expenses were identified. These were measures based on the dollar value of expenses and measures based on ratios of the dollar value of expenses to family (or societal) income. In keeping with this perspective, this section examines six possible measures of financially burdensome health care expenses: three dollar measures and three ratio measures. First, the measures are defined and their significance described. Then data from NMCUES for the three populations of interest will be presented for each measure.

Six Basic Measures: Definitions

Three Dollar Measures

1. *Total charges for health care services*—Total family charges for health care services are defined as the total amount billed to families for health care services, whether these amounts are paid out-of-pocket, paid by health care coverage, or go unpaid. Total charges are the underlying expense that families and societies face. To the extent that they are so high as to be burdensome, they constitute a potential financial (and health) problem for both individual families and for modern societies. Private health insurance and public health care programs are social mechanisms designed to reduce the burden of total charges on the family. These social mechanisms convert infrequent, large, unexpected, burdensome expenses into expenses that are routine, smaller, and more planned for. As a result, health expenses become more manageable. However, the existence of these programs usually does not completely eliminate all family health expenses. There are still the planned expenses associated with premium payments as well as unexpected, unplanned expenses that result from the fact that health care coverage programs rarely cover all health care costs.
2. *Out-of-pocket expenses for health care services*—Out-of-pocket expenses for health care services are defined as amounts paid by families for health care services that are not reimbursed by public health care coverage programs or by private insurance. (Such expenses exclude premiums paid for these public and private health care coverage programs.) Out-of-pocket expenses for health care services are a measure of the expenses for health care that families actually bear. They tell what the family actually had to pay, whether or not it had health care coverage and regardless of what the actual total charges by providers were. In large part, these are unplanned expenses. They are therefore a measure of what the cost problem is if, and when, the family uses health care services. They do not tell what this cost would be if there were no public health care programs or private health insurance. For this value, one must look to total charges, as discussed in number 1 above.
3. *Total out-of-pocket health expenses*—Total out-of-pocket expenses for health are defined as the *total* amount families pay from their own resources for health care. They include both out-of-pocket expenses for health care services (as described in number 2 above) *and* family-paid premiums for private insurance or public health care coverage programs. While it is true that premium expenses are expected, usually voluntary, and so presumably cannot constitute a financial disaster, premium expenses are genuine health expenses and should be included when measuring a family's total health cost burden. Indeed, if premium costs are large, they can be a substantial part of a health expenditure level that, in total, *is* quite burdensome. Moreover, in a very real sense, families face a trade-off between premiums and out-of-pocket expenses for health care services. If families forgo insurance, their premiums are zero, but their out-of-pocket costs for health care services (or risk of these out-of-pocket costs) is high. At the other extreme, families might obtain very extensive health care coverage—for example, through a health maintenance organization (HMO). They would thereby be sure of cutting out-of-pocket costs for care almost to zero, but the premium for their coverage would then be relatively large. Thus, to obtain a realistic measure of the cost burden that families *actually* bear, it is desirable to include both out-of-pocket costs for health care services and family-paid premiums.

Three Ratio Measures

4. *Total charges for health care services as a percent of family (or societal) income*—The financial burden

that a family (or society) experiences at any given level of expenditure depends very heavily on its income. For example, an expense that is a severe burden for a low-income family might be a relatively minor burden for a higher-income family with, say, five times the income. Hence, in order to measure the burden that health expenses represent for these two types of families, it is useful to examine the ratio of these expenses to family income. For families, total charges as a percent of income measure what a family's financial burden would be if it had to pay its total cost for health care services out-of-pocket without the aid of private health insurance or public health care coverage programs. For society, this measure tells what the aggregate financial burden of health care is for the total society when compared with some measure of total societal income (GNP, net national product, and so forth).

5. *Out-of-pocket expenses for health care services as a percent of family income*—As with the previous measure, the argument for this measure is that the actual financial burden that any given expense generates depends heavily on income. This measure, like measure number 2, focuses on the often unexpected out-of-pocket spending for health care services that a family experiences after private health insurance and public health care coverage programs have made their contribution to paying bills for health care services.
6. *Total out-of-pocket expenses as a percent of family income*—This measure is derived from measure number 3, total out-of-pocket expenses for health care. It is the ratio of total out-of-pocket expenses (out-of-pocket expenses for health care services plus family-paid premiums) to family income. It is a measure of the total actual financial burden of health care expenses on a family. It takes into account the tradeoff discussed earlier between planned premium costs and generally unplanned out-of-pocket costs for health care services. In addition, this measure approximates the measure used in the federal income tax to quantify how burdensome a family's health costs are.

Each of these measures is valuable in identifying financially burdensome costs. None is the "right" measure—or the "wrong" measure—in any universal sense. Some, however, are more useful than others for specific purposes. For example, if a study's objective is to model the effects of tax law changes, it should use the ratio of total out-of-pocket expenses for health care to total family income, measure 6, or a variant thereof.

What Level of Expense is Financially Burdensome?

Apart from the question of what measure to use, identifying financially burdensome health expenses, as previously pointed out, involves another basic question: What level of expenses is financially burdensome? To take examples from some of the studies cited above, is the threshold of burden $5,000, or $2,000, or something in between? Is it 5 percent of income, 20 percent of income, or somewhere in between—or perhaps higher or lower? One approach to answering this question (mentioned in Berki, et al., 1985) is to look to the Federal tax law, since it is the closest thing the United States has to a nationwide standard. However, this approach does not provide a clear answer. As of 1987, the Federal income tax threshold for an itemized deduction for medical expenses became 7.5 percent of income, but before that it was 5 percent of income, and only a few years earlier it was 3 percent of income. Therefore, it appears that even for a specific purpose, such as tax law, the expense level that is burdensome is the subject of changing views. In short, there is no consensus, national or otherwise, as to what level of health care expenses is financially burdensome.

In light of this reality, a number of the studies have evolved a useful pragmatic approach. (See, for example, Koretz, 1982 and Berki, et al., 1985.) This approach is based on the view that financially burdensome health expenses are—almost by definition—uncommon; and, therefore, possible thresholds to measure them should be taken from the upper part of the distribution of expenses. The pragmatic approach then selects multiple thresholds, widely spaced from one another, and taken from the upper part of the distribution of expenses. For example, Berki et al. (1985) uses a range of threshold values from 5 percent of income to 20 percent of income. The proportion of the U.S. population of families included by these thresholds ranges from the highest 20 percent of U.S. families in out-of-pocket health care spending to the highest 4 percent of U.S. families. The purpose of a widely spread set of thresholds is to present data and show relationships across a range of values that includes what most (if not all) observers would choose as a cutoff for "burdensome" if they had to choose a single cutoff. This approach offers readers specific data and simultaneously illustrates relationships across a broad range of values of interest.

Six Basic Measures: Data Presentation

In light of the preceding discussion of six possible alternative measures of financially burdensome health care expenses, it seems appropriate to present basic data on the distribution of U.S. families in 1980 by each of the six measures. For each measure, a set of thresholds is used that divides families into dichotomous distributions according to whether or not their health expenses reached the various thresholds. Following the pragmatic approach just described, the thresholds are chosen from the upper part of the distribution of families on each of the six measures of health expenses, and are widely spread. The data for all U.S. multiple-person families

and for the three socioeconomic family categories that are of concern in this report are found in Tables A through F.

Three Dollar Measures

1. *Total charges for health care services*—The 1980 annual rate of total family charges for health care services for multiple-person families is shown in Table A. This table shows that 18 percent of all U.S. multiple-person families had total charges in 1980 of $3,000 or more, 9 percent had total charges of $5,000 or more, 3 percent had total charges of $10,000 or more, and only 1 percent had total charges of $20,000 or more.

   Two patterns emerge from the data on the socioeconomic categories presented in the table. First, the percent of older families with high levels of total charges was twice or more the percent for either category of younger families. For example, 18 percent of older families had total charges of $5,000 or more in 1980 compared with 7 or 8 percent of families in the two categories of younger families. Similarly, 8 percent of older families had total charges of $10,000 or more in 1980 compared with 2 or 3 percent of families in the younger categories.

   The second pattern is that there are no statistically significant differences between the two socioeconomic categories of younger families in the proportion of families at or exceeding the various expense thresholds shown in Table A. This suggests that both categories of younger families have the same proportions of families with high total charges for health care.

2. *Out-of-pocket expenses for health care services*—The 1980 annual rate of out-of-pocket spending for health care services for multiple-person families is shown in Table B. This table shows that 16 percent of all multiple-person U.S. families had out-of-pocket expenses for health care services of $1,000 or more in 1980, 4 percent had out-of-pocket expenses of $2,000 or more, and only about 1 percent had out-of-pocket expenses of $4,000 or more.

   Because the amounts spent out of pocket for health care services are much less than the amounts billed for total charges, the thresholds chosen for Table B tend to be lower than those chosen for Table A. There is, however, one comparable threshold, $3,000 or more. A comparison of Tables A and B for this expense threshold indicates that the proportion of families with an expense at or

Table A

Total charges for health care for multiple-person families, by age and status relative to the poverty level: United States, 1980

Age and status of multiple-person family	Sample size	Number of families in thousands	Total charges			
			$3,000 or more	$5,000 or more	$10,000 or more	$20,000 or more
			Percent of families (standard error)			
All	4,845	58,135	17.5 (0.6)	9.2 (0.5)	3.3 (0.3)	0.9 (0.1)
Older	860	10,809	26.9 (1.7)	17.8 (1.5)	7.7 (1.0)	2.3 (0.6)
Younger:						
Lower income	1,057	13,595	15.5 (1.2)	8.0 (0.9)	2.9 (0.5)	*0.8 (0.3)
Better off	2,928	33,732	15.4 (0.7)	7.0 (0.6)	2.1 (0.3)	*0.5 (0.2)

NOTES: Older families are families with member(s) 65 years of age or over. Younger families are families with no member 65 years of age or over. Lower-income families are families with income below 200 percent of the poverty level. Better-off families are families with income of 200 percent of the poverty level or more.

Table B

Out-of-pocket expenses for health care services for multiple-person families, by age and status relative to the poverty level: United States, 1980

Age and status of multiple-person family	Sample size	Number of families in thousands	Out-of-pocket expenses				
			$1,000 or more	$1,500 or more	$2,000 or more	$3,000 or more	$4,000 or more
			Percent of families (standard error)				
All	4,845	58,135	15.5 (0.6)	8.0 (0.5)	4.2 (0.3)	1.6 (0.2)	0.7 (0.1)
Older	860	10,809	20.7 (1.6)	10.6 (1.1)	5.8 (0.7)	2.4 (0.4)	0.8 (0.3)
Younger:							
Lower income	1,057	13,595	12.0 (1.1)	6.4 (0.7)	3.8 (0.5)	1.8 (0.4)	1.2 (0.4)
Better off	2,928	33,732	15.2 (0.8)	7.8 (0.6)	3.9 (0.4)	1.3 (0.2)	0.5 (0.1)

NOTES: Older families are families with member(s) 65 years of age or over. Younger families are families with no member 65 years of age or over. Lower-income families are families with income below 200 percent of the poverty level. Better-off families are families with income of 200 percent of the poverty level or more.

Table D

Total charges for health care as a percent of annual income for multiple-person families, by age and status relative to the poverty level: United States, 1980

Age and status of multiple-person family	Sample size	Number of families in thousands	Total charges of— 10 percent of income or more	25 percent of income or more	50 percent of income or more	75 percent of income or more	100 percent of income or more
			Percent of families (standard error)				
All	4,824	57,825	29.0 (0.8)	13.0 (0.6)	5.7 (0.4)	3.3 (0.3)	2.5 (0.3)
Older	858	10,766	43.0 (1.9)	24.7 (1.7)	12.7 (1.3)	8.6 (1.2)	5.6 (0.9)
Younger:							
Lower income	1,038	13,327	47.3 (1.7)	25.3 (1.4)	10.8 (1.1)	6.1 (0.9)	5.3 (0.8)
Better off	2,928	33,732	17.3 (0.7)	4.4 (0.5)	1.4 (0.3)	0.5 (0.1)	*0.3 (0.1)

NOTES: Older families are families with member(s) 65 years of age or over. Younger families are families with no member 65 years of age or over. Lower-income families are families with income below 200 percent of the poverty level. Better-off families are families with income of 200 percent of the poverty level or more. Excludes very low income families—those reporting annual income under $1,000 or less than 20 percent of the poverty level.

their income, 6 percent of multiple-person families had total charges that amounted to 50 percent or more of their income, and 3 percent of families had total charges that equaled or exceeded 100 percent of their income.

A comparison among the three socioeconomic categories into which multiple-person families were divided for this report shows that for most of the thresholds presented in Table D, older families and younger, lower-income families had similar distributions. That is, approximately the same proportion of older families and of lower-income younger families reached or exceeded a given threshold of percent of total charges to income. For example, approximately 25 percent of both older families and younger, lower-income families had total charges for health care that amounted to 25 percent of their income or more. By comparison, a much smaller percent of younger, better-off families were found to have reached or exceeded each threshold. For example, only 4 percent of these families had total charges that reached or exceeded the 25-percent-of-income threshold.

A comparison of Table A with Table D suggests that the similarities found in Table D for older families and younger, lower-income families are found because the health and socioeconomic factors that affect the numerators of the ratios in Table D are different from the health and socioeconomic factors that affect the denominators. It appears, in particular, that the relatively high proportion of older families at each percent-of-income threshold in Table D is the result of the higher total charges for health care these families had (Table A). In contrast, the similar, relatively high proportion of younger lower-income families at each percent-of-income threshold in Table D appears to be the result of the lower family incomes of these families, since their actual dollar charges for health care were relatively low (Table A).

Table D indicates that if families had to pay their total charges for health care out of pocket, only a minority of all families would have financially burdensome charges. Looking at the two lowest percent thresholds, from 29 percent to 13 percent of all families could be considered to have financially burdensome total charges. However, these proportions are not miniscule and probably are much larger than desired either by the families or society. Moreover, the burden of total charges would fall most heavily on older families and on younger, lower-income families. For these families, the two lowest thresholds in Table D show that from 47 percent to 25 percent of the families had ratios that indicate total charges that if paid out of pocket would be financially burdensome to the families.

It should be noted, however, that family behavior would change considerably if families had to pay their total charges out of pocket rather than having much of these charges paid by health care coverage plans. Previous research has shown that use of care and total charges would be lower if care had to be completely paid for on an out-of-pocket basis (Newhouse et al., 1981).

5. *Out-of-pocket expenses for health care services as a percent of family income*—The percent of family income accounted for by out-of-pocket expenses for health care services in 1980 is shown in Table E for U.S. multiple-person families. This table shows that 16 percent of U.S. multiple-person families had out-of-pocket expenses for health care services that amounted to 5 percent or more of their income in 1980, 6 percent of multiple-person families had such out-of-pocket expenses that amounted to 10 percent or more of their income, 2 percent of the families had out-of-pocket expenses that amounted to 20 percent or more of their income, and only 1 percent of families had out-of-pocket expenses for health care services totalling 25 percent or more of their income.

A comparison of Tables D and E indicates that the proportion of families at a particular percent-of-income threshold in 1980 was much smaller when

Table E

Out-of-pocket expenses for health care services as a percent of annual income for multiple-person families, by age and status relative to the poverty level: United States, 1980

Age and status of multiple-person family	Sample size	Number of families in thousands	Out-of-pocket expenses of— 5 percent of income or more	10 percent of income or more	15 percent of income or more	20 percent of income or more	25 percent of income or more
			Percent of families (standard error)				
All	4,824	57,825	16.3 (0.7)	6.4 (0.4)	3.4 (0.3)	2.1 (0.2)	1.4 (0.2)
Older	858	10,766	29.9 (1.9)	13.3 (1.4)	6.9 (1.0)	4.3 (0.8)	2.9 (0.6)
Younger:							
Lower income	1,038	13,327	26.5 (1.4)	11.8 (0.9)	7.5 (0.8)	5.1 (0.7)	3.4 (0.6)
Better off	2,928	33,732	7.9 (0.6)	2.1 (0.3)	0.7 (0.2)	*0.3 (0.1)	*0.2 (0.1)

NOTES: Older families are families with member(s) 65 years of age or over. Younger families are families with no member 65 years of age or over. Lower-income families are families with income below 200 percent of the poverty level. Better-off families are families with income of 200 percent of the poverty level or more. Excludes very low income families—those reporting annual income under $1,000 or less than 20 percent of the poverty level.

the numerator for the ratio measure was out-of-pocket expenses for health care services than when the numerator for the measure was total charges for health care. For example, 29 percent of U.S. families had total charges for health care that amounted to 10 percent or more of their income (Table D) compared with 6 percent that had out-of-pocket expenses for health care services that were this high in proportion to income (Table E). The difference amounts to almost a fivefold decrease in the proportion of families that spent 10 percent or more of their income on health care. At higher expense thresholds, the decrease in the proportion of families reaching or exceeding the threshold was even greater. For example, 13 percent of U.S. families had total charges for health care that amounted to 25 percent or more of income (Table D) compared with only 1.4 percent that had out-of-pocket expenses for health care services that were this high (Table E). This is almost a tenfold decrease in the probability of having health care expenses of 25 percent or more of income. These large decreases in the probability of having high expenses for health care can be attributed to the general availability in the United States of private and public health care coverage programs that convert high total charges for health care into moderate or low out-of-pocket expenses for health care services.

This phenomenon is also found for the three socioeconomic family categories that are of interest in this report. In each category, the relative risk of reaching or exceeding a particular percent-of-income threshold was much lower in 1980 for out-of-pocket expenses for health care services than for total charges. The size of the difference, however, varied with the particular threshold and the population category. For example, 43 percent of older families had total charges for health care that amounted to 10 percent or more of income (Table D) compared with 13 percent that had out-of-pocket expenses for health care service that were this high (Table E). This was approximately a threefold decrease. For younger, better-off families, the corresponding statistics were 17 percent (Table D) and 2 percent (Table E). This was an eightfold decrease. For younger, lower-income families the relevant statistics were 47 percent (Table D) and 12 percent (Table E), a fourfold decrease.

As the above statistics indicate, and as was previously found when examining total charges as a percent of income, when out-of-pocket expenses for health care services are analyzed as a percent of income, both older families and younger, lower-income families are much more likely to have had a high financial burden for health care costs than younger, better-off families. Moreover, the proportion of families with high levels of burden appears to be about the same for younger, lower-income families and for older families. This was true even at very high ratios. For example, at the expense-to-income threshold of 20 percent or more, 4 percent of older families and 5 percent of younger, lower-income families spent this proportion of income out of pocket for health care services compared with less than 1 percent of younger, better-off families.

It is not obvious from a comparison between the dollar data in Table B and the ratio data in Table E why the above distributions are the way they are. A higher proportion of older families had out-of-pocket expenses for health care services of $1,000 or more and $1,500 or more than did younger families (Table B). This suggests that the high proportions of older families with high expense-to-income ratios result from the higher dollar levels of out-of-pocket expenses for health care services that they have. In contrast, the high proportions of younger, lower-income families with high expense-to-income ratios appears to result from the lower incomes of these families. This appears to support some of the conclusions of Berki et al. (1985). On the other hand, there were no consistent, statistically significant differences between older and

younger families in the proportions of families with dollar expenses of $3,000 or more and $4,000 or more (Table B). This suggests that a simple explanation involving only lower income for younger, lower-income families and only higher expenses for older families is not a complete explanation for the distribution found in Table E.

6. *Total out-of-pocket health expenses as a percent of family income*—Total out-of-pocket health expenses are the sum of out-of-pocket expenses for health care services and family-paid premiums for private and public health care coverage(s). Statistics on this measure are presented in Table F. This table shows that 29 percent of U.S. multiple-person families had total out-of-pocket expenses for health care that amounted to 5 percent or more of their income, 12 percent of multiple-person families had total out-of-pocket expenses of 10 percent or more of income, 4 percent of families had total out-of-pocket expenses of 20 percent or more of income, and two percent of families had total out-of-pocket expenses of 25 percent or more of income.

A comparison of the distributions in Table F and Table E indicates that the addition of out-of-pocket premium payments for health care coverage approximately doubles the proportion of families found at each of the ratio thresholds. For example, Table F shows 12 percent of families had total out-of-pocket expenses that were 10 percent or more of their income compared with only 6 percent of families in Table E. For expenses of 20 percent or more of income, Table F shows 4 percent of U.S. multiple-person families at or above this threshold compared with 2 percent in Table E. The approximate doubling effect is also found for the three socioeconomic categories of families found in the tables. Clearly, the addition of out-of-pocket premiums to other out-of-pocket health care expenses has a sizable impact on the distribution of families according to thresholds of financial burden. To omit out-of-pocket premiums from out-of-pocket health expenses therefore runs the risk of seriously understating the full financial burden of out-of-pocket health care costs on U.S. families. In absolute numbers, for example, Table E indicates there were 3.7 million U.S. multiple-person families with out-of-pocket health care expenses of 10 percent or more of income in 1980. Table F, in contrast, yields the estimate that there were 6.9 million such families.

When Table F (showing total out-of-pocket expenses as a percent of income) is compared with Table D (showing total charges as a percent of income), the expense-reducing effect of private health insurance and public health care programs is found to be strongly attenuated versus the effect found when Table E (showing out-of-pocket expenses *only* for health care *services*) is compared with Table D. For example, consider the threshold of expenses of 10 percent or more of income. Table D shows that 29 percent of U.S. multiple-person families are at this threshold, whereas Table F shows 12 percent of families at this threshold. This is less than a threefold decrease in the proportion of families reaching the threshold compared with the fivefold decrease found when the comparison was between total charges (Table D) and out-of-pocket expenses for health care services (Table E). At the higher threshold of expenses of 25 percent or more of income, Table D shows 13 percent of families at this threshold compared with 2.2 percent for Table F. This is only a sixfold decrease in the probability of reaching the threshold compared with the nearly tenfold decrease found when the comparison involved only the out-of-pocket expenses for health care services shown in Table E. This attenuation of the effect of health care coverage programs in reducing high out-of-pocket costs is also found when the families in the three socioeconomic categories in the tables are examined separately. The preceding comparison involving Tables D, E, and F suggests that while health care coverage programs do have a major effect in relieving financially burdensome out-of-pocket costs for health care, the effect of these programs

Table F

Total out-of-pocket health expenses as a percent of annual income for multiple-person families, by age and status relative to the poverty level: United States, 1980

Age and status of multiple-person family	Sample size	Number of families in thousands	Total out-of-pocket expenses of— 5 percent of income or more	10 percent of income or more	15 percent of income or more	20 percent of income or more	25 percent of income or more
			Percent of families (standard error)				
All	4,824	57,825	29.3 (0.9)	12.0 (0.5)	5.8 (0.4)	3.6 (0.3)	2.2 (0.2)
Older	858	10,766	54.6 (1.7)	27.2 (1.7)	14.2 (1.5)	8.2 (1.1)	5.1 (0.8)
Younger:							
Lower income	1,038	13,327	40.7 (1.7)	20.3 (1.2)	11.6 (1.1)	7.6 (0.9)	4.6 (0.7)
Better off	2,928	33,732	16.8 (0.8)	3.8 (0.4)	0.9 (0.2)	0.5 (0.1)	*0.4 (0.1)

NOTES: Older families are families with member(s) 65 years of age or over. Younger families are families with no member 65 years of age or over. Lower-income families are families with income below 200 percent of the poverty level. Better-off families are families with income of 200 percent of the poverty level or more. Excludes very low income families—those reporting annual income under $1,000 or less than 20 percent of the poverty level.

appears much weaker after family-paid premiums for private health insurance and public health care coverage programs are added to the other out-of-pocket health care expenses of families.

The relationships in Table F among the three socioeconomic populations show a finding that has been consistent regardless of which of the ratio measures is examined. This finding is that when health expenses are calculated as a percent of income, both older families and younger, lower-income families more often have a high financial burden for health care than younger, better-off families. For example, 27 percent of the older families and 20 percent of the younger, lower-income families had total out-of-pocket health care expenses of 10 percent or more of income in 1980 compared with only 4 percent of the younger, better-off families. At the threshold of expenses of 20 percent or more of income, the corresponding statistics were, respectively, 8 percent, 8 percent, and less than 1 percent. Clearly, Tables D, E, and F indicate that no matter which ratio measure is used, older families and younger, lower-income families are much more at risk for financially burdensome health care expenses than are younger, better-off families. At the threshold of total out-of-pocket expenses of 20 percent of income or greater, this risk is more than eight times greater for the former categories of families than for the latter category.

Although both older families and younger, lower-income families are shown in Table F to have been at relatively high risk of having financially burdensome health care expenses in 1980, the distributions for these two types of families diverge significantly in Table F at the lower percent-of-income thresholds. Older families were more at risk of exceeding the 5-percent threshold or the 10-percent threshold than were younger, lower-income families. At higher thresholds, these differences disappeared and both socioeconomic categories of families were about equally at risk in 1980. (See the distributions in Table F for 15 percent or more of income, 20 percent or more of income, and 25 percent or more of income.)

Table C indicates that when total out-of-pocket *dollar* expenses are examined, at thresholds of $1,000 and $1,500 the younger, better-off families are more at risk of reaching or exceeding the threshold than are the younger, lower-income families. At higher dollar thresholds, the two categories of younger families had about the same proportion of families reaching or exceeding the threshold. In contrast, Table F shows that at every percent-of-income threshold the younger, lower-income families were at far greater risk of financially burdensome total out-of-pocket expenses than were the younger, better-off families. This suggests that among younger families, the lower income of the lower-income families outweighs the higher dollar expenses of the better-off families in creating financially burdensome total out-of-pocket expenses.

The generally higher *dollar* total out-of-pocket expenses of the older families (Table C) suggest that it is their high health expenses that principally create a high financial burden index for them (Table F). However, lower income probably also plays a role here, particularly at the higher thresholds. At $4,000, the highest dollar threshold of total out-of-pocket expenses in Table C, the proportion of older and younger families reaching or exceeding the threshold was about the same—approximately 1 percent. Thus, a relatively high proportion of older families would not be expected at the highest levels of the financial burden index (Table F) unless low income was also a factor. In short, as with the ratio measure for out-of-pocket expenses for health care services (Table E), a simple explanation involving only low income for lower-income families and only high expenses for older families is not the whole explanation of the patterns found for the ratio measure of total out-of-pocket expenses.

Summary—The three ratio measures suggest that financially burdensome expenses can result from high dollar expenses for health care, from low family income, from combinations of these two factors, and possibly from other sources. These findings support the conclusions of Berki et al. (1985) that families with financially burdensome health care costs may differ greatly from one another in terms of their relationship to the health care system. Some may enter the circle of families with financially burdensome expenses by having members with devastating illnesses that result in large expenses. Others may have members with only moderate illness and moderate expenses, but they may lack the financial resources to pay readily for even a modest amount of health care. It is also probable that other aspects of the health care system, such as type of insurance available, affect the distribution of financially burdensome expenses. In the next section, regression analyses are carried out to test whether the above hypotheses remain viable and to search for other possible determinants of financially burdensome health expenses.

The Determinants of Financially Burdensome Health Expenses: A Regression Analysis

The Index of Financially Burdensome Health Care Expenses

If only one of the measures discussed in the preceding chapter were to be chosen as the single best measure of financially burdensome family health care expenses, it would be the ratio of total family out-of-pocket expenses to total family income. Because total out-of-pocket expenses include both family-paid premiums for health insurance coverage and out-of-pocket costs for health care services, this measure, as previously pointed out, takes into account the trade-off families may make between out-of-pocket expenses for health insurance coverage and out-of-pocket expenses for physicians' care, hospital care, and so forth. By including income, it relates these out-of-pocket expenses of all types to a family's ability to pay (as measured by total family income from all sources and from all family members). In the remainder of this report, this ratio is labeled the index of financially burdensome family health care expenses or, for short, the financial burden index.

Because the index is the best available measure of the burden of family health care expenses, an extensive study of the determinants of the level of the index among U.S. multiple-person families in 1980 was undertaken. This study follows.

Methods

Multiple regression analysis was used to identify the statistically significant determinants of the level of the index and to estimate the effect of different family characteristics on this level. Appendix I presents a technical description of the analytic procedures followed. For the readers' convenience, a summary of this material follows.

Multiple regression analysis is a statistical technique for estimating the effect on a single dependent variable of each of a set of independent (or causal) variables. The effect of each independent variable is estimated while controlling for the effect of all of the other independent variables in the set. Multiple regression analysis readily incorporates a large number of independent variables, including both continuous and categorical variables, but requires assumptions to be made about the functional form of the relationship among the variables. When multiple regression analysis using many independent variables shows a statistically significant association between the dependent variable and a particular independent variable, the analyst may assume that a relationship exists between the two variables and that it possibly is a causal one, at least in the population sampled. One reason for this is that the analysis controls for the effects of all the other independent variables in the variable set. However, misleading results can still occur, particularly if causally important variables are omitted from the analysis.

The Model

The set of independent variables assumed to cause changes in the dependent variable and the functional form of this hypothesized relationship are usually referred to as the model. Note that a regression analysis examines how a particular set of independent variables organized into a particular model affect the value of the dependent variable for a particular population. That is, the analysis is both model-specific and population-specific, although inferences are often made to a broader population and to other models.

Variables used. The model used in this report to analyze the relationship between family characteristics and differing levels of the index of financially burdensome family health care expenses was derived from a general conceptualization of the health care system. This conceptualization suggests that generalized health status, specific health conditions (illnesses), and special health events (births, deaths, hospitalizations, and so forth) interact with family demographic factors to produce a family potential for the use of health care services. The actual use of care and the final level of out-of-pocket expenses results from a further interaction of the above health factors with social factors such as sociocultural use patterns, family economic status, prices of health care, general economic conditions, and family health care coverage. Therefore, variables were selected for the model that were representative of the above types of health and social factors. It was assumed that a properly selected set of such variables would include some variables that would affect levels of the index of financially burdensome family health expenses.

Besides this general conceptualization of the health care system, the review of the literature reported above and the findings in the section on measurement suggested two hypotheses to be tested. First, Berki et al. (1985) and Wyszewianski (1986) suggested that the level of family income should be an important determinant of financially burdensome health expenses. Second, Berki suggested that high cost illnesses should also be a determinant. To test these hypotheses, the model included variables for family income and types of illnesses that family members could have.

Testing hypotheses depends on controlling for variables that under alternative hypotheses could be the cause of the outcome actually found. A number of the independent variables in the model are of this type. Finally, the extensive literature on the importance of health insurance in affecting both total expenditures and out-of-pocket payments suggests that health insurance variables be included in the model.

The independent variables selected could easily be arranged in categories of an Andersen-Newman model as presented in Buczko (1986). An Andersen-Newman model calls for health status variables (such as perceived health status), enabling variables (such as income), and predisposing variables (such as age). The model used here includes these types of variables, and one of the strengths of the NMCUES is that it allowed for the inclusion of all these types of variables in the model. The specifics of making the model operational are found in Appendix Table I. This table gives the actual operational form of the dependent variable and of each of the 47 independent variables used in the regression analyses reported here.

It should be noted that many of the variables in Table I are imperfect indicators of the underlying concepts that they represent. As a consequence, an actual variable can fit into the underlying conceptual scheme in more than one way. For example, the variable D9, which identifies families with a black head of family, may fit into the scheme in at least three ways. For one, it may be a demographic factor affecting health status. (Black persons, for example, have particularly high rates of hypertension.) Second, if racially based discrimination exists, D9 would denote a smaller supply of care available. Third, it may mark sociocultural differences in habits and preferences in the use of health care. Note that D9 does *not* represent overall economic difference associated with the different races, for such differences are controlled for by the use of income as an independent variable in the regressions.

The functional form of the relationship hypothesized to exist between the dependent variable, the financial burden index, and the independent variables is multiplicative. That is, it was assumed that a specified change in an independent variable will multiply the index by a constant amount. For example, having a family member with heart disease might multiply a family's index by 1.3—that is, increase it by 30 percent—as compared with what the index would be if no member had such an illness. (The multiplicative model was chosen in preference to an additive model for reasons detailed in Appendix I, which also describes how an additive model would work.) This hypothesized form of the functional relationship between the dependent variable and the independent variables calls for the dependent variable and several of the continuous independent variables to be used in logarithmic form. (Again, Appendix I explains why this is so.) Finally, the model does not take into account interaction effects between variables. In order to take such effects into account, special variables to measure interactions would have to be included in the model.

Procedures Followed

Regression analysis was carried out separately for the three socioeconomic multiple-person family populations focused upon in this study. As previously stated, these are older families (families with a member 65 years of age of over); younger, lower-income families (families with no member 65 years of age or over and with family income below 200 percent of the poverty level); and younger, better-off families (families with no member 65 years of age or over and with family income equal to or greater than 200 percent of the poverty level).

There were several steps in the analysis, as detailed in Appendix I. In brief, the steps were as follows. First, a small number of the initial 47 independent variables were excluded from each regression as not suitable. For example, a variable primarily used to distinguish between families with all members 65 years of age or over and families with only some members 65 years of age or over was omitted from the regressions for the two younger family populations. The initial exclusion left 43 to 45 independent variables in the regressions, with the number depending on the family population involved.

Next, stepwise regression was carried out using PC SAS (SAS Institute, 1985). A major reason for using stepwise regression was to eliminate possible multicollinearity (strong correlation) among variables, as several variables were sometimes used to operationalize a single basic concept. For example, four different sets of variables were used separately to operationalize the concept of family general health status. These were (1) total family bed days due to illness, (2) total family work loss days due to illness, (3) a family-level scale of reported health status, and (4) a family-level scale of limitations in main activity. The result of the stepwise regressions was a much smaller preferred regression model for each of the three socioeconomic family populations. These preferred models contained 17 to 24 independent variables.

However, PC SAS does not properly estimate variances of regression coefficients for samples with a complex survey design, such as that found in the National Medical Care Utilization and Expenditure Survey (NMCUES). Therefore, the three preferred models were rerun as (ordinary, nonstepwise) regressions using SURREGR (Holt and Shah, 1982). SURREGR is a regression program that appropriately estimates variances in a sample with a complex design, but it cannot carry out stepwise regression analysis.

The results of the SURREGR regressions were used to identify which independent variables were statistically significant. These results are shown in detail in Appendix Tables II, III, and IV. Text Tables G, H, and J (one for each of the three family populations) show the statistically significant independent variables in each preferred model and the estimated effect of each significant variable on the financial burden index. Only about one-third to one-half of the independent variables in the preferred models were found to be statistically significant.

A SURREGR multiple regression on the full 43- to 45-variable models was carried out for each family population in order to check that the PC SAS stepwise regressions did not omit a statistically significant variable because of their deficiencies in variance estimation. Results are shown in Appendix Tables V, VI, and VII. No omissions were found by this procedure.

Findings

This section presents findings on the determinants of the financial burden index for multiple-person families in the civilian noninstitutionalized population of the United States during the year 1980. It discusses each of the three family populations in turn.

Older Families

Statistically significant results from the regression analysis for older multiple-person families are shown in Table G, with more details of the regression analysis found in Appendix Table II. The multiple correlation coefficient (R^2) for the regression equation for this older population was 0.53, which means the independent variables shown in Table II explained 53 percent of the variance in the dependent variable (which was the natural logarithm of the financial burden index). Seven of the 22 variables in the preferred model were found to be statistically significant determinants of the financial burden index for older, multiple-person families.

A particularly strong inverse association was found between the financial burden index and family income among older families. The F statistic, which measures the statistical significance of this association, is far larger for family income than for any other independent variable. Moreover, the numerical value of the regression coefficient for the family income variable—minus 0.85—is such that a large income difference has a large effect on the level of the financial burden index. The coefficient means that each 1-percent increase in family income produced in 1980 approximately a 0.85 percent decrease in the level of the index. This is equivalent to saying that if one of two otherwise similar families had twice the income of the other in 1980, the financial burden index of the higher-income family would have been about 45 percent less than that of the lower-income family. Another implication of the regression coefficient merits noting. A coefficient of minus 0.85 means that in 1980 for each 1-percent increase in family income, *total family out-of-pocket health expenses increased* by approximately 0.15 percent—that is, much less than proportionately.

Older families with a head but no spouse present during 1980 had a financial burden index about 32 percent lower than otherwise comparable families with different head-spouse structures. (These latter families were mostly head-and-spouse families, but they also included the 4 percent of older families that had an unstable head-spouse structure during the year.) As with all regression coefficients, this is the estimated effect of this variable *after* controlling for the influence of all other variables in the regression. In particular, since family size was included in the regression, smaller family size is not the explanation for the lower index found for head-only families.

Older families with a member having heart or circulatory disease in 1980 had a significantly higher financial burden index level than similar families that did not have any members with these diseases. The index for the former families averaged 24 percent greater, and it should again be noted that this is the difference that was found after controlling for the effects of other factors included in the regression—in particular, after controlling for general health status and hospitalization. In contrast to this statistically significant finding for a category of illness, no statistically significant effects were found for hospitalization or for variables measuring general health status.

Very low values of the financial burden index were found for older families in 1980 with three types of health care coverage: Medicaid coverage only, Medicare and other public coverage, and an unknown source of coverage. The families with Medicaid-only coverage typically had a financial burden index level about 67 percent lower than that of otherwise comparable families not in a coverage category explicitly listed in the regression. (Most of these "not listed" families had Medicare and private coverage.)

Most of the families in the "Medicare and other public" coverage category probably had both Medicare and Medicaid coverage, since about one in six elderly persons in the United States has Medicaid coverage in addition to Medicare. The financial burden index level for families in the "Medicare and other public" coverage

Table G

Significant regression findings for the financial burden index for older multiple-person families

Significant factor	Effect (all other factors assumed constant)
Family structure	The regression coefficient for families with a head but no spouse was − 0.384. This implies a multiplication by 0.68. Thus, families with a head but no spouse had a financial burden index approximately 32 percent lower than other families. (These others were predominantly head-and-spouse families.)
Type of illness: Heart and circulatory diseases	The regression coefficient for families with member(s) having heart or circulatory disease was 0.218. This implies a multiplication by 1.24. Thus, these families had a financial burden index approximately 24 percent higher than families with no members having these diseases.
Family income	The regression coefficient for family income was − 0.853. This means that each 1.0-percent increase in family income was associated with approximately a 0.85-percent decrease in the financial burden index.
Type of insurance: Medicaid	The regression coefficient for families with health care coverage exclusively from Medicaid was − 1.119. This implies a multiplication by 0.33. Thus, families with health care coverage exclusively from Medicaid had a financial burden index approximately 67 percent lower than families not in a coverage source category explicitly listed in the regression. (Overwhelmingly, the unlisted families had both Medicare and private insurance.)
Medicare and other public coverage	The regression coefficient for families with health care coverage from Medicare plus other public sources was − 2.162. This implies a multiplication by 0.12. Thus, families with health care coverage from Medicare plus other public sources had a financial burden index approximately 88 percent lower than families not in a coverage source category explicitly listed in the regression. (Overwhelmingly, the unlisted families had both Medicare and private insurance.)
Unknown coverage source	The regression coefficient for families with health care coverage from unknown coverage source(s) was − 1.151. This implies a multiplication by 0.32. Thus, families with health care coverage from unknown coverage source(s) had a financial burden index approximately 68 percent lower than families not in a coverage source category explicitly listed in the regression. (Families with unknown coverage source were predominantly families all of whose members had part-year or no coverage. The unlisted families predominantly had full-year coverage of all members, with coverage from both Medicare and private insurance.)
Region	The regression coefficient for families residing in the South was 0.288. This implies a multiplication by 1.33. Thus, families residing in the South had a financial burden index approximately 33 percent higher than families living elsewhere in the U.S.

NOTES: For further details of the regression, see Appendix Table II. The probability for the 0.05 level of significance for the preferred model for older families using the multiple F-test discussed in Appendix I is 0.0023.
For an explanation of the above interpretations of the regression coefficients, see Appendix I.

category was about 88 percent lower than the level for otherwise comparable older families.

Because of coding conventions established earlier in the processing of NMCUES data, families reported as having an unknown coverage source were generally families in which no members had full-year health care coverage. The 3 percent of older families with an unknown coverage source had an index about 68 percent less than that for otherwise similar older families not in a listed coverage category.

Finally, the regression shows older families residing in the South in 1980 with a higher index level than otherwise comparable families living elsewhere in the United States. The index for families in the South was about one-third higher.

Younger, Lower-Income Families

Statistically significant results from the regression analysis for younger, lower-income multiple-person families are shown in Table H, with more details of the regression analysis found in Appendix Table III. For this population of families, the independent variables shown in Table III explained 27 percent of the variance in the dependent variable (which was the natural logarithm of the financial burden index). In total, 9 of the 17 variables in the preferred model were found to be statistically significant determinants in 1980 of the level of the financial burden index for younger, lower-income families.

Two demographic variables were found to be significant. First, families with a head but no spouse present during 1980 had lower index levels than families with other head-spouse structures. (Most of these "other" families had a head and spouse, but they also included the 4 percent of younger, lower-income families with an unstable head-spouse structure.) The head-only families had a financial burden index approximately 39 percent lower than that for otherwise comparable families with a different head-spouse structure.

Table H

Significant regression findings for the financial burden index for younger, lower-income multiple-person families

Significant factor	Effect (all other factors assumed constant)
Family structure	The regression coefficient for families with a head but no spouse was − 0.501. This implies a multiplication by 0.61. Thus, families with a head but no spouse had a financial burden index approximately 39 percent lower than other families. (These others were predominantly head-and-spouse families.)
Age	The regression coefficient for age of head was 0.025. This means that each additional year of age of the family head was associated with an increase of approximately 2.5 percent in the financial burden index.
Education	The regression coefficient for education of head was 0.049. This means that each additional year of education of the family head was associated with an increase of approximately 5 percent in the financial burden index.
General health status: family work-loss days due to illness	The regression coefficient for family work-loss days due to illness was 0.117. This means that each 1-percent increase in the quantity (family work-loss days + 1) was associated with an increase of approximately 0.12 percent in the financial burden index.
Family income	The regression coefficient for family income was − 0.556. This means that each 1-percent increase in family income was associated with approximately a 0.56 percent decrease in the financial burden index.
Type of insurance: Medicaid	The regression coefficient for families with health care coverage exclusively from Medicaid was − 1.674. This implies a multiplication by 0.19. Thus, families with health care coverage exclusively from Medicaid had a financial burden index approximately 81 percent lower than families not in a coverage source category explicitly listed in the regression. (Overwhelmingly, the unlisted families were covered by private insurance only.)
Other public and private coverage	The regression coefficient for families with health care coverage from private insurance plus public source(s) other than Medicare alone was − 0.394. This implies a multiplication by 0.67. Thus, families with health care coverage from private insurance plus public source(s) other than Medicare alone had a financial burden index approximately 33 percent lower than families not in a coverage source category explicitly listed in the regression. (Overwhelmingly, the unlisted families were covered by private insurance only.)
Other public coverage	The regression coefficient for families with health care coverage solely from public sources other than (1) Medicare with or without other public programs or (2) Medicaid alone was − 1.972. This implies a multiplication by 0.14. Thus, families with health care coverage from these "other public" sources had a financial burden index approximately 86 percent lower than families not in a coverage source category explicitly listed in the regression. (Overwhelmingly, the unlisted families were covered by private insurance only.)
Unknown coverage source	The regression coefficient for families with an unknown health care coverage source (which were predominantly families all of whose members had part-year or no coverage) was − 0.564. This implies a multiplication by 0.57. Thus, families with an unknown health care coverage source had a financial burden index approximately 43 percent lower than families not in a coverage source category explicitly listed in the regression. (Overwhelmingly, the unlisted families were covered by private insurance only, with at least some members having full-year coverage.)

NOTES: For further details of the regression, see Appendix Table III. The probability for the 0.05 level of significance for the preferred model for younger, lower-income families using the multiple F-test discussed in Appendix I is 0.0029.
For an explanation of the above interpretations of the regression coefficients, see Appendix I.

Second, the financial burden index increased in 1980 with increasing age of the family head. Each additional year of age of the family head increased the index by approximately 2.5 percent. While this may sound like a small effect, it becomes large with substantial age differences. For example, in 1980 a family with a head 20 years older than the head of an otherwise similar family would have had a financial burden index about 65 percent higher than that of the family with the younger head.

Among younger, lower income families in 1980, more education was associated with a higher financial burden index. Each additional year of education completed by the head of the family increased the index by about 5 percent. Thus, a family with a head who completed high school would have had a financial burden

index some 20 percent higher than that of an otherwise similar family with a head who dropped out of school after completing only eighth grade.

General family health status had an effect in 1980 on the financial burden index of younger, lower-income families. Family work-loss days due to illness was the general health status variable that was found to be statistically significant. However, if this variable had been omitted from the regression, it is quite possible that some other variable measuring general family health status would have been significant in its place. The estimated effect of work-loss days was such that each 1-percent increase in the quantity (family work-loss days due to illness plus 1) increased the level of the financial burden index by approximately 0.12 percent.

In contrast to the significant effect on the financial burden index that was found for general health status (as measured by work-loss days), no significant effect on the index was found for hospitalization in 1980 or for the presence in 1980 of any of the disease categories used as variables in the regression. (These disease categories included cancer and heart or circulatory disease.)

Even within the younger, lower-income family population (all of which, by definition, had a 1980 income below 200 percent of the poverty level), higher-income families had a lower financial burden index in 1980 than lower-income families. Unlike the situation with older families, however, for younger, lower-income families the statistical significance of this finding, as measured by the F test, was not stronger than that for all other statistically significant variables. The regression coefficient of minus 0.56 for the family income variable means that each 1-percent increase in family income produced an approximate 0.56-percent decline in the index. This is equivalent to saying that if one of two otherwise similar younger, lower-income families had twice the income of the other in 1980, the financial burden index of the higher-income family would have been about 32 percent lower than that of the poorer family. A regression coefficient of minus 0.56 also implies that each 1-percent increase in family income in 1980 was associated with an increase in total family out-of-pocket health expenses of approximately 0.44 percent, which was again less than proportionate.

For younger, lower-income families in 1980 the most prominent features of the regression findings were the number of statistically significant health insurance variables and their very large effect on the level of the financial burden index. No fewer than four insurance coverage variables were significant. These insurance variables were coverage by medicaid only, coverage by public programs other than medicare or medicaid, coverage by both private insurance and public programs (other than medicare), and, finally, coverage by a "source unknown."

For the one in eight younger, lower-income families with coverage from medicaid only in 1980, the index was about 81 percent lower than for families not in a listed coverage category. (These families not listed overwhelmingly had coverage only from private insurance.) In other words, the financial burden index for medicaid-only families was only about one-fifth of that for similar families with (predominantly) private insurance only.

Less than 1 percent of younger, lower-income families had coverage solely from public programs other than medicare or medicaid in 1980, but for the few families that had such coverage the index was reduced by about 86 percent. That is, their financial burden index was only about one-seventh of that for comparable families differing only in source of health care coverage.

About one in four younger, lower-income families had coverage from a combination of private insurance and public health care coverage programs (other than medicare) in 1980. For these families, the index was approximately one-third lower than for families not in a listed health coverage category.

Finally, the 28 percent of younger, lower-income families with "source unknown" coverage had an index about 43 percent lower than that for families not in a listed health care coverage category. (As noted, the "source unknown" families in this report generally were families with partial or no health care coverage.)

Younger, Better-Off Families

Statistically significant results from the regression analysis for younger, better-off multiple-person families are shown in Table J, with more details of the regression analysis found in Appendix Table IV. The R^2 for the regression equation for this population was 0.23, which means the independent variables shown in Table IV explained 23 percent of the variance in the dependent variable (which was the natural logarithm of the financial burden index). In total, 10 of the 24 variables in the preferred model were found to be statistically significant determinants in 1980 of the level of the financial burden index for younger, better-off families.

As with older families, a particularly strong statistical association was found between the financial burden index and family income. The F statistic, which measures the statistical significance of this association, was far larger for family income than for any other independent variable in the regression. Again, the numerical value of the regression coefficient—minus 0.87—is such that a large income difference had a large effect on the financial burden index in 1980. Each 1-percent increase in family income in 1980 produced an approximate 0.87-percent decrease in the index. This means that if one of two otherwise similar younger, better-off families had twice the income of the other in 1980, its financial burden index would have been about 45 percent less than that of the lower-income family. A regression coefficient of minus 0.87 also implies that each 1-percent increase in family income in 1980 was associated with an increase in total family out-of-pocket health expenses of approximately 0.13 percent, a much less than proportionate increase.

Three sociodemographic variables were found to be statistically significant for younger, better-off families in 1980. First, increased age of the family head was

Table J

Significant regression findings for the financial burden index for younger, better-off multiple-person families

Significant factor	Effect (all other factors assumed constant)
Age	The regression coefficient for age of head of family was 0.015. This means that each additional year of age of the family head was associated with an increase of approximately 1.5 percent in the financial burden index.
Race	The regression coefficient for families with a black head of family was − 0.335. This implies a multiplication by 0.72. Thus, families with a black head of family had a financial burden index approximately 28 percent lower than that of families with heads of other races.
Education	The regression coefficient for education of family head was 0.034. This means that each additional year of education of the family head was associated with an increase of approximately 3.5 percent in the financial burden index.
Special health event: hospitalization	The regression coefficient for families with one or more hospitalizations (variable H13) was 0.256 and the regression coefficient for number of discharges (variable H14) was 0.069. Together, these imply a multiplication of the financial burden index by 1.38 for families with one discharge. Thus, families with one discharge had a financial burden index about 38 percent higher than families with no hospitalization.
General health status: family illness days in bed	The regression coefficient for family illness days in bed was 0.069. This means that each 1-percent increase in the quantity (family illness days in bed + 1) was associated with an increase of approximately 0.07 percent in the financial burden index.
Family income	The regression coefficient for family income was − 0.868. This means that each 1-percent increase in family income was associated with approximately a 0.87-percent decrease in the financial burden index.
Completeness of health care coverage	The regression coefficient for families with no members having any health care coverage was − 0.529. This implies a multiplication by 0.59. Thus, families with no member having any health care coverage had a financial burden index approximately 41 percent lower than families not in a coverage completeness category explicitly listed in the regression. (Families in the unlisted categories predominantly had all members with full-year coverage.)
Type of insurance:	
Other public and private coverage	The regression coefficient for families with health care coverage from private insurance plus public source(s) other than Medicare alone was − 0.228. This implies a multiplication by 0.80. Thus, families with these other public and private coverage mixes had a financial burden index approximately 20 percent lower than families not in a coverage source category explicitly listed in the regression. (Overwhelmingly, the unlisted families were covered by private insurance only.)
Other public coverage	The regression coefficient for families with health care coverage solely from public sources other than (1) Medicare with or without other public programs or (2) Medicaid alone was − 0.893. This implies a multiplication by 0.41. Thus, families with health care coverage from these "other public" sources had a financial burden index approximately 59 percent lower than families not in a coverage source category explicitly listed in the regression. (Overwhelmingly, the unlisted families were covered by private insurance only.)
Region	The regression coefficient for families residing in the South was 0.273. This implies a multiplication by 1.31. Thus, families residing in the South had a financial burden index approximately 31 percent higher than families residing elsewhere in the U.S.

NOTES: For further details of the regression, see Appendix Table IV. The probability of the 0.05 level of significance for the preferred model for younger, better-off families using the multiple F-test discussed in Appendix I is 0.0021.
For an explanation of the above interpretations of the regression coefficients, see Appendix I. Variable numbers such as H14 refer to variables in Appendix Table I.

associated with a higher financial burden index level. Each 1-year increase in age produced an approximate 1.5-percent increase in the index. Second, families with a black head of family had a relatively low financial burden index. It was about 28 percent lower than that for comparable families with a head of family of another race. Finally, education of the family head was associated with a higher index level, with each additional year of education of the head increasing the index by about 3.5 percent.

Hospitalization in 1980 had a significant effect on the financial burden index for younger, better-off families. A hospitalization increased the index by about 38 percent relative to what it would have been for families with no hospitalization. One generalized health status variable, family illness days in bed in 1980, also had a statistically significant effect on the financial burden index for younger, better-off families. The estimated effect of bed days in 1980 was that each 1-percent increase in the quantity (annual family bed days plus 1)

increased the financial burden index by approximately 0.07 percent. In contrast, none of the variables involving specific categories of illness was statistically significant.

Three health insurance variables were found to be statistically significant. These insurance variables were "no health care coverage," "coverage by private insurance and public sources (other than medicare)," and, finally, "coverage only by public sources other than medicare or medicaid." For the 2 percent of younger, better-off families in 1980 that had no health care coverage, the financial burden index was about 41 percent lower than for otherwise comparable families (most of whom had full year health care coverage for all family members). For the one in seven younger, better-off families with health care coverage from a combination of private insurance and public programs (other than medicare) in 1980, the financial burden index was about 20 percent lower than for otherwise comparable families not in a coverage source category explicitly listed in the regression. (These "not listed" families overwhelmingly had coverage solely from private insurance.) Finally, less than 1 percent of younger, better-off families had coverage only from public programs other than medicare or medicaid, but the financial burden index of these families was about 59 percent lower than that of otherwise comparable families.

One regional effect was found. Younger, better-off families residing in the South in 1980 were found to have values on the index about 31 percent higher than for similar families living elsewhere in the United States.

Discussion

The financial burden index, it should be recalled, is the ratio of total family out-of-pocket health expenses to total family income. Total family out-of-pocket health expenses include both out-of-pocket expenses for health care *and* family-paid premiums for health care coverage. Separate regressions were carried out for three family populations that, combined, represented the whole multiple-person family population of the United States in 1980. These three populations were older multiple-person families (those with a member 65 years of age or over); younger, lower-income multiple-person families (those with all members under 65 years of age and with income below 200 percent of the poverty level); and younger, better-off multiple-person families (those with all members under 65 years of age and with income of 200 percent of the poverty level or more).

Explanatory Power

R^2, the multiple correlation coefficient squared, is a measure of the overall explanatory power of the entire regression. R^2 is equal to the proportion of the variance in the dependent variable that is explained by all the independent variables in combination. In order to provide a better understanding of the relative explanatory power of the regression equations reported in Tables II, III, and IV, R^2 for these equations is compared here with the R^2 reported in other, similar regression studies. As shown in the first column of Table K, R^2s in the regressions for the financial burden index ranged from 0.23 to 0.53, depending on the socioeconomic family category involved.

This is a relatively high R^2 compared with that reported in most studies. For example, three recent papers that use NMCUES data or data from the similar 1977 National Medical Care Expenditure Survey (NMCES) in regression equations similar to those presented here report an R^2 of 0.04 to 0.27 (Farley, 1986), 0.31 (Taube, Kessler, and Burns, 1986), and 0.18 to 0.20 (Buczko, 1987). These studies, however, differ from this report in that they deal with individuals, not families, with physician visits, not total health care, and—except for the Buczko study—with number of visits, not spending.

On the other hand, an analysis exactly paralleling that described in this report, but using total charges for health care instead of the financial burden index as the dependent variable, obtained higher R^2s (Sunshine and Dicker, 1987c). As the second column in Table K

Table K

Comparison of multiple correlation coefficients squared, by dependent variable and age and family status relative to the poverty level

Age and status of family	Dependent variable: Index of financially burdensome family health expenses (preferred model)[1]	Dependent variable: Total family health charges (preferred model)[1,2]	Dependent variable: Index of financially burdensome family health expenses (full model)[1]
Older	0.53	0.72	0.54
Younger:			
Lower income	0.27	0.60	0.28
Better off	0.23	0.57	0.24

[1]Dependent variable is natural logarithm of stated statistic.
[2]Regression results not shown in this report.

NOTES: Older families are families with member(s) 65 years of age or over. Younger families are families with no member 65 years of age or over. Lower-income families are families with income below 200 percent of the poverty level. Better-off families are families with income of 200 percent of the poverty level or more.

shows, R^2s in the regression analyses of total charges was 0.57 to 0.72, depending on the family population involved. Thus, factors not included in (or poorly measured by) the 47 independent variables used in this report account for more of the variance in the financial burden index than they do for total family charges for health care.

It is also interesting to compare the R^2s of two alternate versions of the regressions for the financial burden index. Preferred models with 17 to 24 independent variables are the source of the findings reported and discussed in this chapter and are shown in Appendix Tables II, III, and IV. However, full models with 43 to 45 independent variables were also run using SUREGR, and are shown in Appendix Tables V, VI, and VII. (See Appendix I for more information on the two types of models.) The third column of Table K shows the R^2 for the full models. For all three family populations, it is only 0.01 higher than the R^2 for the preferred model. This is a typical finding for preferred models developed with stepwise regression (as the models presented here were), and shows that very little explanatory power is lost by using the preferred models rather than the full models.

Interestingly, for both the financial burden index and total family charges for health care, the independent variable sets used in this report account for more of the variance among older families than among younger family populations. This suggests that factors not included in the regressions (or poorly measured by the included variables) are less important for determining levels of these two health cost measures among older families than among younger families. Why this should be true is not obvious.

Individual Variables

The regressions indicate a very strong role for family income as a determinant of the financial burden index. Income is the only independent variable statistically significant for all three family populations, and for two of these populations its statistical significance, as measured by the F statistic, is stronger than that of any other statistically significant independent variable. This finding accords with the Berki et al. (1985) finding that one large population of families with a high level of the financial burden measure used in that study is composed of families with low incomes. Their out-of-pocket expenses for health care services may not be particularly large in dollar terms, but they are large relative to the small incomes of the families. This finding means that a family's ability to pay, as measured by income, is important in determining the financial burden the family faces from health costs. It indicates that measures of health costs that include an ability-to-pay factor (for example, the three ratio measures discussed in the section "Measuring financially burdensome health expenses") are likely to show quite different patterns than measures that assess health costs solely in dollar terms, without considering differences in ability to pay. However, it should not be surprising that family income is so prominent an explanatory variable. Family income is the denominator of the financial burden index and, thus, family income helps determine the index's level.

The regression coefficients for the family income variable show an interesting pattern. For older families and for better-off younger families, the coefficient indicates that for each 1-percent increase in income, total out-of-pocket expenses for health care increase by about 0.15 percent. In contrast, for lower-income younger families, total out-of-pocket expenses increase by about 0.44 percent for each 1-percent increase in income. There are at least two plausible explanations for this difference. One is that total health spending by lower-income families is severely constrained by their low income and so rises rather rapidly as their income increases. Another explanation is that as income increases for lower-income families, these families no longer receive charity care; and/or they lose the benefits of Medicaid and other public programs that require very small or no out-of-pocket payments. Thus, their total out-of-pocket expenses increase relatively rapidly with increasing income even though their total charges may not. The regression analyzing the determinants of total charges for lower-income younger families (Sunshine and Dicker, 1987c; regression not shown) supports the second of these two explanations.

One other point about income should be noted. Because income is in the regression as an independent variable, the regression estimates of the effects of other independent variables are made with the effect of income controlled. Thus, the effects the regressions show for other variables are not the result of a confounding effect of income. This means that other variables affect the index only by their effect on the numerator, which is the family's level of total out-of-pocket health expenses. While this point sounds obvious, it is important and easily overlooked.

Consider, for example, the statistically significant effect found for education of the family head among both of the younger family categories. Because the regressions control for the effects of income (and the effects of the other independent variables in the models), it is reasonable to conclude that a genuinely causal effect of education has been identified here. Quite possibly, the effect of education found in this study operated through differences in family valuation of health and health care. In contrast, an analysis that showed an apparent effect of education, but did not control for income, could well be finding only a spurious effect, for education and income are strongly correlated.

Age of the family head, like education, is significant in the two younger family populations but not among older families. The effect of age probably results from the well-known poorer health, greater risk for costly

illness, and greater use of health care characteristic of increasing age. The health variables used in the regressions (measures of generalized health status, specific illness, and so forth) probably do not fully capture this phenomenon.

The absence of a statistically significant effect of age among the older family population may well result from a limited variation in the age distribution among the members of older families. These families predominantly have heads 65 years of age or over, and few of the heads or members are more than 20 years older or younger than that. Moreover, the older families among multiple-person families tend to be among the younger families when all older families (both multiple-person and one-person) are taken into account. This would further restrict the age distribution among these families and reduce possible variation in the index associated with age.

A similar situation probably explains why, among lower-income younger families, income is not far more statistically significant than the other statistically significant independent variables. Among lower-income younger families, variation in the income distribution is also quite limited. (The main limit is the 200 percent-of-poverty cutoff that defines the income top of this family category, but very low-income families are also excluded.)

The relatively modest role of hospitalization in the regressions is notable. Hospitalization is statistically significant for only one population—younger, better-off families—and for that population one hospitalization was found to increase the level of the index by about 38 percent above that for otherwise similar families with no hospitalization. In contrast, one descriptive study that did not control for other factors (Kovar, 1986) showed elderly persons with hospitalization having, on average, total charges for health care of more than 1,000 percent of the total charges for elderly persons with no hospitalization. The regressions (previously mentioned) for total family charges for health care that cover the same three populations analyzed here and that, like the present study, control for the effects of many variables, also find an effect of hospitalization that is much larger than 38 percent (Sunshine and Dicker, 1987c; regressions not shown).

Given the difference between the findings of the regressions in this report and those of the parallel regressions for total charges, there seems to be a clear explanation for the modest role hospitalization plays in determining financially burdensome health expenses among families. Inpatient hospital care is particularly well insured and so is much less important in total out-of-pocket expenses than in total charges. Strong support for this conclusion is provided by a comparison between total charges and out-of-pocket expenses for various types of health care. Such a comparison is readily made by using parallel data from two expenditure reports. One (Sunshine and Dicker, 1987a) deals with family out-of-pocket expenses, and the other (Sunshine and Dicker, 1987b) presents corresponding data on total family charges. For multiple-person families, these sources show that only 8 percent of the total charges for inpatient hospital care were paid out-of-pocket in 1980 compared with out-of-pocket expenses ranging from 20 to 65 percent of total charges for all the other types of health care services examined.

The lowered level of the index characteristic of families with incomplete or no health care coverage is puzzling. It is found (in one form or another) for all three socioeconomic populations in the study. (Recall that families with "coverage source unknown" are generally families with incomplete or no coverage.) However, this insurance effect is the reverse of what the literature generally reports. The normal pattern is for limited or absent coverage to lead to reduced *total* charges because of the reduced use of health care resulting from the *high out-of-pocket payments* for health care that are required when coverage is limited or absent (Newhouse et al., 1981). Thus, a high index, not a low index, would be expected for families with incomplete or no coverage. There is some evidence that persons without insurance are disproportionately young persons in good health who, because of their youth and good health, are likely to experience relatively low total charges for health care (Kaspar, Walden, and Wilensky, 1980; Wilensky and Walden, 1981). However, that does not seem to be a likely explanation of the findings of this study, since the regressions control for health status and age.

The relatively high levels of the financial burden index found in the South is a regional effect not previously noted in the literature. However, the phenomenon seems real. It is found for two population groups (older families and younger, better-off families), and the regressions in this study control for a large number of variables, making spuriousness unlikely. In principle, four explanations are possible. First, total charges for health care may be relatively high in the South. Second, the proportion of total charges that have to be paid out-of-pocket may be particularly high there. Third, families may have to pay an unusually large part of the premiums for health care coverage; and, fourth, incomes may be relatively low in the South. As indicated above, the fourth explanation is ruled out because the regressions control for income. The first explanation is eliminated by the regressions for total charges (Sunshine and Dicker, 1987c; regression not shown), as these regressions do not show total charges to be high in the South. Thus, it seems that southerners have to pay a particularly large portion of health care cost out-of-pocket—either as out-of-pocket payments for health care services or as family-paid premiums for health care coverage. As most health care coverage in the United States is employment-based, and as this is particularly true for younger, better-off families—one of the family categories that shows an elevated financial burden index in the South—it is reason-

able to conclude that employment-provided health care coverage is relatively poor in the South. Among older families, however, the explanation is probably partially different. Medicaid coverage, for example, may be more restrictive, so that older families in the South may have to pay a larger amount out of pocket for health care than similar families in other regions of the United States.

Head-only families were found to have a significantly lower index than head-and-spouse families among both older families and younger, lower-income families. This family structure effect is difficult to explain, particularly because it is not found among younger, better-off families or in regressions for total family charges for health care. It is not a matter of smaller family size, the presence of children, or coverage by Medicaid among head-only families, as these factors are controlled for.

Finally, there is the lowered level of the index among younger, better-off families with a black head of family. This race-related finding should be regarded as likely to be a genuine one—and possibly a sign of an important social problem—because the regressions in this study control for many possibly confounding variables such as education, health status, family structure, income, and location of residence. Moreover, a similar finding for the same family population was obtained in the regressions for total family charges for health care. In contrast, statistical associations with race that are found in analyses that do not control for possible confounding factors—for example, many tabular analyses—may only be artifacts of these confounding factors.

Race-related differences were not found among either older families or younger, lower-income families. This was true of regressions for both the financial burden index and for total family charges for health care. Older families and younger lower-income families are usually regarded as more vulnerable than other family populations and certainly have been a focus of attention for governmental health care coverage programs. The absence of racially caused differences among these family categories is a positive sign.

Patterns Among Variables

Perhaps the most important aspect of the findings of this report is the wide variety of factors that are found to be significant determinants of the level of the financial burden index. The study analyzed the effects of family variables that can be classified into seven categories: demographic, sociocultural, specific illnesses and special health events, general health status, economic, health insurance, and geographic. For all three populations analyzed, variables in at least five of these seven categories were found to be significant. Thus, an important conclusion of the study is that the determinants of the level of the index are numerous and of several types.

One way to assess the importance of the findings from these regression analyses (other than by significance levels and the size of regression coefficients) is to examine how the statistically significant variables in the regressions were distributed among the three family populations. Table L shows that 16 of the variables were significant in one or more of the three family socioeconomic populations. However, only one variable was statistically significant in all three populations, eight other variables were statistically significant in two of the populations, and another seven were statistically significant in only one population. Table L thus gives one

Table L

Statistically significant variables from a set of regressions on the index of financially burdensome family health care expenses arranged by the number of family socioeconomic populations in which each variable was statistically significant

Variable	Statistically significant in— 3 populations	2 populations	1 population
Family income	O,L,B		
Head-spouse structure		O,L	
Coverage by Medicaid		O,L	
Source of health care coverage unknown		O,L	
Other public coverage		L,B	
Other public and private coverage		L,B	
Age of head of family		L,B	
Education of head of family		L,B	
Region		O,B	
Medicare and other public coverage			O
Heart and circulatory disease			O
Family work-loss days			L
Family illness days in bed			B
One or more hospitalizations			B
Race of head of family			B
Completeness of health care coverage			B

NOTE: O = Older families
L = Younger, lower-income families
B = Younger, better-off families

measure of the importance of certain variables in affecting values of the financial burden index. This measure is the universality of the effects of the variables in the U.S. multiple-person family population. As noted, from this perspective family income is the most important variable because it is a statistically significant determinant of the index among all segments of the U.S. multiple-person family population.

Second in apparent importance are the eight variables that were found to determine values of the index among two socioeconomic segments of the U.S. multiple-person family population. These eight variables were head-spouse structure, age of the head, education of the head, geographic region of residence, and four health care coverage variables. Clearly, source of health insurance coverage emerges from this list as a very important general category for determining levels of the financial burden index, as four of the eight variables are in this category.

Further evidence of the importance of health insurance variables is the large difference they make in the level of the financial burden index. Public coverage (other than Medicare alone) usually reduces the index by two-thirds or more from what it otherwise would be.

Third in apparent importance are the seven variables that were significant for only one family population. These involve the race of the head, two insurance variables (reinforcing the statement above that insurance is very important for determining values of the index), and a set of four variables that more or less directly measure the family's need for health services (presence of heart or circulatory disease, work loss days due to illness, illness days in bed, and hospitalization). Because these last four variables all have to do with health, it can be concluded that health factors are significant in their effect on the financial burden index, but that their effect is scattered and relatively diffuse.

Indeed, the regressions for the financial burden index, as Table L shows, are conspicuous in the relatively small role of health variables. (These include generalized health status, illness, and special health event variables.) The limited importance of health factors becomes particularly clear from a comparison of Table L with Table M. Table M presents a summary of the results of the regressions that used total family health charges as the dependent variable (Sunshine and Dicker, 1987c; regressions not shown). These regressions for total charges cover the same three family socioeconomic populations and start from the same 47 independent variables as the regressions summarized in Table L. However, in Table M, generalized health status, illness, and special health event variables are major determinants of the dependent variable, total family charges for health care. Of the 15 variables in Table M, 8 come from these categories. By comparison, only 4 of the 16 variables in Table L, which deals with the financial burden index, are of this type. Moreover, in the regressions for total family charges for health care (Table M), six of the eight health variables were significant in two or more populations, with two of them being significant in all three populations. In contrast, in the regressions for the financial burden index (Table L), no health variable was significant in more than one population. It appears, therefore, that health status (in the broad sense) of family members is the major determinant of total family health care charges, but that the major determinants of family financially burdensome health expenses are family income and type and completeness of health insurance coverage.

Table M

Statistically significant variables from a set of regressions on total family charges for health care arranged by the number of family socioeconomic populations in which each variable was statistically significant

Variable	Statistically significant in—		
	3 populations	2 populations	1 population
Hospitalization	O,L,B		
Family illness days in bed	O,L,B		
Cancer		O,L	
Heart and circulatory disease		O,B	
Accidents, poisonings, and injuries		O,B	
Family income		O,B	
Perceived health status		L,B	
Region		L,B	
Completeness of health care coverage		L,B	
Limitation in major activity			L
Family work-loss days due to illness			B
Presence of child			B
Age of head of family			B
Race of head of family			B
Education of head of family			B

NOTE: O = Older families
L = Younger, lower-income families
B = Younger, better-off families

This finding seems to be in partial conflict with the conclusion of Berki et al. (1985) that families with costly illnesses are one of the two major populations of families with high proportions of family income consumed by out-of-pocket expenses for health care services. Although Table L shows some role for such illnesses as determinants of the financial burden index, it is a small role, particularly when compared with the role of illnesses as determinants of total family charges for health care (Table M). Quite likely, the explanation of the relatively small role of illness variables (and other health variables) as determinants of the financial burden index is the same as the explanation given above for the relatively small role of hospitalization. Health care coverage probably pays for a relatively large percent of the cost of major illnesses, making their role in out-of-pocket expenses (which is the numerator of the financial burden index) relatively small.

Concluding Remarks

Table L indicates that different variables from the model were found to be associated with the financial burden index in a statistically significant manner in different socioeconomic populations. Although, with hindsight, some of the associations are obvious, others are not. Some of the obvious associations are Medicare coverage with the older population, Medicaid coverage with lower-income populations (both older and lower-income younger families), and family work-loss days with a younger population (but, surprisingly, not with both younger populations). What these associations suggest is that the effects on the financial burden index of the variables found to be statistically significant are not universal, but rather are specific to particular socioeconomic populations. To some extent, then, policy aimed at alleviating financially burdensome family health care expenses may have to encompass different solutions for different populations.

It should also be noted that there were a number of variables in the preferred models that were not statistically significant in any of the three socioeconomic family populations of interest. Whether these variables would be statistically significant in other populations or with a larger sample cannot be assessed at this time. Moreover, it should be noted that a single regression analysis that encompassed the entire U.S. multiple-person family population might produce different findings about the relative importance of variables than do the three separate regressions reported here. So might an analysis that did not involve the exclusions found in this section on the regressions. (The most important of these exclusions are families with zero or very low income, families with zero as their total out-of-pocket health expense, one-person families, institutionalized persons, and expenses for long-term care.)

Nonetheless, the preceding discussion strongly suggests that to succeed, policies to alleviate financially burdensome health care expenses among U.S. multiple-person families must concentrate on improving health care coverage and/or on increasing income. These are the two types of factors that this analysis has shown are the most important determinants of families' financial burden for health care costs.

References

Aday, L. A., Fleming, G. V., and Andersen, R.: Access to medical care in the U.S.: Who has it, who doesn't. *Research Series*. No. 32. Center for Health Administration Studies, University of Chicago. Chicago. Pluribus Press, 1984.

Andersen, R., and Benham, L.: Factors Affecting the Relationship Between Family Income and Medical Care Consumption, in H. E. Klarman, ed., *Empirical Studies in Health Economics*. Baltimore. The Johns Hopkins Press, 1970.

Anderson, J. M., and Thorne, E.: Estimates of Aggregate Personal Health Care Expenditures in 1980. Comparison of the National Health Accounts and the National Medical Care Utilization and Expenditure Survey Data. Working Paper No. 30, Division of Health Interview Statistics, National Center for Health Statistics, Hyattsville, Md., 1985.

Arnold, S. F.: *Theory of Linear Models*. New York. John Wiley and Sons, 1981.

Berki, S. E.: A look at catastrophic medical expenses and the poor. *Health Affairs* 5(4):138–145, Winter 1986.

Berki, S. E., Wyszewianski, L., Magilavy, L. J., et al.: Families with high out-of-pocket health service expenditures relative to their income. Final Report on Phase II of Contract No. 233–81–3022. Rockville, Md. National Center for Health Services Research and Health Care Technology Assessment, Public Health Service, Jan. 30, 1985.

Binder, D. A.: On the variances of asymptopically normal estimators from complex surveys. *Int. Statistical Rev.* 51:279–292, 1983.

Bonham, G. S.: Procedures and questionnaires of the National Medical Care Utilization and Expenditure Survey. *National Medical Care Utilization and Expenditure Survey*. Series A, Methodological Report No. 1. DHHS Pub. No. 83–20001. Public Health Service. Washington. U.S. Government Printing Office, Mar. 1983.

Buczko, W.: Physician utilization and expenditures in a medicaid population. *Health Care Financ. Rev.* 8(2):17–26, Winter 1986.

Campbell, D. T., and Julian, C. S.: *Experimental and Quasi-Experimental Designs for Research*. Chicago. Rand McNally College Publishing Co., 1980.

Catastrophic Illness Expenses. Department of Health and Human Services report to the President. U.S. Department of Health and Human Services, Nov. 1986.

Cox, G., Parker, A. E., Sweetland, S. S., et al.: *Imputation of Missing Item Data for the National Medical Care Utilization and Expenditure Survey*. Prepared for the National Center for Health Statistics under Contract No. HRA–233–79–2032. Research Triangle Park, N.C. Research Triangle Institute, 1982.

Cox, G., and Sweetland, S. S.: *Imputation of Attribution-Related Missing Data for the National Medical Care Utilization and Expenditure Survey*. Prepared for the National Center for Health Statistics under Contract No. HRA–233–79–2032. Research Triangle Park, N.C. Research Triangle Institute, 1982.

Dicker, M.: Health care coverage and insurance premiums of families, United States, 1980. *National Medical Care Utilization and Expenditure Survey*. Preliminary Data Report No. 3. DHHS Pub. No. 83–20000. Public Health Service. Washington. U.S. Government Printing Office, May 1983a.

Dicker, M.: Panel Surveys and Family Level Measures, Problems and Solutions, in *1983 Proceedings of the American Statistical Association*. Social Statistics Volume. Washington. American Statistical Association, 1983b.

Dicker, M.: Demographic Analysis and a General Theory for Measuring Social Change in Organizations. The Case of the Family. Paper presented at the annual meeting of the Southern Regional Demographic Group. Orlando, Fla., Oct. 19, 1984.

Dicker, M., and Casady, R. J.: A Reciprocal Rule Model for Defining Longitudinal Families for the Analysis of Panel Survey Data, in *1982 Proceedings of the American Statistical Association*. Social Statistics Volume. Washington. American Statistical Association, 1982.

Dicker, M., and Casady, R. J.: Empirical Findings on the Distribution and Construction of Longitudinal Families, Part I, Modeling the Universe, Inscope Changes, and Static and Dynamic Families, in *1984 Proceedings of the American Statistical Association*. Survey Research Methods Volume. Washington. American Statistical Association, 1984.

Dicker, M., and Casady, R. J.: An Introductory Discussion of Issues in the Methodology of Longitudinal Family Surveys. Paper presented at the All Day Measurement Workshop of the Family Health Section at the annual meeting of the National Council on Family Relations. Dallas, Texas, 1985.

Dicker, M., and Sunshine, J. H.: Family Use of Health Care, United States, 1980. *National Medical Care Utilization and Expenditure Survey*. Series B, Descriptive Report No. 10. DHHS Pub. No. 87–20210. Hyattsville, Md. National Center for Health Statistics, Public Health Service, 1987.

Duan, N., Manning, W. G., Jr., Morris, C. N., et al.: *A Comparison of Alternative Models for the Demand for Medical Care*. Contract No. R–2754–HHS. Santa Monica, Calif. Rand Corporation, Jan. 1982.

Farley, P. J.: Private insurance and public programs, Coverage of health services. *National Health Care Expenditures Study*. Data Preview 20. DHHS Pub. No. (PHS) 85–3374. Rockville, Md. National Center for Health Services Research and Health Care Technology Assessment, Mar. 1985.

Farley, P. J.: Hospital and ambulatory services for selected illness. *Health Serv. Res.* 21(5):587–634, Dec. 1986.

Feldstein, P. J.: *Health Care Economics,* 2d ed. New York. John Wiley and Sons, 1983.

Hershey, J. C., Luft, H. S., and Gianaris, J. M.: Making sense out of utilization data. *Med. Care* 13(10):838–852, Oct. 1975.

Holt, M. M., and Shah, B. V.: *SURREGR, Standard Errors of Regression Coefficients From Sample Survey Data*. Research Triangle Park, N.C. Research Triangle Institute, Apr. 1982.

Kaspar, J. A., Walden, D. C., and Wilensky, G. C.: Who are the uninsured? *National Health Care Expenditures Study*. Data Preview 1. Hyattsville, Md. National Center for Health Services Research, Public Health Service, 1980.

Koretz, D.: *Catastrophic Medical Expenses: Patterns in the Non-Elderly, Non-Poor Population*. Washington, D.C. Congressional Budget Office, Dec. 1982.

Kovar, M. G.: Expenditures for the medical care of elderly people living in the community in 1980. *Milbank Mem. Fund Q.* 64(1):100–132, 1986.

Landis, J. R., Lepkowski, J. M., Eklund, S. A., and Stehouwer, S. A.: A statistical methodology for analyzing data from a complex survey, the first National Health and Nutrition Examination Survey. *Vital and Health Statistics*. Series 2, No. 92. DHHS Pub. No (PHS) 82–1366. National Center for Health Statistics, Public Health Service. Washington. U.S. Government Printing Office, Sept. 1982.

Levy, P. S., and Lemeshow, S.: *Sampling for Health Professionals*. Belmont, Calif. Lifetime Learning Publications, 1980.

Mallows, C. L.: Some comments on C(p). *Technometrics* 15:661–675, 1973.

McCarthy, P. J.: Replication, an approach to the analysis of data from complex surveys. *Vital and Health Statistics*. Series 2, No. 14. DHEW Pub. No. (PHS) 79–1269. National Center for Health Statistics, Public Health Service. Washington. U.S. Government Printing Office, Apr. 1966.

McMillen, D. B., and Herriot, R.: Towards a Longitudinal Definition of Households. SIPP Working Paper Series No. 8402. U.S. Bureau of the Census. Washington, D.C., 1984.

Moser, B., Whitmore, R., Frick, G. G., et al.: *National Medical Care Utilization and Expenditure Survey: Analytic Family File Construction Methodology Report*. RTI Project No. 251U–1898–14. Research Triangle Park, N.C. Research Triangle Institute, Oct. 1983.

Neter, J., and Wasserman, W.: *Applied Linear Statistical Models*. Homewood, Ill. Richard D. Irwin, Inc., 1974.

Newacheck, P. W.: *Utilization and Expenditures for Medical Care Services Provided to Children with Activity Limitations*. Grant MCJ–063468–01–0. Prepared for National Maternal and Child Health Resource Center. Institute for Health Policy Studies, University of California, San Francisco, 1985a.

Newacheck, P. W.: *Prevalence and Severity of Chronic Conditions Among Children*. Prepared for National Maternal and Child Health Resource Center. Institute for Health Policy Studies, University of California, San Francisco, 1985b.

Newhouse, J. P., Manning, W. G., Morris, C. N., et al.: Some interim results from a controlled trial of cost sharing in health insurance. *N. Engl. J. Med.* 305:1501–1507, 1981.

Public Use Data Tape Documentation: Family Data. *National Medical Care Utilization and Expenditure Survey, 1980*. Hyattsville, Md. National Center for Health Statistics, Public Health Service. Nov. 1986.

SAS Institute, Inc.: *SAS Users Guide, Basics,* 1982 ed. Cary, N.C. SAS Institute, Inc., 1982.

SAS Institute, Inc.: *SAS/STAT Guide for Personal Computers, Version 6 Edition*. Cary, N.C. SAS Institute, Inc., 1985.

Shah, B. V.: *SESUDAAN, Standard Errors Program for Computing Standardized Rates From Sample Survey Data*. Research Triangle Park, N.C. Research Triangle Institute, Apr. 1981.

Sunshine, J.H.: Medicare, past, present, and future. *Natl. J.* 14(23):1030–1033, June 5, 1982.

Sunshine, J. H.: How many Americans lack outside sources of payment for major health care costs? *J. Health Hum. Resour. Adm.* 6(3):341–360, Winter 1984.

Sunshine, J. H., and Dicker, M.: Family out-of-pocket expenditures for health care: United States, 1980. *National Medical Care Utilization and Expenditure Survey*. Series B, Descriptive Report No. 11. DHHS Pub. No. 87–20211. National Center for Health Statistics, Public Health Service. Washington. U.S. Government Printing Office, Aug. 1987.

Sunshine, J. H., and Dicker, M.: Total family expenditures for health care: United States, 1980. *National Medical Care Utilization and Expenditure Survey*. Series B, Descriptive Report No. 15. DHHS Pub. No. 87–20215. National Center for Health Statistics, Public Health Service. Washington. U.S. Government Printing Office, Sept. 1987.

Sunshine, J. H., and Dicker, M.: Determinants of total family charges for health care: A multiple regression analysis. 1987c (unpublished).

Taube, C. A., Kessler, L. G., and Burns, B. J.: Estimating the probability and level of ambulatory mental health services use. *Health Serv. Res.* 21(2, Pt II):321–340, June 1986.

Whitmore, R. W., Cox B. G., and Folsom, R. E.: *Family Unit Weighting Methodology for the National Household Survey Component of the National Medical Care Utilization and Expenditure Survey*. Contract No. HRA–233–79–2032. Prepared for the National Center for Health Statistics. Research Triangle Park, N.C. Research Triangle Institute, 1982.

Wilensky, G. R., and Walden, D. C.: Minorities, Poverty, and the Uninsured. Paper presented at the 109th Annual Meeting of the American Public Health Association. Los Angeles, Nov. 1981.

Wyszewianski, L.: Financially catastrophic and high-cost cases: Definitions, distinctions, and their implications for policy formation. *Inquiry* 23(4):382–394, Winter 1986a.

Wyszewianski, L.: Families with catastrophic health care expenditures. *Health Serv. Res.* 21(5):617–634, Dec. 1986b.

Appendixes

Contents

List of Appendix Tables

Appendix I
Technical Notes on Regression Methods

Introduction

Multiple regression analysis is a statistical method for examining the effect of a set of independent (or causal) variables on a continuous dependent variable. It permits the analysis of a large number of independent variables, both continuous and categorical, and provides separate estimates for the effects of each. The set of variables used in the regression analysis, together with the functional form of the equation that relates the independent variables to the dependent variable, is called the regression model. By analyzing the coefficients in the regression model, one can explore relationships between the dependent and independent variables and make predictions about the future behavior of the independent variable.

This appendix first presents a technical description of multiple regression analysis with special attention given to the application of this technique to complex surveys (such as the National Medical Care Utilization and Expenditure Survey, NMCUES, which is the source of data used in this report). The appendix then presents a series of somewhat less technical sections. These cover, in turn, the regression model used in this report, the interpretation of regression coefficients and of means of variables, and the analytic procedures followed in this report.

Technical Description of Multiple Regression Analysis

Introduction

This section discusses the use of multiple regression analysis with special emphasis on its application to complex survey data and, in particular, on its use in analyzing the NMCUES data of this study. The following topics are covered:

- The basic structure of the regression model and how it is usually applied in finite population sampling,
- Estimation of regression parameters using complex survey data, and
- Estimation of variances of parameter estimates.

The Regression Model in Finite Population Sampling

In most statistical literature, the regression model is stated as:

$$y_i = \underline{x}_i' \underline{b} + e_i \qquad (1)$$

where y_i is an observable random variable, $\underline{x}_i$ is a $p \times 1$ vector of independent variables, $\underline{b}$ is a $p \times 1$ vector of unobservable regression coefficients, and e_i is a random variable with $E(e_i) = 0$ and $\text{Var}(e_i) = \sigma^2$. The sequences $y_1, y_2, \ldots$ and $e_1, e_2, \ldots$ are generally assumed to be independent and identically distributed. In the normal regression model, it is further specified that e_i has a normal distribution.

In finite population sampling, the regression model looks similar but is formulated in a fundamentally different way. When sampling from a finite population, it is assumed that there is a population of N pairs $(y_i, \underline{x}_i)$:

$$P = \{(y_1, \underline{x}_1), (y_2, \underline{x}_2), \ldots, (y_N, \underline{x}_N)\} \qquad (2)$$

from which a sample of n pairs is selected. Note that y_i is considered to be fixed in finite population sampling and the randomness is introduced by the selection process, whereas each y_i in the infinite population regression model is considered to be a random variable. The same is true of the e_i. In the finite population setting, the expression "independent and identically distributed" has no real meaning.

The regression coefficients are determined in the finite population setting by the least squares equation:

$$\underline{b} = (X'X)^{-1}X'\underline{y}, \qquad (3)$$

where X is the matrix whose rows are made up of the $\underline{x}_i'$. In the infinite population situation neither the $X'X$ nor $X'\underline{y}$ exist in any meaningful sense, since each would consist of divergent infinite sums. (Equation (3) would, however, be used in a sample from an infinite population to estimate the regression coefficients.)

Finally, the error terms e_i in the finite population setting are defined as the residuals from the least squares equation in (3):

$$e_i = y_i - \underline{x}_i' \underline{b}. \qquad (4)$$

Thus, in sampling from a finite population, the e_i can be seen as a population of fixed values, from which a sample of size n is drawn. Again, this contrasts with the usual regression model, where each e_i is considered to be random.[1]

Estimation of Regression Parameters

In equation (3), it was indicated that $\underline{b}$ is given by:

$$\underline{b} = (X'X)^{-1} X'\underline{y}.$$

The elements of the matrix X'X are given by:

$$z_{ij} = \sum_k x_{ki}x_{kj} \tag{5}$$

while the elements of the vector $X'\underline{y}$ are given by:

$$t_i = \sum_k x_{ki}y_k. \tag{6}$$

To estimate $\underline{b}$, then, it is necessary to estimate z_{ij} and t_i. Let:

$$w_i = \text{sampling weight of } i\text{th sample unit.} \tag{7}$$

Unbiased estimates of z_{ij} and t_i are given by:

$$\hat{z}_{ij} = \sum_k x_{ki}x_{kj}w_k \tag{8}$$

and

$$\hat{t}_i = \sum_k x_{ki}y_k w_k. \tag{9}$$

It must be noted that while these estimates are unbiased, the estimate of $\underline{b}$ obtained by forming

$$\hat{\underline{b}} = \hat{Z}^{-1}\hat{\underline{t}} \tag{10}$$

is not. Nor is the variance of this estimate easy to calculate, as the next section shows.

Estimating Variances of Regression Coefficients

When sampling from an infinite population, the covariance matrix of $\hat{\underline{b}}$ is given by:

$$\text{Cov}_\infty(\hat{\underline{b}}) = (X'X)^{-1}\sigma^2 \tag{11}$$

where σ^2 is the variance of the independent variable y. When sampling from a finite population, this formula is somewhat more complicated.

Define the vector $\underline{u}$ by

$$u_i = \sum_k (y_k - x_k'\underline{b})\, x_{ki}\, w_k. \tag{12}$$

Notice that u_i is a Horvitz-Thompson type sum whose variance can be calculated using the familiar rules of stratified and/or cluster sampling. We can now write a formula for the approximate covariance matrix of $\hat{\underline{b}}$ as:

$$\text{Cov}(\hat{\underline{b}}) = (X'X)^{-1}\hat{\Sigma}(\underline{u})\,(X'X)^{-1} \tag{13}$$

where $\hat{\Sigma}(\underline{u})$ is an estimate of $\text{Cov}(\underline{u})$. (See Binder (1983), pp. 279–292.)

This method of variance estimation is often called "Taylorizing" or "linearizing," since the Taylor expansion is used to develop the linear approximation on which equation (13) is based. SURREGR (Holt and Shah, 1982), the computer software package used in the final steps of the regression analysis of this report, uses this "Taylorizing" or linearizing technique.

The Model

Introduction

In this section, the model used in the regression analysis reported in this study is discussed. Three topics are covered in turn: first, the variables used in the regression analysis; second, the functional form of the hypothesized relationship between the dependent variable and the independent variables; and, third, the rationale for using this functional form.

Variables Used in the Regression Analysis

The variables used in the regression analysis are listed and described in Table I, which appears at the end of this appendix. The 48 variables in this table were either taken directly from or constructed from the variables on the NMCUES family data tape. The index of financially burdensome family health care expenses was used as the dependent variable because it is judged to be the best available measure of the burden of health expenses on families. The 47 independent variables were chosen from the larger set of family variables available from the NMCUES on the basis of previous research and the desire to include the variables believed likely to have the greatest explanatory power. In developing a regression model, the number of variables is limited by the amount of data available; screening out unnecessary variables greatly facilitates the analysis. The 48 variables are discussed further in the text of this report.

[1] It should be noted that this is the "classical" point of view in survey sampling. Increasingly, superpopulation models are used in survey inference. They assert that the finite population under study is simply a large "sample" from an infinite "superpopulation."

in the section titled "The Model" in the chapter, "The Determinants of Financially Burdensome Health Expenses."

Functional Form of the Regression Equation

The functional form of the relationship underlying the regression equation used in this report is

$$y_i = \prod_{j=1}^{s} (x_{ij}^{b_j}) \text{EXP}\{b_o + \sum_{j>s} b_j x_{ij} + e_i\}, \quad (14)$$

where s is the number of independent variables transformed into their natural logarithm in the estimating equation (see below). As in equation (1), y_i is the dependent variable (in this report, it is the financial burden index, defined as the ratio of annualized total out-of-pocket health expenses to annual family income), the x_{ij} are the independent variables, the b_i are the regression coefficients, and e_i is the error term with $E(e_i) = 0$. (The notation EXP means that "e," the base of natural logarithms (=2.71828...), is to be raised to the power indicated by the expression in braces that follows EXP.) Expanding the products in equation (14) yields

$$y_i = (x_{i1}^{b_1})(x_{i2}^{b_2})(x_{i3}^{b_3}) \ldots \quad (14a)$$
$$(\text{EXP}\{b_o\})(\text{EXP}\{b_{s+1} x_{i,s+1}\})$$
$$(\text{EXP}\{b_{s+2} x_{i,s+2}\}) \ldots (\text{EXP}\{e_i\}).$$

Equation (14) and equation (14a), which are mathematically equivalent, are not linear in the regression coefficients (the b_i) and so cannot be estimated in the fashion described earlier in this appendix. However, these equations have a number of desirable features, as described below, and were selected because of these features.

One desirable feature is that they are easily transformed into equations that *are* linear in the regression coefficients. (For more on transformation of variables in regression analysis, see Neter and Wasserman, 1974, pp. 123–127.) The transformation necessary to achieve linearity consists of taking the natural logarithm of both sides of the equations. Taking the natural logarithm of both sides of equation (14) yields

$$\ln(y_i) = b_o + \sum_{j=1}^{s} (b_j \ln(x_{ij})) + \sum_{j>s} (b_j x_{ij}) + e_i \quad (15)$$

Similarly, taking the natural logarithm of both sides of equation (14a) yields

$$\ln(y_i) = b_o + b_1 \ln(x_{i1}) + b_2 \ln(x_{i2}) + \ldots + b_{s+1} x_{i,s+1} + b_{s+2} x_{i,s+2} + \ldots + e_i. \quad (15a)$$

Equation (15) and equation (15a), which are algebraically the same, *are* linear in the regression coefficients (the b_j) and are the regression equation used in the analyses in this report. Their parameters and the variances of these parameters were estimated by the techniques described in the first section of this appendix.

A point to note here with respect to equation (15) is that in this equation, some of the original NMCUES variables are transformed into their natural logarithms. That is, these variables are replaced by their natural logarithms. Table I indicates the variables for which the natural logarithm, rather than the untransformed variable, was used.

Also, equations (14) and (15) require any categorical independent variables to be expressed in numerical form. This is readily accomplished for variables which take on only two values. For these variables, it is accomplished by assigning "1" to one of the values and "0" to the other. For example, sex of the head of the family was coded as female = 1 and male = 0. For categorical variables that take $k \geq 3$ values, an extension of this procedure was used. Such categorical variables were represented in the regression equation by a series of $k-1$ "dummy" variables, each of which can take the value of "1" or "0." Each $k-1$ value of the original categorical variable was associated with one dummy variable, which was assigned the value of 0 or 1 for each family in the sample depending on whether the value (of the original categorical variable) was true (dummy = 1) or not (dummy = 0). For example, the head-and-spouse structure of a family takes on three values: (1) head and spouse always present, (2) family always has only a head, and (3) changing head-and-spouse structure. Dummy variables, D1 and D2, were created from the second and third of these three values, while the first value was the omitted value.

In creating dummy variables, one of the original k values of the original variable must be omitted, since the kth dummy variable would be a linear combination of the first $k-1$ dummy variables. (If one independent variable in a regression is a linear combination of others, the matrix $X'X$ (see equation (3)) cannot be inverted and the regression coefficients are thus undefined.) Typically, the omitted value was the one regarded conceptually as the "base case" or was the most common state. Table I shows the dummy variables that were used and indicates which value was omitted.

Rationale for the Functional Form

The dependent variable. The functional form shown in equation (14) requires that the dependent variable be used in logarithmic form when estimating the regression model using equation (15). This functional form was chosen for three reasons.

First, it is believed that the relationship between the independent and dependent variables is primarily multiplicative. For example, it was expected that the reduction in the financial burden index associated with Medicaid coverage would be multiplicative rather than additive. To simplify greatly, in an additive model, if Medicaid coverage generally reduced the index from 8 percent to 2 percent for some families, it should reduce it from 6 percent to 0 percent for other families and reduce it from 4 percent to minus 2 percent—an obvious absurdity—for yet others. In contrast, a multiplicative relationship would, following this example, generally reduce the index to about one-fourth of its value—that is, reduce it from 8 percent to 2 percent for the first category of families, from 6 percent to 1.5 percent for the second category, and from 4 percent to 1 percent for the third. A multiplicative effect of this type is what the authors believed likely. When underlying relationships are multiplicative, a functional form with a logarithmic dependent variable in the regression model (that is, in equation (15)) is appropriate because such a functional form is multiplicative in its untransformed version (that is, equation (14)).

Second, a logarithmic dependent variable was used because prior research in evaluating the appropriateness of different functional forms for regression analysis of medical expenditure data indicates that a logarithmic dependent variable should be used in the estimating equation. Duan et al. (1982) carried out an extensive analysis of residuals from various models and found that they approximated a normal distribution (as assumed by the linear regression model) more closely when a logarithmic dependent variable was used than when an untransformed dependent variable was used in the regression model.

Third, the literature on health expenditures, perhaps because of the preceding two reasons, almost always uses a logarithmic dependent variable in regression equations. Some examples, which are similar to this report in the data bases they use or the subjects they investigate, are Taube et al. (1986) and Farley (1986). By following the literature, comparability of results is enhanced.

Independent variables. Given a logarithmic dependent variable in the estimating equation (equation (15)), there remains the question of whether or not to transform the (continuous) independent variables in the equation into their logarithms. The alternatives of carrying out such a transformation or not doing so imply different relationships. The following paragraphs first describe in nontechnical terms the relationships implied by each alternative and then describe the choices made.

When the dependent variable is logarithmic, a logarithmic independent variable in the estimating equation implies constant elasticity. In nontechnical terms, this means that a 1 percent increase in the untransformed independent variable produces a fixed percent increase (or decrease) in the untransformed dependent variable, with this increase (or decrease) called the elasticity. (Technically, elasticity is defined as $(\partial y_i/\partial x_{ij})\ (x_{ij}/y_i)$, and the nontechnical description in the preceding sentence is usually a close approximation to elasticity as measured by this formula.) A logarithmic independent variable also implies that a fixed percent change in the untransformed independent variable—for example, increasing it by 100 percent (that is, doubling it) produces a uniform percent change in the untransformed dependent variable. In contrast, with a logarithmic dependent variable, using a continuous independent variable in nontransformed form in the estimating equation implies that it is a unit increase in the nontransformed independent variable (not a percent increase) that produces a uniform percent change in the nontransformed dependent variable.

Manipulation of equation (14) will show that these relationships hold. [Independent variables used in logarithmic form in the estimating equation (equation (15)) are the first s of the x_{ij}, and these are found in the product term of equation (14). Independent variables used without transformation in the estimating equation (equation (15)) are the remaining x_{ij}, and these are found in the exponential term of equation (14).]

In light of the different relationships that hold true for logarithmic and untransformed continuous independent variables in the estimating equation (equation (15)), a choice was made between these two forms for continuous independent variables. The choice was based on beliefs about the nature of the underlying relationships. For example, it was believed that if an increase from 10 to 20 annualized family illness days spent in bed ("bed days") resulted in, say, a doubling of the financial burden index, then a further increase from 20 to about 40 bed days would be required to produce another doubling of the index. Hence bed days was used in logarithmic form in the estimating equation (equation (15)). Using bed days untransformed would imply that an increase in bed days from 20 to about 30 would produce the second doubling of the index. In contrast, to take a second example, it was believed that if an increase in the age of the family head from 30 to 40 years produced a given percent increase in the index (say, increasing it by 30 percent—that is, multiplying it by 1.3), then an increase in age of the family head from 54 to approximately 64 years would produce an equally large percent change (that is, a 30 percent increase) in the index. Hence, age of the family head was used untransformed. If it were used in logarithmic form, an increase in the head's age from 54 to 72 years would be required to generate the same effect (in percent terms) as the increase from age 30 to 40.

Based on previous research about the underlying relationships, the following four continuous variables were used in logarithmic form in the estimating equation (equation (15)):

- Average family size.
- Annualized family bed days.
- Annualized family work days lost because of illness ("work-loss days").
- Family annual income.

Because family bed days and family work-loss days can take on the value zero, and the logarithm of zero is undefined, the value of these variables was increased by one before performing the logarithmic transformation.

The following three continuous independent variables were used in untransformed form:

- Age of head of family.
- Years of education of head.
- Annualized number of hospital discharges for family members.

Using the PC SAS computer program (SAS Institute, 1985), informal, exploratory tests were conducted on the effect of the choice between no transformation of the above independent variables and a logarithmic transformation of them. There was little difference in R-square and, in general, relatively little difference in tests of statistical significance for each of these independent variables when a logarithmic transformation was substituted for no transformation and vice versa. This would indicate that significant factors can be detected using either the logarithmic or untransformed dependent variable. This can be explained partially by the fact that strong logarithmic effects will also have roughly linear patterns. It should also be noted that the results of significance tests will generally be valid in regressions when using large data sets because of the asymptotic normality properties of least squares estimators. That is, even though the residuals may not be normally distributed (one of the key assumptions on which the regression F tests are based), the F tests used in regression will be valid for large data sets. (See Arnold, 1981.) Thus the results of the significance tests may be similar even if the logarithmic transformation improves the normality of the regression residuals.

Interpreting the Regression Coefficients

Because the estimating equation (equation (15)) involves variables in logarithmic form, interpretation of the regression coefficients is somewhat complex. The reader may find the following explanation helpful in interpreting the regression coefficients, which appear in Appendix Tables II, III, and IV.

When the dependent variable in a regression model is in logarithmic form in the estimating equation, as is the financial burden index in all the regressions in this report, three different types of independent variables can be distinguished.

1. First are dummy variables. (Again, "dummy variable" designates a categorical variable that takes on only the values 0 and 1.) The regression coefficient, b, for a dummy variable has the following interpretation. The presence of the characteristic indicated by the dummy variable is associated with multiplication of the underlying, nonlogarithmic value of the dependent variable by approximately antilog(b), where antilog(b) is the number whose logarithm is b. For example, in Table III the regression coefficient of D1, a dummy variable denoting families with a head but no spouse, is -0.50. The antilog of -0.50 is 0.61. Thus, the regression coefficient of -0.50 means that families with a head and no spouse have, other things equal, a financial burden index about 0.61 times as large as families with other head-spouse structures (or a financial burden index 39 percent smaller than families with other head-spouse structures). Table H, which interprets Table III, presents this finding.
2. Second are continuous independent variables used in logarithmic form in the estimating equation. For such variables, the regression coefficient, b, is the elasticity. This means that each 1-percent increase in the underlying, nonlogarithmic independent variable is associated with approximately a b-percent increase in the underlying, nonlogarithmic dependent variable. For example, I30, the natural logarithm of a family's annual income, is a continuous independent variable used in logarithmic form in Table III. Its regression coefficient there is -0.56. This means that each 1-percent increase in annual family income (the underlying, nonlogarithmic independent variable) is, other things equal, associated with approximately a 0.56 percent *decrease* (a 0.56 percent decrease is a -0.56 percent increase) in the level of the financial burden index (the underlying, nonlogarithmic dependent variable). Again, Table H, which interprets Table III, presents this finding.
3. Finally, there are continuous independent variables used in nontransformed (that is, nonlogarithmic) form. If the regression coefficient for such a variable is b, then each increase of one unit in the independent variable is associated with a multiplication of the underlying, nonlogarithmic form of the dependent variable by approximately antilog(b). Again, Tables III and H can serve to illustrate this point. The age of the family head in years, D6, is a continuous independent variable used in nontransformed form in the regression equation whose results are presented in Table III. Its regression coefficient in that equation is found to be 0.025. The antilog of 0.025 is 1.025. Hence, the interpretation of the regression result is that each increase of one year (the unit in which

age is measured) in the age of a family's head is, other things equal, associated with a multiplication of the level of the family's financial burden index by approximately 1.025, which is an increase of 2.5 percent. Again, Table H, which interprets the results reported in Table III, presents this finding.

These interpretations of regression coefficients can be demonstrated by suitable manipulation of equation (14) (or of equation (14a), which is mathematically equivalent). However, in using these interpretations of regression coefficients, it should be noted that antilog($\hat{b}$), the antilog of the estimated regression coefficient, is a biased estimator of the antilog of the regression coefficient, although statistical significance tests associated with the coefficients are sound. Thus, the regressions correctly indicate which variables are statistically significant. If extensive estimation using antilogs is to be carried out, corrections for the bias are available.

Interpreting Means of Variables

Because of the mixture of dummy variables, untransformed variables, and logarithmic variables in the regressions, readers may find helpful the following information on interpreting the means of variables. (Means of the variables used in the regressions are shown in Tables II, III, and IV.)

The mean value of a dummy variable is the proportion of the population that has the characteristic denoted by the variable. For example, Table III shows that the mean of D1, the dummy variable denoting families with a head but no spouse, is 0.42 for the U.S. population of younger, lower-income multiple-person families included in this table. This means that 42 percent of these families have a head-only family structure.

The mean value of an untransformed continuous variable is simply the (familiar) arithmetic mean of the variable for the population in question. For example, the mean of D6 in Table III is 38.1. This variable measures the age of the family head in years and thus shows that the (arithmetic) mean age of the family head for the U.S. population of younger, lower-income multiple-person families included in the table is 38.1 years.

The mean value of a logarithmic continuous variable is the logarithm of the geometric mean of the variable. Taking the antilogarithm of the mean does not give the arithmetic mean of the untransformed variable. The geometric mean often differs very substantially from the arithmetic mean and should not be confused with it.

Analytic Procedures Used

Introduction

This section describes the analytic procedures used in the regression analyses in this report. Several steps were involved. These were weighting and standardizing the data; selecting the initial variable set; finding a core set of variables through stepwise regression; choosing criteria for evaluating the statistical significance of variables; and estimating the statistical significance of the core variables with SURREGR (Holt and Shah, 1982), a computer program that takes account of the complex sample design of the NMCUES. These steps are described in turn in this section.

Weighting and Standardizing the Data

Before regression analysis (or other analysis) of the data could begin, certain weighting and standardizing procedures had to be carried out. These procedures are described in more detail in Appendix II, but a summary of them is included here in order to present in sequence the procedures followed in the regression analysis.

Weighting of each case (family) in the data set began with a weight that previous reports on NMCUES family data have called FWEIGHT. Described simply, FWEIGHT is the reciprocal of the sampling probability adjusted for undercoverage and nonresponse and smoothed to agree with population totals from the March 1980 Current Population Survey. For each case (family), FWEIGHT was multiplied by the proportion of the survey year (calendar year 1980) that the family was eligible for the survey. This time-adjusted weight, called AWEIGHT, is the weight used in the regression analyses and in other analyses in this report. AWEIGHT differs from FWEIGHT only for families not in the sample for a full year.

For these families, standardization of data on income, health spending, health care use, and other variables that measure rates was carried out. Data on these variables covering the period the family was in the sample were divided by the proportion of the year the family was in the sample in order to derive an annualized rate. For example, a family in the sample for half the year with $10,000 of income and $150 of out-of-pocket health expenses recorded during this half year had its annualized income recorded as $20,000 and its annualized out-of-pocket health expenses recorded as $300. Annualized statistics, like these, were used in the regression analyses as the measure of all variables involving rates.

Selecting the Initial Variable Set

Regressions were run separately for three categories of multiple-person families:

- Older families, those with one or more members age 65 or older.
- Younger, lower-income families, those with no member 65 or older and with income below 200 percent of the poverty level.
- Younger, better-off families, those with no member 65 or older and with income of 200 percent of the poverty level or more.

For each family category, a small number of the 47 independent variables shown in Table I were omitted from the initial regressions because they were not applicable or relevant. Thus, the initial regressions involved the financial burden index as the dependent variable and slightly fewer than 47 independent variables. The omitted independent variables and the reasons for their omission are as follows. First, one dummy variable for source of health care coverage was omitted for each of the three populations, for the reason described above in the section "Functional Form of the Regression Equation." The omitted variable was that representing the most common source of coverage—medicare plus private insurance, I38, for older families; and private insurance only, I34, for both younger family populations. The dummy variable identifying families with all members age 65 or older, D7, was omitted from the regressions for both younger family categories because such families, by definition, have no members age 65 or older. The variable measuring work days lost due to illness, H29, was omitted from the regression for older families because many such families have no working members, which makes this variable meaningless for them. The dummy variable for families with health care coverage entirely from "other public" sources, I40, was omitted from the regression for older families because none of the older families in the sample had such coverage. The dummy variable for families with all members having an unknown perceived health status rating, H25, was omitted from the regressions for older families and for younger, lower-income families because no families in the sample in these two categories had this rating. Thus there were 43 independent variables in the initial regressions for older families, 44 independent variables in the initial regressions for younger, lower-income families, and 45 independent variables in the initial regressions for younger, better-off families.

Identifying a Core Set of Variables Through Stepwise Regression

For each of the three family categories, stepwise regression was used to select a preferred subset of independent variables from among the original 43 to 45 independent variables. The stepwise regression was carried out using PC SAS (SAS Institute, 1985).

Stepwise regression was used for two reasons. For one, a number of factors were operationalized by multiple variables. For example, family health status was operationalized by four sets of variables: (1) family bed days due to illness, (2) family work-loss days due to illness, (3) the excellent-good-fair-poor scale of reported health status, and (4) the limitations in main activity scale. Because of multicollinearity, using multiple variables that operationalize the same concept in a regression equation often yields distorted regression coefficients and large standard errors that indicate none of the variables is significant. Stepwise regression generally selects out a subset of the variables operationalizing a given concept, thus avoiding severe multicollinearity problems.

Second, as more variables are entered into a regression equation, there is a reduction in the precision with which the effects of any one variable can be identified. Standard errors increase, which tends to reduce the number of variables identified as significant. Stepwise regression permits a trade-off between the additional explanatory power obtained by adding more variables to a regression and the loss of precision in identifying the effects of any one of them.

The preferred independent variable set was defined by the step of the stepwise regression that had the lowest value of $C(p)$. (For $C(p)$, see Mallows, 1973.) The variable entered in the step at which $C(p)$ reached a minimum and all variables entered in preceding steps were included in the preferred variable set. All other variables were excluded. However, if $C(p)$ was still decreasing when all independent variables that entered the stepwise regression with probabilities less than 0.20 had been added, then the last step before the entry probability for a variable exceeded 0.20 was chosen as defining the preferred variable set. This step was used to define the preferred variable set in the same fashion that $C(p)$ was otherwise used to define it. Finally, all preferred variable sets were required to contain the variables I30 (natural logarithm of annual family income), H13 (indicating whether or not any family member was hospitalized), and D5 (natural logarithm of average family size). If any of these three variables was missing from the preferred variable set developed by the stepwise regression, the missing variable(s) was added and the variable set that included all three of these variables was considered the preferred variable set. Income and hospitalization were included because the literature has found them very important. Family size was included in order to try to assure that effects of family size were distinguished from effects of family structure variables. (The family structure variables used were whether a family had children, whether it had a head and a spouse or only a head, and whether its composition was stable.)

The PC SAS stepwise regression program weights each case (family) in the regression, but does not take account of the complex sample design of NMCUES. That is, it estimates variances according to equation (11), the formula appropriate for noncomplex samples, while equation (13), which is more involved, is the appropriate formula to use in estimating variances of NMCUES data. There appears to be no stepwise regression program available that takes complex sample design into consideration in estimating variances.

Estimating Statistical Significance Using SURREGR

Because the stepwise regression procedure in PC SAS does not take account of complex sample design in estimating variances, its estimates of statistical significance can involve large errors when it is used in analysis

of data from surveys, such as the NMCUES, which have complex sample designs. Therefore, the next steps in the regression analysis procedure involved the use of a regression software program that does estimate variances of complex samples appropriately, using equation (13). The program used was SURREGR (Holt and Shah, 1982), which runs within the SAS system. Nonstepwise SURREGR regressions were run on the preferred variable sets, and it is the results of these regressions that are shown in Tables II, III, and IV.

It should be noted that identical estimates of regression coefficients and of R-square (the proportion of the variance in the dependent variable explained by the independent variables) are produced by the PC SAS stepwise regression procedure and the SURREGR regression procedure. The differences of concern between the two procedures are in the statistical significance levels that they report for independent variables.

In the regressions shown in Tables II, III, and IV, the probability that the regression coefficient associated with any variable in the regressions was different from zero was computed by SURREGR using the F statistic. This probability is shown in the last column in each table. A regression coefficient was considered significant if its probability of occurring by chance was less than 0.05.

However, because there are 17 to 24 regression coefficients (one for each independent variable) in the preferred variable sets, a simple use of a 0.05 probability test would not be appropriate. Approximately one coefficient meeting a simple 0.05 probability test would be expected for every 20 regression coefficients, and thus approximately one such coefficient would be expected in each of Tables II, III, and IV purely by chance.

The significance test actually used was that the probability associated with any one variable had to be less than $0.05 \div n$, where n is the number of independent variables in the preferred variable set. This test is analogous to the multiple t-test used in frequency tables in this report (see Levy and Lemeshow, 1980, p. 296) and in previous reports on NMCUES family data (Dicker and Sunshine, 1987; Sunshine and Dicker, 1987a; Sunshine and Dicker, 1987b). The actual probability corresponding to $0.05 \div n$ was as follows for the three populations of multiple-person families studied:

Population	*Probability*
Older families (Table II)	0.0023
Younger, lower-income families (Table III)	0.0029
Younger, better-off families (Table IV)	0.0021

Variables with probabilities below these levels were considered significant and appear in Tables G, H, and J, which report significant findings and accompany the text discussions of findings for each of the three family populations.

A Further Check Using SURREGR

The preferred variable sets should include all statistically significant independent variables, for stepwise regression selects the independent variables with the greatest statistical significance first, and then moves to progressively less significant variables. However, in light of possible problems arising because the PC SAS stepwise procedure does not compute variances (and hence significance levels) based on the NMCUES complex sample design, a check for omitted significant variables was performed. For each of the three family populations, the full regression model, with all 43 to 45 independent variables, was run using SURREGR. The results are shown in Tables V, VI, and VII. These results were generally as expected, and no statistically significant variables were found that were not also statistically significant in the (smaller) preferred variable sets.

Table I

Initial set of variables used in the stepwise regression

Variable type	Variable indicator	Description of variable
Dependent variable		
The index of financially burdensome family health expenses (continuous)	Y1	The ratio of annualized total out-of-pocket health expenses to annualized family income, transformed into its natural logarithm.[1]
Independent variables		
Demographic and social		
Head-spouse structure of family (3 categories, 2 dummy variables)[2]		
Head-only family	D1	1 = Family had a head only (no spouse) during time in survey; 0 = All other head and spouse combinations.
Head-spouse change[3]	D2	1 = Family had an unstable head-spouse structure during time in survey; 0 = All families with stable head only or stable head-and-spouse.

See footnotes at end of table.

Table I—Continued

Initial set of variables used in the stepwise regression

Variable type	Variable indicator	Description of variable
Dynamic-static nature of family (3 categories, 2 dummy variables)[4]		
Head-spouse change[3]	D2	1 = Family had an unstable head-spouse structure during time in survey; 0 = All families with stable head only or stable head-and-spouse.
Other change	D3	1 = Family had a stable head-spouse structure during time in survey but other family member changed, or family did not exist full year; 0 = Other family change status.
Presence of children in family (2 categories, 1 dummy variable)	D4	1 = Family had a member 16 years of age or younger; 0 = All family members 17 years of age or older.
Family size (continuous)	D5	Average family size (in persons) during time in survey, transformed into its natural logarithm.
Age of head of the family (continuous)	D6	Age of the head of the family in years, as of January 1, 1980.
Age of family members (2 categories, 1 dummy variable)[5]	D7	1 = All family members are 65 years of age or over; 0 = Some or all family members are less than 65 years of age.
Sex of the head of the family (2 categories, 1 dummy variable)	D8	1 = Female head of family; 0 = Male head of family.
Race of head of the family (3 categories, 2 dummy variables)[6]		
Black	D9	1 = Black head of family; 0 = Head of family of other race.
Other	D10	1 = Other (neither black nor white) head of family; 0 = Head of family either black or white.
Ethnicity of the head of the family (2 categories, 1 dummy variable)	D11	1 = Hispanic head of family; 0 = Head of family of other ethnicity.
Education of head of the family (continuous)	D12	Formal education of the head of the family in years of education (18 was the highest value used).
Health related		
Hospitalization of a family member (2 categories, 1 dummy variable)	H13	1 = Family had one or more members discharged from a hospital during its time in the survey; 0 = No family members discharged from a hospital.
Total number of hospital discharges (continuous)	H14	Annual rate of hospital discharges for all family members.
Institutionalization of a family member (2 categories, 1 dummy variable)	H15	1 = Family had one or more members institutionalized during its time in the survey or, if it did not continue until the end of 1980, at its termination; 0 = No family members were institutionalized.
Death of a family member (2 categories, 1 dummy variable)	H16	1 = Family had one or more members die during its time in the survey or, if it did not continue until the end of 1980, at its termination; 0 = No family member died.
Birth of a family member (2 categories, 1 dummy variable)	H17	1 = Family had one or more members who gave birth to a child during its time in the survey; 0 = No family member gave birth to a child.
Illness in a family member (4 dummy variables)[7]		
Cancer and other neoplasms[7]	H18	1 = Family had one or more members with some type of neoplasm during its time in the survey; 0 = No family member had a neoplasm.
Circulatory and heart disease[7]	H19	1 = Family had one or more members with some type of circulatory or heart disease during its time in the survey; 0 = No family member had circulatory or heart disease.
Accidents, injuries, and poisonings[7]	H20	1 = Family had one or more members with some type of accident, injury, or poisoning during its time in the survey; 0 = No family member had an accident, injury, or poisoning.

See footnotes at end of table.

Table I—Continued

Initial set of variables used in the stepwise regression

Variable type	Variable indicator	Description of variable
Other illnesses[7]	H21	1 = Family members had none of the above illnesses, but one or more members had some other illness during his or her time in the survey; 0 = One or more family members had one of the above illnesses or all family members had no illness.
Perceived health status rating of family (5 categories, 4 dummy variables)[8]		
Good	H22	1 = Worst perceived health status of any family member was reported as "good"; 0 = All family members were reported in excellent health or some family members were reported in fair or poor health.
Fair	H23	1 = Worst perceived health status of any family member was reported to be "fair"; 0 = All family members were reported in excellent or good health or some member was reported in poor health.
Poor	H24	1 = Worst perceived health status of any family member was reported to be "poor"; 0 = No family member had a "poor" rating.
Unknown	H25	1 = Reported health status of all family members is "unknown"; 0 = Reported health status of at least some family members is known.
Limitation in usual activity rating of family (3 categories, 2 dummy variables)[9]		
Secondary limitation	H26	1 = Most severe limitation reported for any family member was either a limitation in secondary activity or a limitation in amount or kind of main activity (work, housekeeping, school, and so on); 0 = No family member was reported to have a limitation or a major limitation was reported for one or more family members.
Major limitation	H27	1 = Most severe limitation reported for any family member was inability to perform a usual major activity (work, housekeeping, school, and so on); 0 = No family member was reported as unable to perform his or her usual major activity.
Family illness days in bed (continuous)	H28	Annual rate of total illness days spent in bed for all family members. One day was added to the annual rate and the resulting statistic then transformed into its natural logarithm.
Family work-loss days (continuous)[10]	H29	Annual rate of total work-loss days due to illness. One day was added to the annual rate and the resulting statistic then transformed into its natural logarithm.
Income and insurance		
Family income (continuous)	I30	Annualized family income in dollars transformed into its natural logarithm.
Completeness of health care coverage (4 categories, 3 dummy variables)[11]		
Partial coverage 1	I31	1 = All family members covered but some or all only part year; 0 = All other types of coverage.
Partial coverage 2	I32	1 = Some family members had coverage, but some had no coverage; 0 = All other types of coverage.
No coverage	I33	1 = No family member covered; 0 = Partial or full coverage.
Source of health care coverage (8 categories, 7 dummy variables used for each population group)[12]		
Private insurance[12]	I34	1 = Family members only had coverage from private health insurance; 0 = All other sources of coverage.
Medicaid	I35	1 = Family members only had coverage from medicaid; 0 = All other sources of coverage.
Medicare	I36	1 = Family members only had coverage from medicare; 0 = All other sources of coverage.
Medicare and other public	I37	1 = Family members only had coverage from medicare and other public programs; 0 = All other sources of coverage.

See footnotes at end of table.

Table I—Continued

Initial set of variables used in the stepwise regression

Variable type	Variable indicator	Description of variable
Medicare and private[12]	I38	1 = Family members only had coverage from medicare and private health insurance; 0 = All other sources of coverage.
Other public and private	I39	1 = Family members had coverage from both (1) public sources other than medicare and (2) private insurance; 0 = All other sources of coverage.
Other public	I40	1 = Family members only had coverage from public source(s) other than those listed above; 0 = All other sources of coverage.
Unknown	I41	1 = Every family member either had no coverage or had coverage from sources not identified; 0 = One or more family members had coverage from identified sources.
Geographic		
Region of United States (4 categories, 3 dummy variables)[13]		
North Central	G42	1 = Head of family resided in North Central census region; 0 = Head resided in other region of U.S.
South	G43	1 = Head of family resided in South census region; 0 = Head resided in other region of U.S.
West	G44	1 = Head of family resided in West census region; 0 = Head resided in other region of U.S.
Urban-rural location (4 categories, 3 dummy variables)[14]		
Metropolitan suburb	G45	1 = Head of family resided in a suburb of a metropolitan statistical area; 0 = Head resided in another location.
Nonmetropolitan urban area	G46	1 = Head of family resided in an urban area that was not a part of a metropolitan statistical area; 0 = Head resided in another location.
Nonurban area	G47	1 = Head of family resided in non-urban area; 0 = Head resided in another location.

[1]Total out-of-pocket health expenses is the sum of out-of-pocket expenses for health care services plus family-paid premiums (see discussion in text).
[2]Omitted category is families with both a head and a spouse during time in survey.
[3]The variable is entered only once in the regression, but functions both as a measure of head-spouse structure and as a measure of dynamic-static nature of the family.
[4]Omitted category is static families—that is, families that had no change in membership and were in the survey the full survey year.
[5]This variable was only used in regressions for older families (families with a member 65 years of age or over). When used with this population, the "0" category designates families with members both over and under 65 years of age.
[6]Omitted category is families with a white head of family.
[7]Omitted category is families in which no family member reported an illness. The dummy variables for cancer; heart and circulatory disease; and accidents, illnesses, and poisonings are not mutually exclusive.
[8]Omitted category is all family members reported to be in excellent health.
[9]Omitted category is families in which no family member reported a limitation. Family members for whom limitation status was unknown were coded as having no limitation.
[10]This variable was not used in regressions for older families (families with a member 65 years of age or over).
[11]Omitted category is families in which all family members had full-year coverage by private health insurance and/or public health care coverage program(s).
[12]For regressions for older families, the omitted category is coverage by both Medicare and private insurance (I38). For regressions for younger families, the omitted category is coverage from private health insurance only (I34).
[13]Omitted category is residence in the Northeast census region.
[14]Omitted category is residence in the central city of a metropolitan statistical area.

NOTE: Further information on the variables in this table may be found either in Appendix III, "Definition of Terms," or in the text.

Table II

SURREGR regression for the index of financially burdensome family health expenses for older families

Independent variable	Mean	Regression coefficient	Standard error of regression coefficient	F Value	Probability
D1 Head only	0.256	−0.384	0.076	25.32	0.0000
D4 Children	0.108	0.288	0.164	3.08	0.0836
D5 Family size	0.865	0.005	.	0.00	0.9731
D6 Age of head	68.089	0.007	0.003	5.27	0.0248
D9 Black race	0.093	−0.431	0.152	7.99	0.0062
D10 "Other" race	0.020	−0.331	0.240	1.89	0.1735
D12 Education of head	9.840	0.020	0.010	4.39	0.0399
H13 Hospitalization	0.388	0.206	0.094	4.86	0.0309
H14 Discharges	0.711	0.076	0.034	4.90	0.0302
H15 Institutionalization	0.018	0.551	0.217	6.44	0.0134
H19 Heart disease, etc.	0.724	0.218	0.067	10.68	0.0017
H20 Accidents, etc.	0.276	0.134	0.066	4.04	0.0483
H26 Secondary limitation	0.080	0.156	0.081	3.70	0.0587
H28 Bed days	1.873	0.060	0.025	5.86	0.0182
I30 Income	9.587	−0.853	0.053	262.58	0.0000
I32 Partial coverage 2	0.108	−0.169	0.125	1.85	0.1781
I35 Medicaid coverage only	0.004	−1.119	0.317	12.43	0.0008
I36 Medicare coverage only	0.095	−0.243	0.079	9.58	0.0028
I37 Medicare and other public coverage	0.053	−2.162	0.255	72.04	0.0000
I41 Coverage source unknown	0.030	−1.151	0.304	14.35	0.0003
G42 North Central region	0.235	0.109	0.067	2.66	0.1076
G43 South region	0.348	0.288	0.063	20.83	0.0000

Intercept = 4.267
Number of observations = 849
Multiple correlation coefficient squared (R^2) = 0.529
F = 54.97 with 22 degrees of freedom

Mean of dependent variable = −2.971
Probability = 0.0000
68 denominator degrees of freedom

NOTES: Older families are families with member(s) 65 years of age or over. A probability of 0.0023 was needed for statistical significance at the .05 level using an F test that is analogous to a multiple *t* test (see Appendix I). Standard error of regression blank because F test had a value of zero after rounding.

Table III

SURREGR regression for the index of financially burdensome family health expenses for younger, lower-income families

Independent variable	Mean	Regression coefficient	Standard error of regression coefficient	F Value	Probability
D1 Head only	0.422	−0.501	0.108	21.52	0.0000
D5 Family size	1.258	−0.298	0.150	3.93	0.0514
D6 Age of head	38.126	0.025	0.004	47.26	0.0000
D12 Education of head	10.802	0.049	0.016	9.91	0.0024
H13 Hospitalization	0.334	0.360	0.134	7.22	0.0090
H17 Birth	0.059	0.433	0.204	4.51	0.0374
H18 Cancer	0.082	0.344	0.177	3.80	0.0555
H20 Accidents, etc.	0.487	0.139	0.097	2.07	0.1551
H22 Good health	0.374	0.168	0.087	3.76	0.0566
H26 Secondary limitation	0.095	−0.213	0.205	1.08	0.3023
H29 Work-loss days	1.190	0.117	0.032	13.49	0.0005
I30 Income	9.104	−0.556	0.120	21.28	0.0000
I32 Partial coverage 2	0.145	0.218	0.128	2.89	0.0938
I35 Medicaid coverage only	0.117	−1.674	0.253	43.76	0.0000
I39 Other public and private coverage	0.230	−0.394	0.121	10.59	0.0018
I40 Other public coverage only	0.007	−1.972	0.490	16.16	0.0001
I41 Coverage source unknown	0.280	−0.564	0.112	25.39	0.0000

Intercept = 0.823
Number of observations = 920
Multiple correlation coefficient squared (R^2) = 0.267
F = 21.13 with 17 degrees of freedom

Mean of dependent variable = −3.342
Probability = 0.0000
69 denominator degrees of freedom

NOTES: Younger, lower-income families are families with no member 65 years of age or over and with incomes below 200 percent of the poverty level. A probability of 0.0029 was needed for statistical significance at the .05 level using an F test analogous to a multiple *t* test (see Appendix I).

Table IV

SURREGR regression for the index of financially burdensome family health expenses for younger, better-off families

Independent variable	Mean	Regression coefficient	Standard error of regression coefficient	F Value	Probability
D1 Head only	0.145	−0.136	0.079	2.94	0.0909
D3 Other change	0.167	−0.151	0.056	7.23	0.0090
D4 Children	0.593	0.135	0.060	5.02	0.0283
D5 Family size	1.135	0.241	0.084	8.20	0.0056
D6 Age of head	42.050	0.015	0.002	61.37	0.0000
D9 Black race	0.063	−0.335	0.100	11.27	0.0013
D12 Education of head	12.665	0.034	0.007	21.59	0.0000
H13 Hospitalization	0.268	0.256	0.072	12.68	0.0007
H14 Discharges	0.423	0.069	0.044	2.52	0.1168
H15 Institutionalization	0.001	−2.017	0.700	8.31	0.0053
H17 Birth	0.037	0.185	0.122	2.30	0.1337
H19 Heart disease, etc.	0.265	0.081	0.053	2.35	0.1302
H20 Accidents, etc.	0.432	0.079	0.037	4.61	0.0353
H23 Fair health	0.135	0.092	0.044	4.28	0.0423
H24 Poor health	0.053	0.243	0.077	10.04	0.0023
H28 Bed days	1.955	0.069	0.017	16.35	0.0001
I30 Income	10.262	−0.868	0.045	369.56	0.0000
I31 Partial coverage 1	0.141	0.074	0.050	2.22	0.1406
I33 No coverage	0.020	−0.529	0.110	22.96	0.0000
I35 Medicaid coverage only	0.004	−1.023	0.488	4.40	0.0395
I39 Other public & private coverage	0.135	−0.228	0.056	16.55	0.0001
I40 Other public coverage only	0.004	−0.893	0.129	47.84	0.0000
G43 South region	0.300	0.273	0.047	33.50	0.0000
G45 Metropolitan suburb	0.447	0.088	0.038	5.46	0.0223

Intercept = 3.197
Number of observations = 2,874
Multiple correlation coefficient squared (R^2) = 0.232
F = 52.92 with 24 degrees of freedom

Mean of dependent variable = −3.942
Probability = 0.0000
69 denominator degrees of freedom

NOTES: Younger, better-off families are families with no member 65 years of age or over and incomes 200 percent of the poverty level or higher. A probability of 0.0021 was needed for statistical significance at the .05 level using an F test analogous to a multiple *t* test (see Appendix I).

Table V

SURREGR regression for the index of financially burdensome family health expenses for older families—full model

Independent variable	Mean	Regression coefficient	Standard error of regression coefficient	F Value	Probability
D1 Head only	0.256	−0.318	0.111	8.21	0.0055
D2 Head-spouse change	0.037	−0.069	0.218	0.10	0.7526
D3 Other change	0.105	0.142	0.151	0.89	0.3492
D4 Children	0.108	0.313	0.171	3.34	0.0721
D5 Family size	0.865	−0.001	.	0.00	1.0000
D6 Age of head	68.089	0.006	0.003	3.55	0.0639
D7 All members 65+	0.385	0.045	0.073	0.38	0.5418
D8 Female family head	0.221	−0.070	0.128	0.30	0.5878
D9 Black race	0.093	−0.410	0.151	7.38	0.0083
D10 "Other" race	0.020	−0.326	0.256	1.62	0.2081
D11 Hispanic	0.031	−0.131	0.243	0.29	0.5943
D12 Education of head	9.840	0.019	0.009	4.20	0.0443
H13 Hospitalization	0.388	0.197	0.093	4.50	0.0375
H14 Discharges	0.711	0.076	0.032	5.60	0.0208
H15 Institutionalization	0.018	0.435	0.243	3.21	0.0778
H16 Death	0.048	−0.044	0.197	0.05	0.8225
H17 Birth	0.002	−0.908	0.492	3.40	0.0696
H18 Cancer	0.164	0.093	0.097	0.91	0.3437
H19 Heart disease, etc.	0.724	0.220	0.103	4.57	0.0362
H20 Accidents, etc.	0.276	0.123	0.065	3.55	0.0637
H21 Other illness	0.172	0.033	0.110	0.09	0.7640
H22 Good health	0.337	0.048	0.085	0.32	0.5746
H23 Fair health	0.287	−0.029	0.080	0.13	0.7178
H24 Poor health	0.242	0.008	0.080	0.01	0.9425
H26 Secondary limitation	0.080	0.115	0.086	1.78	0.1872
H27 Major limitation	0.533	−0.025	0.072	0.12	0.7279
H28 Bed days	1.873	0.066	0.025	7.04	0.0099
I30 Income	9.587	−0.856	0.052	271.54	0.0000
I31 Partial coverage 1	0.620	0.024	0.120	0.04	0.8350
I32 Partial coverage 2	0.108	−0.144	0.107	1.80	0.1847
I33 No coverage	0.007	−0.222	0.906	0.06	0.8125
I34 Private coverage only	0.031	−0.181	0.286	0.40	0.5280
I35 Medicaid coverage only	0.004	−1.117	0.364	9.44	0.0031
I36 Medicare coverage only	0.095	−0.256	0.080	10.25	0.0021
I37 Medicare & other public coverage	0.053	−2.152	0.252	72.67	0.0000
I39 Other public & private coverage	0.012	0.053	0.237	0.05	0.8200
I41 Coverage source unknown	0.030	−1.140	0.379	9.06	0.0037
G42 North Central region	0.235	0.103	0.065	2.50	0.1188
G43 South region	0.348	0.275	0.063	19.34	0.0000
G44 West region	0.210	0.009	0.090	0.01	0.9185
G45 Metropolitan suburb	0.357	0.069	0.068	1.04	0.3120
G46 Nonmetropolitan urban area	0.149	0.147	0.106	1.92	0.1709
G47 Nonurban area	0.190	0.107	0.098	1.19	0.2789

Intercept = 4.253
Number of observations = 849
Multiple correlation coefficient squared (R^2) = 0.537
F = 47.26 with 43 degrees of freedom

Mean of dependent variable = −2.971
Probability = 0.0000
68 denominator degrees of freedom

NOTE: Standard error of regression blank because F test had a value of zero after rounding.

Table VI

SURREGR regression for the index of financially burdensome family health expenses for younger, lower-income families—full model

Independent variable	Mean	Regression coefficient	Standard error of regression coefficient	F Value	Probability
D1 Head only	0.422	−0.255	0.206	1.53	0.2206
D2 Head-spouse change	0.038	0.011	.	0.00	0.9707
D3 Other change	0.203	−0.118	0.155	0.58	0.4481
D4 Children	0.807	0.079	0.177	0.20	0.6586
D5 Family size	1.258	−0.346	0.202	2.92	0.0918
D6 Age of head	38.126	0.024	0.005	25.32	0.0000
D8 Female family head	0.426	−0.234	0.219	1.14	0.2890
D9 Black race	0.201	−0.143	0.140	1.04	0.3102
D10 "Other" race	0.021	−0.396	0.349	1.29	0.2593
D11 Hispanic	0.101	0.126	0.175	0.52	0.4719
D12 Education of head	10.802	0.058	0.017	11.81	0.0010
H13 Hospitalization	0.334	0.203	0.230	0.78	0.3809
H14 Discharges	0.589	0.066	0.110	0.36	0.5510
H15 Institutionalization	0.007	−0.674	0.735	0.84	0.3622
H16 Death	0.007	−0.709	0.793	0.80	0.3733
H17 Birth	0.059	0.452	0.269	2.83	0.0970
H18 Cancer	0.082	0.396	0.198	3.98	0.0500
H19 Heart disease, etc.	0.264	0.200	0.188	1.13	0.2917
H20 Accidents, etc.	0.487	0.256	0.168	2.31	0.1333
H21 Other illness	0.336	0.162	0.223	0.53	0.4708
H22 Good health	0.374	0.234	0.110	4.50	0.0375
H23 Fair health	0.263	0.116	0.136	0.73	0.3970
H24 Poor health	0.141	0.126	0.257	0.24	0.6234
H26 Secondary limitation	0.095	−0.237	0.192	1.53	0.2210
H27 Major limitation	0.158	0.032	0.143	0.05	0.8250
H28 Bed days	2.296	0.035	0.048	0.54	0.4669
H29 Work-loss days	1.190	0.104	0.034	9.54	0.0029
I30 Income	9.104	−0.540	0.118	20.87	0.0000
I31 Partial coverage 1	0.242	0.003	.	0.00	0.9859
I32 Partial coverage 2	0.145	0.243	0.174	1.94	0.1684
I33 No coverage	0.092	−0.012	.	0.00	0.9594
I35 Medicaid coverage only	0.117	−1.598	0.267	35.76	0.0000
I36 Medicare coverage only	0.015	−0.053	0.265	0.04	0.8395
I37 Medicare & other public coverage	0.001	−0.198	0.326	0.37	0.5462
I38 Medicare & private coverage	0.004	−0.170	0.221	0.59	0.4442
I39 Other public & private coverage	0.230	−0.376	0.111	11.42	0.0012
I40 Other public coverage only	0.007	−1.914	0.532	12.93	0.0006
I41 Coverage source unknown	0.280	−0.554	0.172	10.38	0.0019
G42 North Central region	0.234	0.102	0.114	0.80	0.3753
G43 South region	0.349	0.230	0.110	4.35	0.0407
G44 West region	0.221	0.039	0.138	0.08	0.7824
G45 Metropolitan suburb	0.328	0.132	0.131	1.02	0.3162
G46 Nonmetropolitan urban area	0.144	0.167	0.153	1.19	0.2801
G47 Nonurban area	0.195	0.064	0.133	0.23	0.6357

Intercept = 0.177
Number of observations = 920
Multiple correlation coefficient squared (R^2) = 0.282
F = 48.03 with 44 degrees of freedom

Mean of dependent variable = −3.342
Probability = 0.0000
69 denominator degrees of freedom

NOTE: Standard error of regression blank because F test had a value of zero after rounding.

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 person-level analysis. However, data for nonkey persons are used in family analysis because nonkey persons contributed to the family's utilization of and expenditures for health care during the time they were part of the family.

Persons included in the sample were grouped into "reporting units" for data collection purposes. Reporting units were defined as all persons related to each other by blood, marriage, adoption, or foster care status and living in the same dwelling unit. The combined NMCUES sample consisted of 7,244 eligible reporting units, of which 6,599 agreed to participate in the survey. In total, data were obtained on 17,123 key persons. The Research Triangle Institute sample yielded 8,326 key persons, and the National Opinion Research Center sample yielded 8,797.

Research Triangle Institute Sample Design

A primary sampling unit (PSU) is 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 disjoint PSU's exhaust the land area of the 50 States and Washington, D.C. The PSU's are classified as one of two types. The 16 largest standard metropolitan statistical areas (SMSA's) are designated as self-representing PSU's, and the remaining 1,670 PSU's in the primary sampling frame are designated as non-self-representing PSU's.

PSU's are 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 had sufficient population in 1970 to be treated as primary strata. The 1,659 non-self-representing PSU's from the continental United States were stratified into 59 primary strata with approximately equal populations. Each of these primary strata had a 1970 population of about 3⅓ million. One supplementary primary stratum of 11 PSU's, with a 1970 population of about 1 million, was added to the Research Triangle Institute primary frame to include Alaska and Hawaii.

The total first-stage sample for Research Triangle Institute consisted of 59 PSU's, of which 16 were self-representing PSU's. The non-self-representing PSU's were obtained by selecting one PSU from each of the 43 non-self-representing primary strata. These PSU's were selected with probability proportional to 1970 population size.

In each of the 59 sample PSU's, the entire PSU was divided into smaller disjoint area units called secondary sampling units (SSU's). Each SSU consisted of one or more enumeration districts or block groups defined by the 1970 census. Within each PSU, SSU's were ordered and then partitioned to form secondary strata of approximately equal size. Two secondary strata were formed in the non-self-representing PSU drawn from Alaska and Hawaii, and four secondary strata were formed in each of the remaining 42 non-self-representing PSU's. Thus, the non-self-representing PSU's were partitioned into a total of 170 secondary strata. In a similar manner, the 16 self-representing PSU's were partitioned into 144 secondary strata.

In the second stage of selection, 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 SSU's total noninstitutionalized population. The total number of sample SSU's was $2 \times 170 + 144 = 484$.

For the third stage of selection, each SSU was first divided into smaller disjoint geographic areas, and one area within the SSU was selected with probability proportional to the total number of housing units in 1970. Next, one or more disjoint segments of at least 60 housing units were formed in the selected area. One segment was selected from each SSU with probability proportional to the segment housing unit count. In response to the sponsoring agencies' request that the expected household sample size be reduced, a systematic sample of one-sixth of the segments was deleted from the sample. Thus, the total third-stage sample was reduced to 404 segments.

For the fourth stage of selection, 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 overall probability of selection. All of the reporting units within the selected dwelling units were included in the sample.

National Opinion Research Center Sample Design

The land area of the 50 States and Washington, D.C., was also divided into disjoint PSU's for the National Opinion Research Center sample design. 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. 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 approximately 1 million.

A single PSU was selected within each zone with a probability proportional to its 1970 population. It should be noted that this procedure allowed a PSU to be selected more than one time. For instance, an SMSA primary sampling unit with a population of 3 million could be selected as many as four times. The full general-purpose sample contained 204 PSU's. These 204 PSU's

were systematically allocated to four subsamples of 51 PSU's. The final set of 76 sample PSU's was chosen by randomly selecting two complete subsamples of 51 PSU's. One subsample was included in its entirety, and 25 of the PSU's in the other subsample were selected systematically for inclusion in NMCUES.

For the second stage, each PSU selected in the first stage was partitioned into a disjoint set of SSU's defined by block groups, enumeration districts, or a combination of the two types of census units. Within each sample PSU, the SSU's were ordered and then partitioned into 18 zones such that each zone contained approximately the same number of households. 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, the second-stage selection process was repeated as many times as there were first-stage hits. The 405 SSU's were identified by selecting 5 SSU's from each of the 51 PSU's in the subsample that was included in its entirety and 6 SSU's from each of the 25 PSU's in the group for which only one-half of the PSU's were included.

The SSU's selected in the second stage were then subdivided into area segments with a minimum size of 100 housing units each. One segment was then selected with probability proportional to the estimated number of housing units. The final-stage sample, in which a selection of housing units was made, was essentially the same as that used by the Research Triangle Institute.

Collection of Data

Field operations for NMCUES were performed by the Research Triangle Institute and the National Opinion Research Center under specifications established by the sponsoring 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 rounds of interviews to collect data on health, health care, health care charges, sources of payment, and health insurance 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 rounds of interviews to collect data that were not expected to change during the year or that were needed only once. Approximately 80 percent of the third and fourth rounds of 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 older. A proxy respondent not residing in the household was permitted only if all eligible household members were unable to respond because of health, language, or mental condition.

Imputation

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). In general, response rates for NMCUES were excellent. Approximately 90 percent of the sample reporting units agreed to participate in the survey, and approximately 94 percent of the individuals in the participating reporting units supplied complete annual information. Even though the overall response rates are quite high for NMCUES, the estimates of means and proportions may be biased if nonrespondents have different health care experiences than respondents or if there is a substantial response rate differential across subgroups of the target population. Furthermore, totals will tend to be underestimated unless allowance is made for the loss of data because of nonresponse.

Two methods commonly used to compensate for survey nonresponse are data imputation and the adjustment of sampling weights. For NMCUES, 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 is discussed in the section on sampling weights.

A specialized form of the sequential hot-deck imputation method was used for attrition imputation. First, each sample person with incomplete annual data (recipient) was linked to a sample person with similar demographic and socioeconomic characteristics who had complete annual data (donor). Second, the time periods for which the recipient had missing data were divided into two categories, imputed eligible days and imputed ineligible days. Imputed eligible days were those days for which the donor was eligible (that is, in scope), and imputed ineligible days were those days for which the donor was ineligible (that is, out of scope). For the recipient's imputed eligible days, the donor's medical care experiences (such as medical provider visits, dental visits, or hospital stays) were imputed into the recipient's record. 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, then the person was considered to be a total nonrespondent, and all data for the person were removed from the analytic data file.

The data collection methodology and field quality control procedures for NMCUES were designed so that the data would be as accurate and complete as possible subject to budget considerations. However, individuals

cannot report data that are unknown to them, or they may choose not to report the data even if known. This latter situation is especially true for data relating to expenditures, income, and other sensitive topics. Because of the size and complexity of the NMCUES data base, it was not feasible, from the standpoint of cost, to replace all missing data for all data items. The 12-month data files, for example, contain approximately 1,400 data items per person. With this in mind, the NMCUES approach was to designate a subset of the total items on the data base for imputation of the missing data. Thus, for 5 percent of the NMCUES data items, the responses were edited and missing data imputed by a combination of logic and hot-deck procedures to produce revised variables for use in analysis. Items for which imputations were made cover the following data areas.

- Visit charges.
- Source of payment codes and amounts.
- Annual disability days.
- Health insurance premium amount.
- Length of hospital stay.
- Total weeks worked in 1980.
- Average hours worked per week.
- Educational level.
- Hispanic ethnicity.
- Income.
- Age and birth date.
- Race.
- Sex.
- Health insurance coverage.
- Visit dates.

These items were selected as the most important variables for statistical analyses.

Construction of Longitudinal Families

At the time of the initial interview, a group of persons sharing a common housing unit was designated a family if they were related to each other by blood, marriage, adoption, or a formal foster care relationship. An unmarried student 17–22 years of age living away from home was also considered a part of the family, even though his or her residence was in a different location. When, on subsequent interviews, this initial sampled social unit was found to have had changes in membership, it became necessary to find a decision rule (or set of decision rules) for deciding when a family continued, when it ended, and when a new family began.

The decision rule chosen was initially referred to as a principal-predecessor-principal-successor rule (Dicker and Casady, 1982; Whitmore, Cox, and Folsom, 1982; Moser et al., 1983). The term came from the understanding that, at any given point in time, a family may have several predecessor families from which its members came and several successor families into which its members would go. The decisionmaking problem, therefore, was to objectively select only one predecessor family (the principal predecessor) and only one successor family (the principal successor) as representing the family through successive stages in time. If no principal successor family could be found, the initial family had ended. If no principal predecessor family could be found, the current family (at the time of the interview) was a new family. Later discussions in the literature referred to the above rule under a different name. It came to be called a "reciprocal, majority population rule" (McMillen, 1984; Dicker, 1984) because the principal-predecessor-principal-successor rule came to be understood as a rule that linked families on the basis of cross-family majorities. Thus, if two families (as defined above) exist at different but adjacent points in time, they are the same family if and only if a majority of the eligible members of the first family are found in the second family and a majority of the eligible members of the second family are also found in the first family. The reciprocity of the comparison is crucial. A unidirectional majority—either from the first family to the second family or from the second to the first—is not sufficient for the two families to be defined as the same.

Several aspects of the rule as applied in this survey need further elaboration. First, the rule was applied to all families in the longitudinal universe (not only to those in the initial sample) that had cross-membership connections with initially sampled families. Second, only persons eligible over time to be in both families being compared were counted when calculating cross-family majorities. For example, persons in family 1 who died or otherwise left the universe were not eligible for membership in family 2 and were not counted. Likewise, persons who entered family 2 from outside the universe during the interval between interviews, such as a newborn baby or a soldier returning to civilian status, could not have been in family 1 (that is, were not eligible for inclusion in that family) and also were not counted. Third, the reciprocal majority population rule, as stated above, links only two families adjacent in time. However, transitivity between linkages is implied in the rule. This means that given three families (families A, B, and C) existing at three different points in time, if family A is the same as family B and family B is the same as family C, then family A is also the same as family C. A longitudinal family, therefore, is either one or a series of point-interval families linked by the reciprocal majority population rule. Fourth, the final sample of families was limited to initially sampled families and all other families derived from these families that had at least one initially sampled person (a key individual) in them on their beginning date. Thus, the collection of families examined for family construction purposes was divided into key families (a family with

a key individual), which were in the sample and given a positive sampling weight, and nonkey families (a family without a key individual), which were not in the sample and given a sampling weight of zero. One reason for not including nonkey families in the sample is that very little data for them were available. Moreover, assumptions were often required to construct these families. (For more details on this methodology, see Dicker and Casady, 1982, and Whitmore, Cox, and Folsom, 1982.)

The dynamic sample of longitudinal families derived from this process tended to have characteristics that are generally sociologically believed to define the beginning and ending of families. For example, an even merger of two individuals through marriage always produced a new family. Similarly, an even split in a two-person family as the result of divorce or separation always ended the family. On the other hand, an uneven split in a larger family would not necessarily end such a family. In most cases, the original family continued as the larger part of the split. For example, if an adult child left a family of three persons or more to set up a separate household, in most cases the original family continued as the same but smaller family. Such an outcome appears to be in agreement with the sociological consensus that the loss of a single family member, other than the head or spouse, does not usually end the original family. The majority of uneven splits arise from this type of situation.

By the same reciprocal majority rule, however, a separation of husband and wife in a situation where children remained with one of the spouses in most cases continued the old family, now reconstituted as a single-spouse family with children. This result may not appear to be the sociologically preferred one. However, a more detailed review of the class of events of which this is a special case suggests that this result is in line both with the results based on sampling criteria for other members of the class and with sociological expectations of what the result should be for those class members. For example, given a head-spouse family with children, the loss of a head or spouse because of death or institutionalization is rarely thought of sociologically as an event ending the family. Rather, the social consensus appears to be that the original family continues, although in a recognizably changed state. The same may be said for the situation in which a head or spouse enters the military or goes overseas and is absent from the family for long periods. The family is not defined as ended but as continuing with an absent spouse. In this survey, all of the above events are defined as out-of-scope sampling events that cannot affect the identity of the family over time. Therefore, families would not end because of their occurrence. Only when the separating head or spouse remains within the noninstitutionalized U.S. population (the universe of inference) does the dilemma arise from sampling and sociological considerations as to whether the original family has ended. This inscope event, however, is similar in its effect on family functioning as the four previously mentioned out-of-scope events. In all of these situations, the family loses a significant role player. As a consequence, important family role obligations go unfulfilled (or only partially fulfilled). It seemed appropriate, therefore, to treat all of these events in the same manner (as a functionally equivalent happening) for the purpose of constucting longitudinal families. Given the lack of a sociological consensus for treating the above class of events, the reciprocal majority population rule produces an appropriate, if not consensual, decision. When the separating head or spouse or adult child remains within the universe, the reciprocal majority population rule must also be applied to find out if he or she has formed a new family. The decision will depend on whether the person joins a previously existing family in the universe and the size of the family joined.

An uneven merger of two preexisting families also presents some decisionmaking problems from a sociological perspective. Such mergers occur when one or more related persons join another set of related persons or when a marriage occurs and one or more of the marriage partners bring children from a previous marriage (or another related person) with them. The first type of situation presents few problems. Most of these cases involve the entering or reentering of continuing families by elderly parents, adult children, or other relatives. Usually these new family members constitute the smaller of the two merging families. The larger of the two families entering the merger generally has reciprocal majority linkages to the newly merged family. (The smaller family never has.) The two reciprocally linked families are considered one continuing family. Occasionally, an uneven merger may produce a totally new family if the merged family cannot be linked to any preexisting family. The above result appears to be in line with the general sociological consensus that a family's identity is not changed by the addition or return of elderly parents, adult children, etc. Of course, if the additional family members come from out of scope (that is, if they are newborn children, come out of an institution, or return from the military or from overseas), they do not affect the identity of the family. These instances probably represent the majority of uneven mergers. However, there is less sociological consensus as to what the merged family represents when an uneven merger results from a marriage. The reciprocal majority population rule treats this situation in the same manner as the preceding one. For situations in which a single spouse enters an already existing larger family, the result appears appropriate. Where both spouses bring large families into the marriage, the result may be questionable. However, these latter situations represent a very small number of cases.

Construction and Use of Family Weights

Initial Family Weights

The target population of the household survey (HHS) was civilian noninstitutionalized families existing in the United States at any time during 1980. The universe of families existing on any specific day during 1980 was potentially different from that existing on any other day of the year. Conceptually, one could have conducted a census of the eligible population of the United States on January 1, 1980. By following this initial universe of families throughout the year, every unique longitudinal family unit could be identified and labeled. These longitudinal family units are defined by a beginning date, an ending date, and a set of persons who qualify as eligible (civilian and noninstitutionalized) family members. In addition to all family units that can be linked to the initial January 1 family universe, there are persons and families who were ineligible on January 1, 1980, but subsequently returned to the civilian noninstitutionalized population without merging with families containing individuals who were eligible on January 1. Such individuals and families were eligible for the sample but did not have a chance of entering it. Poststratification weight adjustments partially compensated for this undercoverage.

The family weights for longitudinal families in the household sample were developed from the sampling weights for the initially sampled families, which were called originating base reporting units (OBRU's). For each HHS longitudinal family, the key family members all belonged to the same OBRU. Hence, the initial family weight for the j^{th} key HHS longitudinal family was computed as follows.

$$WF_1(j) = [n(j)/g(j)]\, w_o(j),$$

where $n(j)$ is the number of key individuals in family j on its beginning date, $g(j)$ is the total number of members of family j on its beginning date, and $w_o(j)$ is the OBRU initial sampling weight for the key members of family j. Thus, the initial family weight is the OBRU sampling weight adjusted for person-level multiplicity. Essentially, this formula means that the sampling weight of a family beginning on January 1, 1980, is the same as the household sampling weight, regardless of when the family ended or family membership changed in the subsequent 12 months. However, if a family began on some day after January 1, 1980, the household sampling weight was adjusted to take into account the fact that the new family may have had multiple chances of getting into the sample. However, as previously pointed out, positive sampling weights were developed only for key longitudinal families. Further details of the methodology for HHS longitudinal sampling weights are provided by Whitmore, Cox, and Folsom (1982).

Adjustment for Undercoverage and Nonresponse

Poststratification adjustment of the initial HHS family weights to the family counts based on the March Supplement to the 1980 Current Population Survey (CPS) was used to reduce the variance of estimators and the bias from undercoverage. These counts, however, were from estimates based on an updating of the 1970 census. Therefore, NMCUES family counts and estimates may not agree with family counts and estimates based on the 1980 census. The poststratification adjustments and a weighting class adjustment were also used to reduce the bias from nonresponse of longitudinal families.

A key HHS longitudinal family was classified as responding if it satisfied the following three requirements.

1. At least one key family member was classified as a respondent; that is, at least one key family member responded for at least one-third of his or her eligible days in the survey.
2. The total number of responding (known eligible) days during the family's existence summed over all family members is at least one-third of the total number of eligible days during the family's existence summed over all members of the family.
3. The family contained no students who were listed only on the parents' round 1 secondary reporting unit roster and for whom no other data collection instrument was ever received.

This definition of a responding family was felt to be consistent with the definition of person-level response and was used to create the HHS family response indicator variable. Only about 0.1 percent of all longitudinal families were declared to be nonresponding because of condition 3. Imputation of a full year of data for these students was problematic. Hence, inclusion of condition 3 in the definition of a responding family was felt to be cost effective.

The initial multiplicity-adjusted family weight was computed for all longitudinal families from the initial OBRU weight. A poststratification adjustment was then made for nonresponse of families linked to nonresponding OBRU's, producing an adjusted weight. A weighting class adjustment was performed for nonresponding longitudinal families generated by responding OBRU's. This adjusted weight was then truncated to produce a new family weight. The final adjustment was a poststratification and smoothing to the March Current Population Survey family counts to produce the final HHS longitudinal family weight, FWEIGHT. An alternative family weight, AWEIGHT, which was adjusted for each family's eligible days, was also computed from FWEIGHT to facilitate analytic tabulations. AWEIGHT, a time-adjusted family weight, is equal to FWEIGHT times the proportion of 1980 for which the

family existed. (Computationally, it equals FWEIGHT times the family's survey eligibility days divided by 366, the total number of days in 1980.) The time-adjusted family weights, AWEIGHT, sum to the average daily number of HHS-eligible longitudinal families in the United States in 1980.

Estimators

This family weighting scheme produces the adjusted family weight, FWEIGHT, which can be used directly for estimation of annual health care utilization and expenditure. For example, if $Y(j)$ represents the total expenditure of the j^{th} HHS longitudinal family for a particular medical service in 1980, then

$$\Sigma \text{FWEIGHT}(j)Y(j)$$

estimates the total expenditure of all civilian noninstitutionalized families in the United States for this medical service in 1980, where the summation extends over all longitudinal families in the NMCUES HHS sample.

Rates of utilization and expenditure are, however, of more interest than population totals. The rates of annual utilization and expenditure per family for a given family domain, say domain d, are defined at the population level by

$$R(d) = [\sum_{j=1}^{J} X_d(j)Y(j)]/[\sum_{j=1}^{J} X_d(j)PE(j)],$$

where j = 1,..., J indexes the population of all key longitudinal families that ever existed in 1980 (that is, all longitudinal families that had a chance for selection as key NMCUES families);

$X_d(j)$ = 1 if family j belongs to domain d, 0 otherwise;

$Y(j)$ = total utilization or expenditure for family j during the portion of 1980 that family j was eligible for NMCUES; and

$PE(j)$ = proportion of 1980 that family j was eligible for NMCUES, or (FAMEND − FAMBEG + 1)/366, where FAMEND = family ending date (days of 1980 numbered 1 through 366) and FAMBEG = family beginning date.

The family aggregates, $Y(j)$, can be viewed as sums of associated person-level visit counts or expenditures for key and nonkey individuals belonging to family j during the time period in which they were members of the family. The denominator of $R(d)$ is the average daily number of families of type d that existed during 1980. The bracketed portion of the numerator of $R(d)$ is simply the total number of health care visits or the total expenditures of a specified type experienced by NMCUES eligible persons while they belonged to families of type d.

Unbiased estimators for the numerator and denominator of $R(d)$ lead to the ratio estimator $r(d)$, for which the equation is:

$$r(d) = [\Sigma \text{FWEIGHT}(j)X_d(j)Y(j)] / [\Sigma \text{FWEIGHT}(j)X_d(j)PE(j)],$$

where the summation extends over all longitudinal families in the sample. Of course, it is necessary to compute $X_d(j)$ and $PE(j)$ only for responding families because FWEIGHT is zero for all other families. Two alternative formulations of this estimator that may be more convenient for some computations are:

$$r(d) = [\Sigma \text{AWEIGHT}(j)X_d(j)Y(j)/PE(j)] / [\Sigma \text{AWEIGHT}(j)X_d(j)],$$

and

$$r(d) = [\Sigma \text{FWEIGHT}(j)X_d(j)Y(j)] / [\Sigma \text{AWEIGHT}(j)X_d(j)],$$

where the summations extend over all longitudinal families and AWEIGHT(j), as previously noted, is the final time-adjusted weight for family j; that is,

$$\text{AWEIGHT}(j) = \text{FWEIGHT}(j)\,PE(j).$$

Throughout this report, all estimates are based on the first of these two alternative formulations. All counts of expenditures for health care employ as the measure of expenditure used

$$\Sigma \text{AWEIGHT}(j)X_d(j)Y(j)/PE(j),$$

and all counts of families employ as the number of families in question

$$\Sigma \text{AWEIGHT}(j)X_d(j).$$

To be more specific, the statistics presented in the detailed tables of this report are estimated as follows.

The *number of families* with given characteristic(s) is estimated as

$$\Sigma \text{AWEIGHT}(j)X_d(j),$$

where $X_d(j)$ = 1 if family j has the characteristic(s) in question and 0 otherwise.

Note that this estimator estimates the number of family years experienced by families with the given characteristic(s) or, equivalently, the average number of families with the given characteristic(s) that would have been found at a randomly chosen point in time in 1980. It is, in general, less than the cumulative total of distinct longitudinal families with the given characteristic(s) that ever existed at any time in 1980, some of which existed for only part of the year.

The *mean* for use or expenditure is always the mean rate per family year and is estimated as

$$[\Sigma \text{AWEIGHT}(j)X_d(j)Y(j)/PE(j)] / [\Sigma \text{AWEIGHT}(j)X_d(j)].$$

The *percent* of families with a given characteristic is estimated as

$$[\Sigma \text{AWEIGHT}(j)X_d(j)X_u(j)] / [\Sigma \text{AWEIGHT}(j)X_d(j)]$$

where $X_u(j) = 1$ if family j has the given characteristic and 0 otherwise.

Note that this estimator has as its denominator the estimated number of family years experienced by all families in a domain defined by a set of family characteristics and has as its numerator the estimated number of family years experienced by families in the domain that also have the utilization characteristic in question. In other words, the estimator involves a ratio of family years.

Special Requirements for Imputation of Family Data

As noted in the previous section, estimation of utilization and expenditure rates requires family aggregate data, say $Y(j)$, where the aggregates can be obtained as sums of associated person-level visit counts or expenditures. To compute the family aggregate $Y(j)$, it is necessary to sum over all members of family j, both key and nonkey. Moreover, computation of annual utilization and expenditure statistics requires a full year of data for every member of each responding family. Hence, in the attrition imputation, a weighted sequential hot-deck procedure was used to produce complete data for individuals who did not respond for the full year. In the attrition task (Cox and Sweetland, 1982), each individual was first classified as either having complete data or having incomplete data, based on whether the individual had responded for all 366 days in 1980. The data records for individuals who had not responded for the full year were completed by attrition imputation, including imputation of eligibility status (eligible or ineligible) for each day in 1980. The major importance of the attrition task is that it provided a full year of data for every individual from which family aggregates, $Y(j)$, can be computed. The concept of a key responding family was defined in such a way, however, that minimal use of data from the attrition task is required. Of course, missing item data can also lead to missing values for the family aggregate, $Y(j)$. Hence, item imputation procedures (Cox et al., 1982) were performed in addition to attrition imputation to assure the availability of complete data for important analytical variables for every eligible day for each family member.

Reliability of Estimates

Standard Errors

The estimates presented in this report are based on a sample of the target population rather than on the entire population. Thus, the values of the estimates may be different from values that would be obtained from a complete census. The difference between a sample estimate and the population value is referred to as the sampling error, and the expected magnitude of the sampling error is measured by the standard error. Estimated standard errors for the estimates in Tables A–F are generally next to each estimate.

The SESUDAAN (Shah, 1981) standard error estimation software package was used to produce the estimates of standard errors. SESUDAAN is a Taylor Series procedure, developed and released by the Research Triangle Institute. It runs within the Statistical Analysis System (SAS Institute, Inc., 1982).

In addition to sampling errors, the estimates presented in this report are subject to nonsampling errors, such as biased interviewing and reporting, undercoverage, and nonresponse. The standard error does not provide an estimate of nonsampling errors. However, as discussed in preceding sections, every effort was made to minimize these errors.

Confidence Intervals

The estimates in this report are subject to sampling error. The true values are unknown. But the sampling error can be used to determine a range of values such that the true value will be within that range with a known probability. This range is called a confidence interval.

Suppose that $\hat{\theta}$ is an unbiased estimator for the parameter θ, and $S_{\hat{\theta}}$ is a consistent estimator for the standard error of $\hat{\theta}$. Under appropriate central limit theorem assumptions regarding $\hat{\theta}$, the statistic $Z = (\hat{\theta} - \theta)/S_{\hat{\theta}}$ has an approximate standard normal distribution for large samples. Thus, an approximate $(1 - \alpha) \times 100$ percent confidence interval for θ is given by

$$(\hat{\theta} + z_{\alpha/2}S_{\hat{\theta}}, \hat{\theta} + z_{1-\alpha/2}S_{\hat{\theta}}),$$

where $z_{\alpha/2}$ and $z_{1-\alpha/2}$ are the appropriate values from a standard normal table.

As an example, Table A shows that, of all multiple-person families in the civilian noninstitutionalized population of the United States, an estimated 17.5 percent had total charges for health care of $3,000 or more in 1980. The estimated standard error of 17.5 percent is .60 (Table A). As $Z_{.025} = -1.96$ and $Z_{.975} = 1.96$, a 95-percent confidence interval for the percent of all multiple-person families with such charges in 1980 is $17.5 \pm (1.96 \times .60)$, or the interval 16.3–18.7 percent. Approximately 95 percent of the confidence

intervals constructed in this manner will contain the true percent of families with total charges for health care of $3,000 or more in 1980.

Confidence intervals for the difference of two parameters can be constructed in a similar manner. Suppose θ_1 and θ_2 are the values of the parameter of interest in two mutually exclusive population subgroups. If $\hat{\theta}_1$ and $\hat{\theta}_2$ are unbiased estimators of θ_1 and θ_2, respectively, then $\hat{d} = \hat{\theta}_1 - \hat{\theta}_2$ is unbiased for $d = \theta_1 - \theta_2$ and

$$\text{Var}(\hat{d}) - \text{Var}(\hat{\theta}_1) + \text{Var}(\hat{\theta}_2) - 2\,\text{Cov}(\hat{\theta}_1,\hat{\theta}_2).$$

Unfortunately, the estimation of $\text{Var}(\hat{d})$ presents a problem because it is not possible for the National Center for Health Statistics to provide the reader with covariance estimates for all possible pairs of subdomains of potential interest. However, if it is reasonable to assume that $\text{Cov}(\hat{\theta}_1,\hat{\theta}_2) = 0$, the standard error of d can be estimated by

$$S_{\hat{d}} = \sqrt{S^2_{\hat{\theta}_1} + S^2_{\hat{\theta}_2}}.$$

Then, under appropriate central limit theorem assumptions regarding d, the statistic $Z_d = (\hat{d} - d)/S_{\hat{d}}$ has an approximate standard normal distribution for large samples, and the interval

$$(\hat{d} + z_{\alpha/2}S_{\hat{d}},\ \hat{d} + z_{1-\alpha/2}S_{\hat{d}})$$

is an approximate $(1-\alpha)\times 100$ percent confidence interval for the difference d.

For example, suppose we wanted to construct a 95-percent confidence interval for the difference between the percent of older families with total charges for health care of $3,000 or more ($\theta_1$) and the percent of younger, lower-income families with the same amount of total charges (θ_2). It can be seen in Table A that $\hat{\theta}_1 = 26.9$ and $\hat{\theta}_2 = 15.5$, so

$$\begin{aligned}\hat{d} &= \hat{\theta}_1 - \hat{\theta}_2\\ &= 26.9 - 15.5\\ &= 11.4.\end{aligned}$$

From Table A, it can be seen that $S_{\hat{\theta}_1} = 1.7$ and $S_{\hat{\theta}_2} = 1.2$, so

$$\begin{aligned}S_{\hat{d}} &= \sqrt{S^2_{\hat{\theta}_1} + S^2_{\hat{\theta}_2}}\\ &= \sqrt{2.89 + 1.44}\\ &= \sqrt{4.33}\\ &= 2.08.\end{aligned}$$

Then, as $\alpha = .05$, it follows that $z_{\alpha/2} = -1.96$ and $z_{1-\alpha/2} = 1.96$, so the 95-percent confidence interval for the difference of interest is (15.48, 7.32).

The reader should be aware that the assumption that $\text{Cov}(\hat{\theta}_1,\hat{\theta}_2) = 0$ is frequently not true for complex sample surveys. This warning is especially germane for sample designs, such as the NMCUES design, that rely on cluster sampling at one or more stages of sample selection. If $\text{Cov}(\hat{\theta}_1,\hat{\theta}_2)$ is positive, the confidence interval will tend to be too large, and the confidence level will be understated. More seriously, if $\text{Cov}(\hat{\theta}_1,\hat{\theta}_2)$ is negative, the confidence interval will tend to be too small, and the confidence level will be overstated.

Hypothesis Testing

The statistics Z and Z_d can be used to test hypotheses. For example, the size α critical region for the composite hypothesis

$$H_0: d \geq d_0$$

versus

$$H_A: d < d_0$$

is given by

$$Z_{d_0} = \frac{\hat{d} - d_0}{S_{\hat{d}}} \leq z_\alpha .$$

As an example, suppose that before any data were collected one had a reason to believe that the percent of younger, lower-income families with total charges for health care of $5,000 or more ($\theta_1$) was less than the percent of older families with the same amount of total charges (θ_2). Letting $d = \theta_1 - \theta_2$, this can be restated as a formal hypothesis as

$$H_0: d \geq 0$$

versus

$$H_A: d < 0.$$

Note that what is believed to be the true state of nature is reflected by the one-sided alternative.

It can be seen from Table A that

$$\hat{d} = 8.0 - 17.8 = -9.8$$

and

$$\begin{aligned}S_{\hat{d}} &= \sqrt{.81 + 2.25}\\ &= 1.75,\end{aligned}$$

so that $Z_{d_0} = -5.60$. As there are three categories for family socioeconomic status in Table A, a multiple t test based on the Bonferroni inequality (Levy and Lemeshow, 1980) will be used to assess the significance of the comparison. Comparing three categories, two at a time, and not taking sign into account gives three possible comparisons. Use of the table in Levy and Lemeshow (1980, p. 296) gives a one-tail critical value of -2.13. Therefore, H_0 is rejected in favor of H_A as $Z_{d_0} \leq z_\alpha$.

As discussed earlier, the assumption that $\text{Cov}(\hat{\theta}_1, \hat{\theta}_2) = 0$ must be carefully evaluated. If in fact the covariance is positive, the size of the test will be smaller than α, and if the covariance is negative, the size of the test will be larger than α. Readers who want to conduct more sophisticated analysis of the NMCUES data are advised to consult with a statistician knowledgeable in the analysis of data from complex sample surveys.

Appendix III
Definition of Terms

Accidents, injuries, and poisonings—This category includes injuries; wounds; burns; poisonings; toxic effects; complications of medical and surgical care; and early complications, late effects, and impairments due to the previous causes. Conditions were reported by the family respondent and recoded according to the International Classification of Diseases, 1975 revision.

Age of family head—Age is as of January 1, 1980.

Ambulatory physician visit—A visit by a patient to a physician's office, clinic, or similar place is an ambulatory physician visit. Visits are counted whether a physician or only a member of the physician's staff is seen. House calls and visits to school or workplace clinics are also included. Family visits are the sum of all visits by family members during the time they were in the family.

Bed days—Bed days are days spent in bed by a family member because of illness or injury. Family bed days are the sum of all bed days of family members during the time they were in the survey, prorated to the time they were in the family.

Cancer and other neoplasms—This category includes malignant neoplasms of all sites and tissues, benign neoplasms, carcinoma in situ, and other and unspecified neoplasms. Conditions were reported by the family respondent and recoded according to the International Classification of Diseases, 1975 revision.

Circulatory and heart disease—This category includes rheumatic fever, rheumatic heart disease, hypertensive disease, ischemic heart disease, diseases of the pulmonary circulation, other forms of heart disease, cerebrovascular disease, and other diseases of the circulatory system. Conditions were reported by the family respondent and recoded according to the International Classification of Diseases, 1975 revision.

Civilian noninstitutionalized family—This refers to families in which all members are members of the civilian noninstitutionalized population. Families whose heads are members of the military are defined as not being civilian families and are excluded in their entirety from this report, although they were included in the sample and the weighting. In the sample, there were 49 such families (about 0.7 percent). Family members other than the head who were in the military were excluded from the survey even if they resided with the family.

Dental visit—A visit to a dentist's office is a dental visit. A dentist or a member of the dentist's office staff may have provided services. Family visits are the sum of all visits by family members during the time they were in the family.

Education of family head—The years of school completed by family heads 17 years of age and over constitute the education of family heads. Only years completed in regular schools, where persons are given a formal education, are 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 vocational, trade, or business schools outside the regular school system was not counted in determining the highest grade of school completed.

Ethnicity of family head—The ethnicity of family heads 17 years of age and over is as reported by the family respondent. The ethnicity of family heads under 17 was imputed. Ethnicity is classified as (1) Hispanic, which includes Puerto Rican, Cuban, Mexican, Mexicano, Mexican American, Chicano, other Latin American, and other Spanish or (2) non-Hispanic.

Family—A family is a group of people who share a common housing unit and are related to each other by blood, marriage, adoption, or a foster care relationship. An unmarried student 17–22 years of age living away from home is also considered part of a family even though his or her residence was in a different location. The group of people who compose the family may change composition over time, causing the family to take on one or a combination of the following time-related states: existing over time without change in membership; existing over time with change in membership; going out of existence before the end of the survey; coming into existence after the beginning of the survey; or existing for the whole survey. For more detail, see Appendix II.

Family dynamics—A family is considered unchanging, or static, if it existed for the whole of 1980 and its membership was unchanged. Families that had changes in membership and/or did not exist for the whole of 1980 are considered changing, or dynamic, families.

Family illness days in bed—See bed days.

Family income in 1980—For each person in the family, data were collected on 12 categories of income. These included income from employment for persons 14 years of age and over; income from various

government programs; income from pensions; alimony or child support; interest income; and net rental income. When information was missing, income was imputed. The total income of persons who were members of more than one family was allocated to each family they were in, in proportion to the amount of time they were in that family. Person-level incomes in each family were summed to create a family-level total. If a family did not exist for an entire year, the family income was adjusted to an annual basis by dividing actual income by the proportion of the year the family existed.

Family-paid premiums—The amount paid for premiums by a family and not reimbursed. Much of the cost of premiums in the United States is paid by employers or by public funds rather than by families.

Family size—The time-weighted average number of persons in a family determines the size. Family size was computed by (1) summing the number of days in the family for each person who was ever a family member and (2) dividing this sum by the number of days the family was in existence. For example, if a family existed for 200 days and had two persons who were members throughout its existence and one person who was a member for 80 days, the family size is 2.4.

Family structure—Family structure refers to the presence or absence of family head, spouse, and children under 17, and whether these persons were present for the family's entire duration or part of its duration.

Family work-loss days—See work-loss day.

Family years—Family years refers to the length of time that a family, or a collection of families, existed as a unit of analysis in (were eligible for) the survey, as measured in units of a year or fractions of such units.

For an individual longitudinal family in the NMCUES sample, the number of family years equals the number of days the family was eligible for the NMCUES sample divided by 366, the number of days in 1980 (the NMCUES sample period). For such a family weighted to represent a group of families in the NMCUES universe, the number of family years is AWEIGHT(j), which is equal to FWEIGHT(j), the basic adjusted weight, times $PE(j)$, the proportion of the year the family was eligible for the sample. For a group of sample families, the associated number of family years is the sum of the AWEIGHT's. For further details and fuller definitions of variables, see the section on estimators in Appendix II.

Group quarters—This is a structure occupied by five or more unrelated people who lived or ate together, or for whom there was neither direct access from the outside or through a common hall nor complete kitchen facilities. Only noninstitutional group quarters were included in the NMCUES sample frame. Each unrelated person in a group-quarter household was considered a separate one-person family, unless he (or she) was a student away from home. (See definition of family.)

Head of family—A person was designated as the family head by the respondent at the time of the first interview. If no head was designated or this information was missing, a family head was imputed. Among families in which the person designated as head changed over time, the characteristics of the person who was designated head the longest were used for all head-of-family variables.

Health care coverage—Health care coverage refers to the situation in which a private insurance plan or public health care coverage program (Medicare, Medicaid, and so forth) can be used to pay all or a part of a family's or person's health care costs. For this report, completeness of health care coverage was coded into four categories: (1) "full coverage," meaning all family members had health care coverage during their entire survey eligibility period; (2) "partial coverage 1," meaning all family members had health care coverage at some time, but some or all had coverage for only part of their survey eligibility period; (3) "partial coverage 2," meaning some family members had health care coverage for at least part of their survey eligibility period, but some never had coverage; and (4) "no coverage," meaning no family members had health care coverage at any time during their survey eligibility period.

For this report, a family was coded as having a particular source of health care coverage (such as private insurance, Medicare, Medicaid, or a particular combination of coverages) on the basis of the known coverage of family members. Only when the source of coverage was unknown, or not assignable, for *all* family members was the family coded as having source of coverage unknown. However, the coding categories for individuals upon which the family health care coding in this report was based do not identify the coverage source(s) for individuals with part-year or no coverage. Thus, most families coded as having source of coverage unknown are families with no members having full-year coverage. Also, the coding categories for individuals upon which the family health care coverage coding in this report was based are different from the categories used in a previous family report (Dicker, 1983a) that dealt with only a part of the survey year. As a result, there may be differences in coverage estimates between the reports.

In this report, coverage by CHAMPUS is classified as coverage by private insurance.

Health care services—NMCUES includes information on eight types of health care services. These are (1) inpatient hospital care, (2) inpatient physician care, (3) ambulatory physician visits, (4) hospital emergency room and outpatient visits, (5) dental visits, (6) prescription acquisitions, (7) services of other *independent* medical providers such as chiropractors, speech therapists, faith healers, and psychologists (unless such providers are working as part of a physician's staff, in which case their services are counted in physicians' care) and (8) acquisition of other health care supplies such as eyeglasses, orthopedic items, hearing aids, ambulance services, and diabetic items. Excluded from the data

in this report are nonprescription medicines, nursing home care, and care in other long-term care institutions.

Hospital admission—Hospital admission is the formal acceptance by a hospital of a patient who is provided room, board, and regular nursing care in a unit of the hospital, including patients admitted for childbirth. A patient admitted to the hospital and discharged on the same day is included as a hospital admission. A hospital stay resulting from an emergency department visit is also included. Family hospital admissions are the sum of all admissions of family members during the time they were in the family.

Hospital discharge—A hospital discharge is the formal release by a hospital of a patient who was provided room, board, and regular nursing care in a unit of the hospital. A patient admitted to the hospital and discharged on the same day is included as a hospital discharge. A hospital stay resulting from an emergency room visit and subsequent admission of the patient is also included. Family hospital discharges are the sum of all discharges by family members during the time they were in the family.

Hospital emergency room—The emergency room is a facility within a hospital organized to provide medical services to people needing immediate medical or surgical intervention. People receiving care in the emergency room may be admitted to a hospital.

Hospital emergency room visit—This is a face-to-face encounter between a patient (not necessarily ambulatory) and a medical person in the hospital emergency room. Encounters by patients transported to the emergency room by police or by emergency medical service are included. The visit may result in a hospital admission. Family emergency room visits are the sum of all emergency room visits by family members during the time they were in the family.

Hospital outpatient department—This is a hospital-based ambulatory care facility organized to provide non-emergency medical services. Persons receiving services do not receive inpatient nursing care. Examples of outpatient departments or clinics are pediatrics, obstetrics and gynecology, eye, and psychiatric.

Hospital outpatient department visit—This is a face-to-face encounter between an ambulatory patient and a medical person in a hospital outpatient department. 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 was visited on a single trip, each department or clinic visited was counted as a separate visit. Family outpatient department visits are the sum of all hospital outpatient department visits by family members during the time they were in the family.

Household—This refers to occupants of a housing unit or group quarters included in the sample. A household can be one person, a family of related people, a number of unrelated people, or a combination of related and unrelated people. Therefore, a household can contain more than one family. (See definition of family.)

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

Inpatient hospital care—This is health care provided to a patient by a hospital during the period from the patient's admission to the patient's discharge. This includes admissions for deliveries of babies.

Inpatient physician care—This care is provided to a patient by a physician (or a physician's staff) during the period from the patient's admission to a hospital to the time of the patient's discharge from the hospital. Such care was only recorded in NMCUES if a physician's bill, separate from the hospital bill, was rendered for such care. Otherwise, a patient was not regarded as having received inpatient physician care.

Institution—An institution is a place providing room, board, and certain other services for residents or patients. Correctional institutions, military barracks, and orphanages were always considered institutions in NMCUES. Places that provide long term health care were also identified as institutions if they provide either nursing or personal care services. Certain other facilities licensed, registered, or certified by a State agency or affiliated with a Federal, State, or local government agency were also defined as institutions. People residing in institutions were not included in the household sample.

Key person—See the discussion under "Sample Design" in Appendix II.

Limitation in major activity—Four categories were developed for classifying limitation in major activity:

1. Cannot perform usual major activity (such as working, going to school, or keeping house).
2. Can perform usual major activity but limited in kind or amount.
3. Can perform usual major activity but limited in kind or amount of other activity.
4. Not limited.

People 6 years of age and over were classified into any of the categories; children 1–5 years of age were classified into categories 1, 2, and 4; and children under 1 year of age were classified into categories 1 and 4. In this report, categories 2 and 3 are combined into the category "secondary limitation." The NMCUES classified persons with unknown limitations as not limited.

Longitudinal family—A longitudinal family is a family identified as the same family over a time period. It may or may not have had changes in family membership during the time period. (See the definition of family.)

Marital status—Marital status for each person 17 years of age and over is as indicated by the household respondent.

Metropolitan area—See urban-rural location.

Metropolitan center city—See urban-rural location.

Metropolitan suburb—See urban-rural location.

Multiple-person family—A family with an average size of 1.5 members or more is a multiple-person family.

National household component—One component of NMCUES, this consists of multiple household interviews with an area probability sample of people in the noninstitutionalized population of the United States in 1980.

Nonkey person—See the discussion under "Sample Design" in Appendix II.

Nonmetropolitan urban area—See urban-rural location.

Nonurban area—See urban-rural location.

Number of families—This refers to the average number of families with a given set of characteristics that would have been found at a randomly chosen point in time in 1980. This is equal to the number of family years experienced during 1980 by families with the given characteristics. It is, in general, less than the cumulative total number of distinct longitudinal families with the given characteristics that ever existed at any time in 1980, some of which existed for only part of the year.

One-person family—A family with average size less than 1.5 is a one-person family. More than one individual may be involved.

Out-of-pocket expenses—This is the amount paid by a family and not reimbursed by either insurance or other health care payment programs.

Out-of-pocket expenses for health care services—This is the amount paid out-of-pocket by a family for all eight of the types of health care services covered by NMCUES. (See definition of health care services, above.) This does not include family-paid premiums.

Perceived health status—This is the family respondent's rating on a 4-point scale of the health of a family member compared with the health of other persons of the same age, as reported at the time of the first interview. The categories are "excellent," "good," "fair," and "poor." When a family consisted of only one member, this was a self-rating.

Point-interval family—A point-interval family is a family with exactly the same family membership over a time period. A change in family membership ends one point-interval family and begins another.

Poverty status of family—The poverty status in 1980 was calculated by dividing the family's income in 1980 by the appropriate 1980 poverty level threshold and converting it to a percent. For example, a family with income between two and three times the poverty level threshold that corresponds to its size and other characteristics would be classified in the 200–299 percent category. The poverty level thresholds, as used by the U.S. Bureau of the Census, are determined by the age and sex of the family head and the average number of persons in the family. In 1980, average poverty level thresholds by family size (weighted for the mix of families by sex and age of head) were: 1-person, $4,190; 2-person, $5,368; 3-person, $6,565; 4-person, $8,414; 5-person, $9,966; 6-person, $11,269, 7-person, $12,761; 8-person, $14,199; 9-person and larger, $16,896.

Premium—The premium is the amount paid for private health insurance or other health care coverage.

Prescription acquisition—This describes the obtaining of a medication by a family member requiring a prescription from a doctor or dentist. Both initial fillings of prescriptions and refills are counted as acquisitions. Family prescription acquisitions are the sum of all acquisitions by family members during the time they were in the family.

Principal respondent—This is the member of the reporting unit who provided most of the information for the people in the reporting unit.

Proxy respondent—As used in this survey, a proxy respondent was a person who provided information for people in the reporting unit but who was not a member of the reporting unit. A proxy respondent was used only when no member of the reporting unit could supply the information because of physical or mental incapacity.

Race of family head—The race of the family head is as reported by the family respondent or imputed. Race is classified as "white," "black," or "other." The "other" race category includes American Indians, Alaskan Natives, Asians, Pacific Islanders, and people not identified by race. The category "all other" includes the categories "black" and "other."

Rate per family year—Amount of care used or dollars expended by a family or group of families is divided by the number of family years experienced by these families while eligible for the NMCUES sample. All data on the use of care or on health expenditure in this report are presented in terms of a rate per family year. For a given family, the rate per family year equals $Y(j)/PE(j)$,

where $Y(j)$ = use of care or expenditure during family's period of eligibility for NMCUES sample, and

$PE(j)$ = proportion of year family was eligible for the NMCUES sample.

The section on estimators in Appendix II presents more details of calculations.

REF. DATE—The reference date was the date of the previous interview in most cases. For the first interview, however, it was January 1, 1980. For new persons, it was the date they joined the reporting unit. For the final interview, it spanned the time between the next-to-last interview and December 31, 1980.

Region—The 50 States and the District of Columbia are categorized into four regions by the U.S. Bureau of the Census. This classification by region was used in this report. The regions and their constituent parts are as follows. 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 (RU)—A reporting unit is the basic unit for collecting data in the household components of NMCUES at the time of each interview. A reporting unit consisted of all related people residing in the same housing unit or group quarters during the reference period covered by a particular interview. One person could give information for all members of the reporting unit.

Secondary reporting unit—Unmarried students 17–22 years of age usually living in a sampled household but away from home as full-time students were considered secondary reporting units. Also, in a household with multiple families, the reporting unit with the largest number of individuals was usually designated the primary reporting unit, and all other families were designated secondary reporting units.

Sex—Sex was recorded by the interviewer in the initial NMCUES interview.

Spouse—The spouse is the person designated by the respondent as the spouse of the head of the family.

Total charges—This is the full amount billed (either actual or imputed) to a family for all eight types of health care services covered by NMCUES, whether these amounts are paid out-of-pocket by the family, paid by health care coverage, or go unpaid.

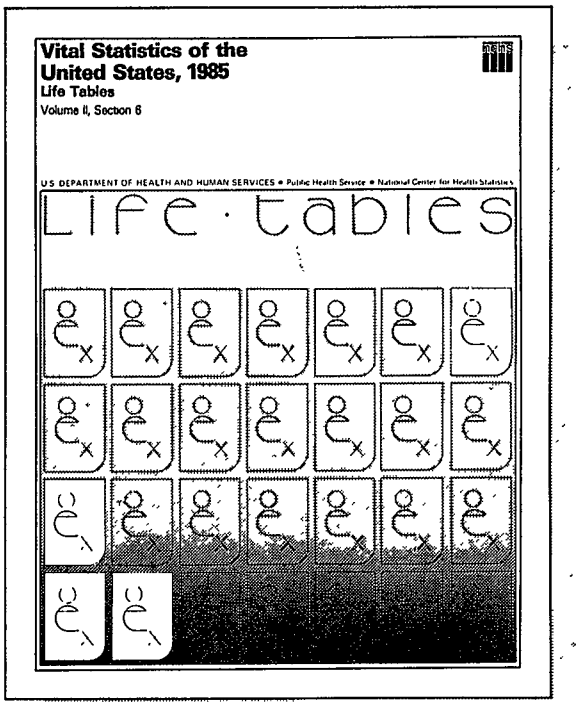
Total out-of-pocket expenses—This is the amount paid by a family for out-of-pocket expenses for health care services plus the amount of family-paid premiums.

Urban-rural location—Households were identified as located either within or outside of a Standard Metropolitan Statistical Area (SMSA). The definitions and titles of SMSA's are established by the U.S. Office of Management and Budget with the advice of the Federal Committee on Standard Metropolitan Statistical Areas. Households located inside SMSA's are further classified as being located within the SMSA's central city (called "central city") or not ("other"). Households located outside of SMSA's are classified as "urban" if they are located in (1) places of 2,500 inhabitants or more that are incorporated as cities, villages, boroughs (except Alaska), and towns (except in New England, New York, and Wisconsin), but excluding persons living in the rural portions of extended cities; (2) unincorporated places of 2,500 inhabitants or more; or (3) other territory, incorporated or unincorporated, included in urbanized areas. Otherwise, households located outside of SMSA's are classified as "rural."

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. Family work-loss days are the sum of all work-loss days of family members during the time they were in the survey, prorated to the time they were in the family.

Available Only from the Government Printing Office...

Certified Life Tables



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

nchs

Publication Order Form

Order processing code: * 6247

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

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

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

Please Type or Print

_____
(Company or personal name)

_____
(Additional address/attention line)

_____
(Street address)

_____
(City, State, ZIP Code)

(____) _____
(Daytime phone including area code)

Please choose method of payment:

☐ Check payable to the Superintendent of Documents

☐ GPO Deposit Account ☐☐☐☐☐☐☐☐–☐

☐ VISA, MasterCard or Choice Account

☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐

_____
(Signature)

_____
(Credit card expiration date)

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

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

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

937 Bestsellers

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

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

Office of Research and Demonstrations
Judith B. Willis, Director

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

Division of Program Studies
Carl Josephson, Director

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

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

Centers for Disease Control
James O. Mason, M.D., Dr.P.H., Director

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

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

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

Utilization and Expenditure Statistics Branch
Robert A. Wright, Chief

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