VITAL & HEALTH STATISTICS

Sample Design, Sampling Variance, and Estimation Procedures for the National Ambulatory Medical Care Survey

This report presents the details of the sample design and estimation procedures used in the National Ambulatory Medical Care Survey from its inception in 1973 through 1985. The survey produces estimates for visits made to office-based physicians for medical care.

Data Evaluation and Methods Research
Series 2, No. 108

DHHS Publication No. (PHS) 88–1382

U.S. Department of Health and Human Services
Public Health Service
Centers for Disease Control
National Center for Health Statistics
Hyattsville, Md.
September 1988

Suggested citation

National Center for Health Statistics, E. Bryant and I. Shimizu. 1988. Sample design, sampling variance, and estimation procedures for the National Ambulatory Medical Care Survey. *Vital and Health Statistics.* Series 2, No. 108. DHHS Pub. No. (PHS) 88–1382. Public Health Service. Washington: U.S. Government Printing Office.

Library of Congress Cataloging-in-Publication Data

Bryant, E. Earl (Elmer Earl)
Sample design, sampling variance, and estimation procedures for the National Ambulatory Medical Care Survey.

(Series 2, Data evaluation and methods research ; no. 108) (DHHS publication ; no. (PHS) 88–1382)
"The survey was conducted each year [1973–1985] by the National Opinion Research Center (NORC) of the University of Chicago under a contractual arrangement with the National Center for Health Statistics (NCHS)"—Pref.
"March 4, 1988."
Bibliography: p.
Supt. of Docs. no.: HE 20.6209:2/108
1. Health surveys—United States—Statistical methods. 2. Sampling (Statistics) 3. Ambulatory medical care—United States—Utilization—Evaluation—Statistical methods. 4. National Ambulatory Medical Care Survey (U.S.)—Statistical methods. I. Shimizu, Iris. II. National Center for Health Statistics (U.S.) III. NORC (Organization) IV. Title. V. Series: Vital and health statistics. Series 2, Data evaluation and methods research ; no. 108. VI. Series: DHHS publication ; no. (PHS) 88–1382. [DNLM: 1. National Ambulatory Care Survey (U.S.) 2. Ambulatory Care. 3. Data Collection—methods. 4. Health Survey—United States. 5. Research Design. 6. Sampling Studies—methods. 7. Statistics—United States. W2 A N148vb no. 108]
RA409.U45 no. 108 362.1'0723 s 88–15120
ISBN 0–8406–0394–0 [362.1'0723]

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

Preface

This report presents a detailed description of the sample design and estimation procedures used in the National Ambulatory Medical Care Survey (NAMCS) between 1973 and 1981, during 1985, and planned for use in 1989 and thereafter. The survey was conducted each year by the National Opinion Research Center (NORC) of the University of Chicago under a contractual arrangement with the National Center for Health Statistics (NCHS).

The initial sample was designed under the supervision of Martin Frankel and Benjamin King of NORC, and by Dwight Brock and Earl Bryant of NCHS. The primary stage of the 1985 sampling plan was designed by Steven Heeringa and Judith Connor of the Institute for Social Research, University of Michigan. Secondary stages of sampling were the responsibility of Iris Shimizu of NCHS. Thomas McLemore, Linda Tompkins, and Nonie Atkinson, all of NCHS, assisted with research leading to the secondary stage designs.

Much of this report is based on published and unpublished manuscripts and on internal NCHS memoranda. Those documents were prepared primarily by the individuals cited above. Earl Bryant, in collaboration with Iris Shimizu, prepared the report for publication under a professional service agreement with NCHS. Thomas McLemore, James DeLozier, and James Massey, all of NCHS, carried out the peer review of this report and provided constructive suggestions.

Symbols

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

Contents

Appendixes

List of text figures

List of text tables

Sample Design, Sampling Variance, and Estimation Procedures for the National Ambulatory Medical Care Survey

by Earl Bryant, formerly Associate Director, Interview and Examination Statistics Program, and Iris Shimizu, Ph.D., Office of Research and Methodology

Introduction

The National Ambulatory Medical Care Survey (NAMCS) is an integral part of the U.S. National Health Survey Program that was authorized by Congress in 1956. NAMCS, which provides detailed statistics on the provision and use of ambulatory medical care services in the United States, complements other data systems maintained by the National Center for Health Statistics, such as the National Health Interview Survey, the National Health and Nutrition Examination Survey, the National Nursing Home Survey, and the National Hospital Discharge Survey, to provide a broad picture of the health of Americans and the health services that they receive. NAMCS began in April 1973 after more than 6 years of research to determine the feasibility of the survey and to test alternative methods for conducting the survey. The survey was continuous each year through 1981 when it was temporarily discontinued. It was conducted again in 1985. It is planned that the survey will be continuous again beginning in 1989.

The primary objective of this report is to provide a detailed description of the NAMCS sample design, estimation procedures, and procedures for estimating and computing sampling errors. The report describes the basic design implemented in 1973, the changes that were made during 1973–81, and the redesign for the 1985 and subsequent samples.

History of survey development

In 1967, a study was conducted to determine the adequacy of existing office records as a source of national ambulatory medical care information. Results indicated that, although the vast majority of physicians kept records, variations in their form, style, content, completeness, and accessibility were extensive. In about 20 percent of the cases, records contained illegible terms, abbreviations, and symbols which could be understood by only the recording physician and staff. Also, alphabetical filing systems were used in 80 percent of the records, which made it difficult to relate the records to specific time periods. It was concluded that only the practicing physicians themselves could provide a wide range of information about their patients, and that information would need to be collected on an "encounter form" at the time of visits to physicians' offices.

Two feasibility studies were conducted; the first was in 1968–69 and the other in 1970–71.

The purpose of the first study (Phase 1) was to evaluate ambulatory patient data collection by resident physicians versus collection by lay interviewers and to test the effect of the length of forms on physicians' willingness to participate in the survey.

The results of the study were tentative, nonresponse rates were unacceptably high (around 50 percent), and the feasibility of the study had not been proven. It was necessary, therefore, to conduct another phase of the study, incorporating information accumulated during Phase 1. The improvements were aimed specifically at reducing the data collection work load and physician practice interference, increasing the participants' awareness of the purpose of the survey, and strengthening previously established levels of professional interest and support. Methods were developed to make the medical profession at large, and particularly the sample physicians who were requested to participate, more aware of the purpose and significance of an ambulatory medical care survey. Endorsements were obtained from the American Medical Association (AMA), and before interviewer telephoning began a letter from the Executive Secretary of AMA was sent to all sample physicians indicating full organizational and professional support for the request to participate. Nineteen medical and professional specialty societies subsequently endorsed the survey in principle.

The result of Phase 2 was a significantly improved response rate (80 percent) and a fully tested methodology to be implemented in the national survey. A detailed description of the feasibility and methodological studies has been published (NCHS, 1974b).

Overview of survey procedures

Prior to their randomly assigned week for data collection, physicians selected in the sample were first contacted by mail to briefly describe the survey and to ask for their cooperation and participation. The letter highlighted endorsements by the American Medical Association, the American Osteopathic Association (AOA), and many of the medical specialty professional associations in the United States.

About a week later, field representatives contacted the physicians by telephone to briefly explain the survey and to arrange for an appointment to see them in their offices. During the office visit, the survey was explained in detail and the physicians' participation in the survey was sought. For those

agreeing to participate, survey materials were reviewed with the physicians and anyone else in the office who might assist in the survey such as a nurse or receptionist.

Before the beginning of, and again during, the randomly assigned week for data collection, the interviewers telephoned the physicians to answer questions that may have arisen and to ensure that the procedures were going smoothly. At the end of the survey week, the physicians mailed the finished survey materials to the interviewer, who edited the forms for completeness. Problems of missing or incomplete data were resolved by telephone followups to the physician before the forms were sent to the central office for processing.

Data collection was carried out by the physicians, aided by their office staffs when possible. Two data collection forms were used: The patient log and the patient record, facsimiles of which are shown in appendix I.

The patient log is a sequential listing of patients seen in the doctor's office during the assigned reporting week. The list serves as the sampling frame to indicate the visits for which data were to be recorded. The list is separated from the patient record and retained in the physicians' office.

Based on the physician's estimate of the number of office visits expected during the week, and the number of days the physician expected to see patients, a patient sampling rate was assigned. The rates were so designed that about 30 patient record forms would be completed during the assigned reporting week. The sampling rates ranged from unity for offices expecting fewer than 10 visits per day to one-fifth for offices expecting more than 30 visits per day. These procedures minimized the data collection work load and maintained approximately equal reporting levels among sample physicians regardless of practice size.

Design specifications

A fundamental fact that conditions the design of a multipurpose survey such as NAMCS is that no single factor will determine uniquely the design. Rather, a balance must be sought taking the leading objectives into consideration. Some of the more important specifications that shape the design of the survey can be grouped in the following areas.

Target population—The target population would consist of all visits in the physicians' offices in the conterminous United States to physicians who were engaged in office-based practice and direct patient care at the time of the survey with the following exceptions:

- Visits to specialists in anesthesiology, pathology, clinical pathology, forensic pathology, radiology, diagnostic radiology, pediatric radiology, and therapeutic radiology.
- Visits to physicians in Federal service.
- Visits to all physicians not in office-based practice. These include hospital-based physicians, teachers, or researchers who do not practice in private offices, as well as publicly employed physicians who work for the Veterans Administration, the Public Health Service, or a local public health department.

These exclusion criteria are all based on the characteristics of the physicians' practices and not the characteristics of the patients they see. Therefore, all patient visits involving direct patient care would be considered in scope as long as the physician involved fits the prior definitions for inclusion in the target population.

Sampling frame—The sampling frame of physicians would consist of those listed in the master files of the American Medical Association (AMA) and the American Osteopathic Association (AOA), who are licensed to practice in the United States, and who are identified in the files as being engaged primarily in office-based practice.

Study domains and precision requirements—The sample would be designed to produce estimates with a specified precision for the Nation as a whole and for a calendar year. Of secondary importance would be the ability to estimate visits and their characteristics for each calendar quarter and for the four major census regions of the United States, excluding Alaska and Hawaii. In addition to the annual, national estimates, estimates would be required for the major specialties including at least the four broad groupings of general practice, medical specialties, surgical specialties, and "all other specialties." The types of parameters to be estimated included the total number of visits cross-classified by characteristics of visits, physicians, patients, and geography. Approximate precision requirements for a variety of estimates are shown in tables A and B.

Burden on physician—Generally, physicians are very busy. Therefore no physician would be asked to participate in the survey for longer than 7 days a year and no more often than every third year. The number of patients selected in the sample for a physician's office would be small enough to keep the burden to a minimum, yet assure a sufficiently large sample for producing national estimates.

Allocation of resources—To minimize cost, especially to reduce the high cost of inducting physicians, the survey would be conducted in a sample of primary sampling units (PSU's). A PSU is a geographic area consisting of either a county (or parish or independent city) or a group of counties, except for some PSU's in New England which are groups of townships. The physician files to be sampled would include only physicians whose addresses were associated with sample PSU's. Addresses are handled differently in the AMA and the AOA files. AMA uses either the physicians' offices or the places they prefer to receive medically related mail, which may or may not be their offices. AOA uses the office address. Thus, it is necessary to define the sampling units to be the physicians' names, not their offices. Therefore all sample patients seen in a physician's office or offices must be enumerated, wherever the offices may be located.

Table A. Expected precision of some sample estimates based on nonmedical data from the National Ambulatory Medical Care Survey pilot study by selected domains

Domain of interest	*Approximate relative error of estimate* $V_{P'}$
ALL PHYSICIANS[1]	
Size of estimate P'[2]	Percent
0.08	8.3
0.16	6.0
0.28	5.7
0.40	1.7
0.55	2.3
0.88	1.5
GENERAL PRACTICE[3]	
Size of estimate P'[2]	
0.08	15.0
0.16	11.0
0.28	8.5
0.40	3.5
0.55	3.5
0.88	1.9
GENERAL SURGERY[4]	
Size of estimate P'[2]	
0.08	22.0
0.16	14.0
0.28	12.0
0.40	5.0
0.55	4.9
0.88	2.3
OBSTETRICS AND GYNECOLOGY[5]	
Size of estimate P'[2]	
0.08	23.0
0.16	15.0
0.28	13.0
0.40	5.5
0.55	5.2
0.88	2.5
PEDIATRICS[6]	
Size of estimate P'[2]	
0.08	30.0
0.16	20.0
0.28	16.0
0.40	7.0
0.55	6.3
0.88	3.0

[1]Number of sample physicians = 1,200; number of sample patients = 65,000.
[2]These proportions refer to the following characteristics: (1) Patient not examined: $P' = 0.08$; (2) visit lasted more than 20 minutes: $P' = 0.16$; (3) general history taken: $P' = 0.28$; (4) visits by males: $P' = 0.40$; (5) patient seen before and for same problem: $P' = 0.55$; (6) visits by white persons: $P' = 0.88$.
[3]Number of sample physicians = 300; number of sample patients = 16,000.
[4]Number of sample physicians = 115; number of sample patients = 6,000.
[5]Number of sample physicians = 90; number of sample patients = 5,000.
[6]Number of sample physicians = 60; number of sample patients = 3,200.

Table B. Expected precision of some sample estimates based on medical data from the National Ambulatory Medical Care Survey pilot study by selected domains

Domain of interest	*Approximate relative error of estimate* $V_{P'}$
ALL PHYSICIANS[1]	
Size of estimate P'[2]	Percent
0.05	6.54
0.06	6.99
0.08	9.68
0.13	6.06
0.16	7.63
0.20	5.62
0.40	2.29
0.53	2.54
GENERAL PRACTICE[3]	
Size of estimate P'[2]	
0.05	12.80
0.06	10.34
0.08	13.57
0.13	8.31
0.16	9.22
0.20	7.47
0.40	3.84
0.53	3.63
GENERAL SURGERY[4]	
Size of estimate P'[2]	
0.05	20.50
0.06	15.20
0.13	9.08
0.16	11.92
0.20	10.36
0.40	5.93
0.53	5.26
OBSTETRICS AND GYNECOLOGY[5]	
Size of estimate P'[2]	
0.05	23.11
0.06	16.96
0.08	21.49
0.13	12.99
0.16	12.92
0.20	11.41
0.40	6.65
0.53	5.85
PEDIATRICS[6]	
Size of estimate P'[2]	
0.05	28.30
0.06	20.41
0.13	15.55
0.16	15.05
0.20	13.58
0.53	7.01

[1]Number of sample physicians = 1,200; number of sample patients = 65,000.
[2]These proportions refer to the following characteristics of patients' visits: (1) Secondary diagnosis of diseases of the circulatory system: $P' = 0.05$; (2) primary diagnosis of diseases of the upper respiratory system: $P' = 0.06$; (3) patient problem of delivery and complications of pregnancy: $P' = 0.08$; (4) laboratory tests for screening purposes: $P' = 0.13$; (5) specimen taken by physician's staff: $P' = 0.16$; (6) disposition to return to physician only as needed: $P' = 0.20$; (7) age of patient 15–44 years: $P' = 0.40$; (8) patient administered drug treatment: $P' = 0.53$.
[3]Number of sample physicians = 300; number of sample patients = 16,000.
[4]Number of sample physicians = 115; number of sample patients = 6,000.
[5]Number of sample physicians = 90; number of sample patients = 5,000.
[6]Number of sample physicians = 60; number of sample patients = 3,200.

Sampling plan

General description of NAMCS sampling plan

NAMCS is based on a three-stage, stratified, cluster design. The first-stage unit, or primary sampling unit (PSU), was a county (or parish or independent city) or group of contiguous counties, except for some PSU's in New England which were groups of townships. The sample PSU was selected with a probability proportional to population size. The second-stage unit was a physician who was selected with a probability inversely proportional to the number of physicians in sample PSU's such that the product of the first- and second-stage probabilities is a constant. The third-stage unit was a visit to a physician's office for ambulatory medical care. The probability of selection was based on the number of patients a doctor expected to see during the randomly selected week assigned to the doctor for reporting visits, and the number of days during the week that the doctor expected to see patients. The sampling rate, which ranges from unity for small practices to one in five for large practices, yields an average sample of about 30 patients per physician.

Each year during 1974–81, the sample included 87 PSU's, 3,000 total office-based physicians, about 1,925 responding physicians, and about 51,000 ambulatory patient visits. In 1973, the sample was somewhat smaller—about 1,100 responding, in-scope physicians; this sample was large enough to generate some useful statistics, yet small enough to allow a careful cost-effective evaluation of the survey methodology. The sample was redesigned in 1985 to make possible separate estimates in each of the physician specialties. The 1985 sample included 84 PSU's and about 5,000 total physicians, 2,900 responding physicians, and 70,000 patient visits. Beginning in 1989 the plan is to modify the 1985 design, with expected samples of about 112 PSU's, 2,500 total physicians, 1,600 responding physicians, and 42,000 visits, annually.

1973–81 (1973) design of NAMCS

Optimum allocation of sample

The estimator is based on a three-stage sample design, and each stage of the design contributes to the variance of the estimator. Thus the sampling variance can be minimized if the proper allocation of the samples can be made.

The estimator for the variance of a unit mean $\bar{\bar{x}}_k$ for the kth physician specialty group is

$$\sigma^2_{\bar{\bar{x}}_k} = \frac{1-f_1}{a}\hat{S}^2_{1x_k} + \frac{1-f_2}{m}\hat{S}^2_{2x_k} + \frac{1-f_3}{n}\hat{S}^2_{3x_k}$$

where f_1, f_2, and f_3 are the first-, second-, and third-stage sampling fraction, respectively; a is the total number of sample PSU's; m is the total number of physicians in the sample; and n is the total number of patient visits in the sample. $\hat{S}^2_{1x_k}$, $\hat{S}^2_{2x_k}$, and $\hat{S}^2_{3x_k}$ are estimated population variances for the three stages of sampling.

Using this formula and NAMCS pilot study data, sampling errors were computed based on three PSU sample sizes, five physician samples sizes, and four patient visit sample sizes as shown in tables C and D. This format was used for a variety of statistics as shown in the footnotes of tables A and B.

The patterns were similar for each of these statistics. The relative sampling errors for the proportion of visits with general history taken and the proportion of visits with primary diagnosis of disease of upper respiratory system are shown as examples in tables C and D, respectively.

As indicated in both tables, the overwhelming portion of the variance is contributed by the between-physician and between-PSU components.

Table C. Relative sampling errors for proportion of visits with general history of patient taken ($P' = 0.28$) by number of PSU's and physicians based on National Ambulatory Medical Care Survey pilot study data

Number of PSU's and physicians	*Number of sample visits*			
	1,000	*2,000*	*6,000*	*12,000*
	Relative sampling error in percent			
Simple random sampling.....	5.12	3.62	2.09	1.48
50 PSU's				
50 physicians	17.50	17.43	17.39	17.38
100 physicians	13.37	13.28	13.22	13.20
200 physicians	10.72	10.61	10.53	10.51
600 physicians	8.50	8.36	8.27	8.25
1,200 physicians.....	...	7.72	7.62	7.57
100 PSU's				
50 physicians	16.84	16.76	16.71	16.70
100 physicians	12.47	12.38	12.31	12.30
200 physicians	9.58	9.45	9.37	9.35
600 physicians	7.02	6.85	6.73	6.70
1,200 physicians.....	...	6.02	5.89	5.85
200 PSU's				
50 physicians	16.48	16.40	16.36	16.34
100 physicians	11.99	11.90	11.83	11.81
200 physicians	8.95	8.82	8.73	8.70
600 physicians	6.13	5.93	5.80	5.76
1,200 physicians.....	...	4.98	4.79	4.75

NOTE: PSU is primary sampling unit.

Table D. Relative sampling errors for proportion of visits with primary diagnosis of disease of upper respiratory system ($P' = 0.06$) by number of PSU's and physicians based on National Ambulatory Medical Care Survey pilot study data

Number of PSU's and physicians	Number of sample visits			
	1,000	2,000	6,000	12,000
	Relative sampling error in percent			
Simple random sampling	12.41	8.78	5.07	3.59
50 PSU's				
50 physicians	23.13	22.77	22.54	22.48
100 physicians	17.58	17.12	16.80	16.72
200 physicians	14.01	13.42	13.01	12.91
600 physicians	11.00	10.24	9.70	9.56
1,200 physicians	...	9.27	8.67	8.52
100 PSU's				
50 physicians	22.60	22.24	21.99	21.93
100 physicians	16.88	16.39	16.06	15.98
200 physicians	13.11	12.48	12.05	11.93
600 physicians	9.84	8.98	8.36	8.20
1,200 physicians	...	7.86	7.15	6.96
200 PSU's				
50 physicians	22.30	21.93	21.69	21.62
100 physicians	16.48	15.98	15.64	15.55
200 physicians	12.59	11.93	11.48	11.36
600 physicians	9.13	8.20	7.51	7.33
1,200 physicians	...	6.96	6.13	5.91

NOTE: PSU is primary sampling unit.

When there are relatively few physicians in the sample, for example, fewer than 200, there is not an appreciable improvement in the precision by increasing the number of PSU's in the sample. However, when the sample contains 600 or more physicians, the gain from increasing the number of PSU's becomes more evident. For example, if the sample contains 100 physicians and 6,000 patients, the estimated gain in precision from increasing the number of PSU's from 50 to 200 ranges from about 3 percent to about 20 percent for primarily medical data (patient visit characteristics shown in table B); the estimated gain increases from about 10 percent to about 15 percent for primarily nonmedical data (characteristics shown in table A). Now, for a sample of 1,200 physicians and 6,000 patients the estimated gain is about 35 to 40 percent for nonmedical data and about 40 to 60 percent for medical data. For the statistics shown in these tables the between-patient or within-physician variance components appear to have a negligible effect on the total variance. The number of sample patient visits for the first year of the survey was expected to be about 65,000 (but, in fact, it was much smaller—around 40,000 visits). In succeeding years, the number of sample visits was reduced to about 50,000 expected visits per year and the number of physicians was increased to 3,000, or about 1,925 responding physicians. This design is much closer to optimum than that used in the first year of the survey.

The precision that can be expected from the survey, based on a sample of 87 PSU's, 1,600 physicians, 1,200 respondents, and about 65,000 patient visits, is shown in tables A and B.

The tables provide some insights into the amount of detail that might be expected from a survey of 1,200 physicians and 87 PSU's. One interpretation of the information in table A is that a condition affecting about 15 percent of the patients within a domain containing 5 percent of the total sample of physicians can be estimated with a relative error of about 20 percent. In table B, a condition affecting some 6 percent of the patients within the 5-percent physician domain could be estimated with a 20-percent relative error.

These sampling errors are probably larger than those computed for the national survey, because a poststratified estimator is used for the national survey and a simple inflation estimator was assumed for the estimates based on the NAMCS pilot study. Also, sampling errors of national estimates should be reduced by the use of the ratio estimator for proportions, where both the numerator and denominator are random variables, if the correlation between the random variables is 0.5 or larger.

Stratification and selection of PSU's, 1973 design

The first-stage sample for NAMCS was designed by the National Opinion Research Center (NORC) of the University of Chicago as part of its 1972 Master Probability Sample (King and Richards, 1972). The NORC master sample was designed primarily for general purpose population surveys where the ultimate sampling unit is a person.

Techniques were employed that assured design efficiency, such as stratification, clustering, and sampling with probability proportional to size.

The first stage of the NORC design was adapted to NAMCS for economic reasons because NORC was chosen as the contractor to conduct the survey and already had a trained field staff located in the selected PSU's. Also, because there is a high correlation between population and the number of physicians in an area, there should be little loss in design efficiency by selecting the first-stage sample with probability proportional to population size rather than proportional to the number of physicians in an area.

The NORC first-stage sample design was also used in the design of the 1973–81 National Survey of Family Growth. A description of that design has been published (NCHS, 1979) and is summarized here for the reader's convenience.

The sampling frame consisted of 246 Standard Metropolitan Statistical Areas (SMSA's) as defined by the U.S. Bureau of the Census in March 1971 and all other counties, parishes, independent cities, and townships in the conterminous United States. Each of these geographic areas will be referred to as PSU's.

The SMSA frame was ordered as follows:

- Within each of the nine census geographic divisions as shown in figure 1, the SMSA's were sorted into three size groups, 1 million persons or more, 200,000 to 999,999, and fewer than 200,000 persons.
- In divisions 1, 2, 3, 4, 8, and 9 the SMSA's in each size class were sorted by State with States ordered geographically from northwest to northeast to southeast to southwest. Within each State the SMSA's were ordered the same way. In divisions 5, 6, and 7, the SMSA's in each size class were put in descending order of the number of residents other than white persons, to increase the chance

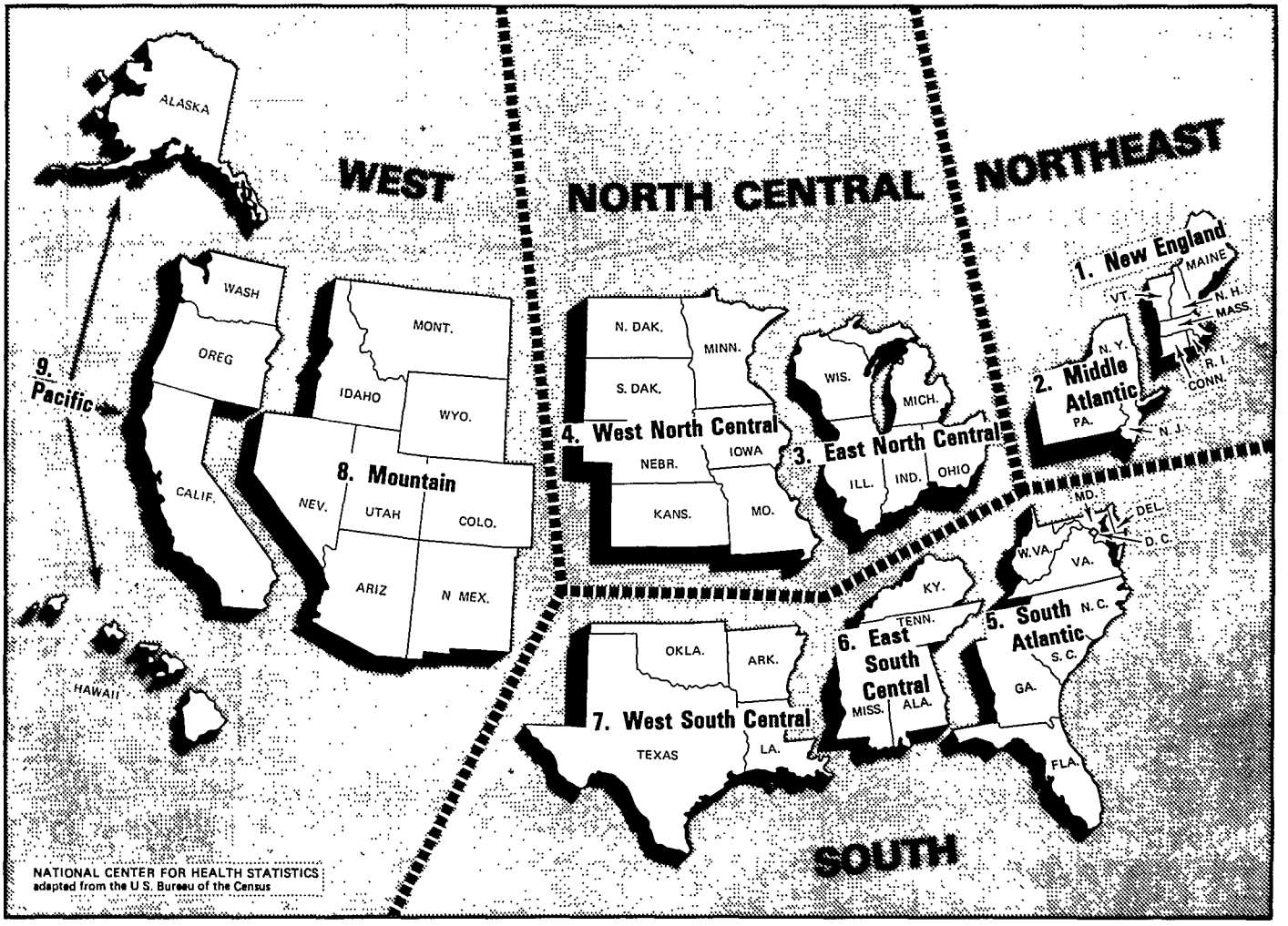


Figure 1. Census geographic divisions of the United States, and their order for first-stage sample of 1973–81 design

of selecting an appropriate number of southern SMSA's with large black populations.

The total population of the SMSA frame was 138,789,636, based on preliminary 1970 census data. The next step was to divide the frame into 139 sequential zones of 1 million people each. (The last zone contained 789,636 individuals and 210,364 "blanks.") In each zone the numbers from 1 to 1 million were assigned in ordered intervals to the SMSA's that were totally or partially included, each SMSA receiving an interval equal to its population within the zone. A "hit" number between 1 and 1 million was randomly and independently selected for each zone, and the SMSA whose assigned interval included that number was the PSU selected to represent the zone.

After the PSU's were selected, the 139 zones were systematically separated into four random groups. Two of the groups were combined to form 56 distinct SMSA's, which served as the SMSA sample for NAMCS. The remaining two groups were held in reserve by NORC to be used as needed.

Under this scheme, a large SMSA could be spread across several zones and, therefore, be selected more than once. By the conventional method of probability proportional to size (PPS), a random number determines each of the PSU's to be selected from each zone and the probability of selecting a PSU is simply the ratio of the PSU population to the zone or stratum population. For the scheme used for NAMCS, the method for determining the probabilities is not exactly PPS and is more complicated. The procedure is described in appendix II.

The outside SMSA sampling frame consisted of a total population of 63,456,729 according to the 1970 census. The frame included all areas of the conterminous United States not classified as an SMSA. These areas can be classified as counties, parishes, independent cities, or townships. PSU's in this frame consisted of individual areas outside of SMSA's or groups of areas outside of SMSA's. Small areas outside of SMSA's were linked with other areas that were not SMSA's to form PSU's with minimum size of 10,000 persons.

The outside SMSA frame was stratified by census division, PSU population size, percent of the population living in urban areas, and percent of the population that consisted of black persons; the PSU's were sorted in the frame as shown in figure 2.

Similar to the procedure described for the SMSA frame, 64 zones of 1 million people each were formed except that the last zone contained 543,271 "blanks." The method of selecting a PSU to represent a zone was the same as that described for SMSA's. However, the method for dividing the PSU's into four random groups was different.

The 64 zones were divided into 16 sets of 4 consecutive

Census division	Population of PSU	Percent urban	Percent black	Order of PSU's in frame
1 and 2	≥50,000	-	-	DOP
	<50,000	-	-	DOP
3 and 4	≥60,000	-	-	Sort by DOP within State
	30,000–59,999	≥40	-	Sort by DOP within State
		<40	-	Sort by DOP within State
	<30,000	<50	-	Sort by DOP within State
		≥50	-	Sort by DOP within State
5, 6, and 7	≥30,000	≥30	<20	Sort by DOP within State
			≥20	Sort by DOP within State
		<30	<20	Sort by DOP within State
			≥20	Sort by DOP within State
	<30,000	-	<20	Sort by DOP within State
		-	≥20	Sort by DOP within State
8 and 9	Any size	-	-	Sort by DOP

NOTE: DOP is descending order of population size.

Figure 2. Frame stratification for areas outside of SMSA's (Standard Metropolitan Statistical Areas) and PSU (primary sampling unit) sort in frame

zones each. Each set was randomly permuted, and the PSU's representing the first two zones in the sequence were the ones used in NAMCS. Only 31 PSU's were selected from the outside SMSA frame because a blank was selected as the "hit number" in zone 64. Thus the first-stage, outside SMSA sample consisted of 31 distinct PSU's with none of the PSU's appearing in the sample more than once.

When added to the SMSA sample, these 31 PSU's produced a total first-stage sample of 87 distinct localities.

Redesign of 1973 outside SMSA sample

One important constraint on the NAMCS design is that physicians are not asked to participate in the survey more often than every third year. This had a large effect on the first-stage NAMCS design, especially for PSU's with very few physicians. Prior to this redesign which was implemented in 1976, the physician universe was exhausted in 20 of the 31 outside SMSA PSU's. Rather than continue replacing these PSU's on an annual basis, as needed, it was decided to redesign and reselect the entire outside SMSA sample.

The basic features of the new design were that three samples of outside SMSA PSU's would be rotated annually. One sample would be used only every third year, ensuring that no private or group practice physician would be contacted more frequently than every 3 years, unless the doctor transferred from one sample PSU to another.

The sampling frame and stratification procedures were the same as used for the original outside SMSA sample. A sample of 93 PSU's was selected with a probability proportional to size.

After the 93 PSU's were selected, they were divided into 31 groups of 3 consecutively listed PSU's. The order within each group was randomized. Then within each group, the first PSU (according to the randomized order) was assigned to the first sample, the second PSU to the second sample, and the third PSU to the third sample.

Last, the three samples were randomly assigned to the year that they would be in the survey, beginning with 1976 and continuing through 1981 when the survey was discontinued.

Second-stage selection and assignment of physicians

After the PSU's were selected in the sample, a sample of physicians was chosen from each sample PSU with a probability inversely proportional to the probability of selecting a PSU so that the joint probability of selecting a PSU and a physician was approximately a constant for all PSU's.

The second-stage design involved the development of the sampling frame of physicians, stratification and ordering of the file, and the process of selecting the sample, which are described next.

Development and evaluation of the sampling frame—The sampling frame consisted of the combined master files of the American Medical Association (AMA) and the American Osteopathic Association (AOA).

The AMA master file contains current and historic data on all physicians in the United States, including members and nonmembers of the association, and graduates of foreign medical schools who are in the United States and meet education standards for primary recognition as physicians.

A file is started on each individual upon entry into medical school or, in the case of foreign or Canadian medical graduates, upon entry into the United States. In addition to other information, the file contains current address, type of practice (patient care, nonpatient care), specialties, and present employment (medical schools, hospitals, government, and other).

Every 4 years, questionnaires are sent to all physicians residing in the United States as well as to all U.S. physicians residing temporarily overseas. Data collected include professional activity, specialization, and current employment. Between census years, a computerized weekly update system keeps the master file current. Each physician's record is updated as information becomes available to reflect changes in address, specialty, or professional activity. Sources of information indi-

cating change include AMA mailing lists or publication requests, and correspondence to physicians, hospitals, government agencies, medical societies, specialty boards, and licensure agencies. Depending on the source of information and type of change, a questionnaire may be sent to the physician to verify the change (Roback et al., 1984).

AOA uses a similar procedure for maintaining its master file which includes the names of all persons who attend and graduate from osteopathic schools. AOA also keeps its master file current by conducting surveys of persons in the file every 18 months. In 1985, 21,000 people were counted.

Several studies have been conducted to evaluate the completeness or coverage of the AMA file, but none was conducted recently and none is comprehensive.

As part of a study in 1972 to evaluate NCHS, a four-county study was conducted to evaluate the coverage of the AMA master file (Committee to Evaluate the National Center for Health Statistics, 1972). The physicians in each county were identified through State and county medical society membership lists, personal interviews with local authorities, hospital and medical school lists, and telephone directories. The overall error rate comparing all physicians located in the four counties with their master file entries was approximately 3 percent "overcoverage," assuming that the number of physicians found in the four counties was "truth."

In 1975, another study on the completeness of the AMA master file was conducted by the staff of *Medical Marketing and Media* (1975). Data in the master file for four adjacent counties in Connecticut were compared with a list of physicians that was determined from hospital rosters, county medical society membership lists, and telephone directories. Of 329 physicians in the master file, all but 7 were verified as practicing in the area.

A study was conducted in 1974 by AMA, comparing a list of physicians licensed to practice medicine in Washington State with the AMA master file for Washington State (Cherkin and Lawrence, 1977). Of 5,467 physicians on the Washington licensure file, 302 or 5.2 percent were not in the master file.

In 1980 an intensive national study was conducted by NCHS to determine the volume of office-based ambulatory care provided by non-office-based physicians included in the AMA and AOA files. The study indicated that the NAMCS estimates undercount the number of office visits to physicians of all types by about 10 percent (NCHS, 1984).

Stratification of sampling frame—The AMA and AOA computer tape files did not have the same formats, so it was necessary to process them separately before they were combined to select the sample.

For the AMA files within each PSU, physicians were sorted into groups according to specialty. Then they were sorted by primary specialty within each group and ordered alphabetically within each specialty. These groups and corresponding specialties are as follows:

- *General practice:* General practice, family practice.
- *Medical specialties:* Allergy, broncho-esophagology, cardiovascular diseases, dermatology, diabetes, endocrinology, gastroenterology, hematology, infectious diseases, internal medicine, neoplastic diseases, nephrology, nutrition, pediatrics, pediatric allergy, pediatric cardiology, pulmonary diseases, rheumatology.
- *Surgical specialties:* Abdominal, cardiovascular, colon and rectal, general, gynecology, head and neck, hand, laryngology, neurological, obstetrics and gynecology, obstetrics, ophthalmology, orthopedic, otology, otorhinolaryngology, pediatric, plastic, rhinology, traumatic, thoracic, urology.
- *Other specialties:* Aerospace medicine, child neurology, child psychiatry, geriatrics, general preventive medicine, hypnosis, legal medicine, neurology, occupational medicine, other specialties (those not designated as one already listed), psychiatry, clinical pharmacology, public health, physical medicine and rehabilitation, psychoanalysis, psychosomatic medicine, unspecified.

For the AOA file all physicians not listed as being engaged in private practice were removed from the file as were those physicians practicing outside the conterminous United States. (This step was not necessary for the AMA tape because such physicians had been edited from the tapes by AMA.)

AOA physician specialty coding was different from AMA's, so recoding was necessary to make the two files compatible. AOA physicians were assigned to "general practice" if they spent 50 percent or more of their time in general practice, but they were assigned to a specialty if they spent 50 percent or more of their time in that specialty. Ineligible or out-of-scope specialists were removed from the file, and those AOA specialists not falling into one of the specific specialty categories were assigned to "other specialty."

Within each PSU the physicians were arranged into specialty groups, into specialties within groups, and alphabetically within each specialty.

The two tapes were then combined into a single file, with ordering by PSU, specialty groups within PSU's, specialties within specialty groups, and alphabetically within specialties. A check of names and addresses was made to identify and remove duplicates from the file before the sample was selected.

Selection of sample physicians—From the final sorted file of physicians listed in the 87 PSU's, a sample was chosen in such a way that the overall probability of any physician's being selected was approximately a constant for each PSU. Because all PSU's were not selected with the same probabilities, physicians were selected at differential rates within PSU's, in an attempt to preserve the overall constant sampling rate.

The sample was selected systematically within each PSU across all specialties, with the frame ordered as described above. This helped to ensure coverage of as many different specialties in the sample as possible for each PSU. The sampling interval for PSU h was

$$k_h = \frac{M_h}{m_h},$$

where M_h and m_h are population and designated sample sizes, respectively.

A random start between 1 and k_h was chosen, and the physician in the position of the random start in the file and every k_hth physician thereafter was chosen. In some small PSU's the sampling interval was unity. In this situation all

physicians in the PSU were taken, even though m_h may have been smaller than M_h (as integers).

Thus a sample of 1,609 physicians was selected in the 87 PSU's for the 1973 NAMCS.

In 1974, the sample was increased to about 3,000 physicians and remained at that level through 1981.

Assignment of physicians to panels—After the sample was selected in PSU, specialty, and alphabetical order, the sample physicians were numbered serially. Then they were assigned to two random subsets, or panels. Panel 1 consisted of physicians whose serial numbers ended with an odd digit and panel 2 consisted of physicians whose serial numbers ended in even digits. The division of the sample into panels was done originally to provide a ready mechanism for research studies. The panels have also been used in the computation of sampling variances.

Assignment of physicians to reporting periods—The assignment to reporting periods was carried out independently by panel, and each of the two panels was randomly distributed over all weeks of the year.

Each panel of physicians was divided into 52 random groups. This was done systematically by assigning the first listed physician in the panel and every 52d thereafter to a group, the second listed and every 52d thereafter to a second group, and so on until all 52 groups were formed.

The 52 groups were randomly ordered and assigned to calendar weeks according to the random ordering. For example, the first group of physicians was assigned to week number 39, the week beginning September 24, 1973 (survey weeks are considered to run from Monday through Sunday).

Sampling of patients, 1973 design

A sequential listing of all patient visits made to a physician's office during the randomly assigned reporting week served as the sampling frame. Individual sample physicians were assigned a sampling rate based on their expected work load during their assigned week, that is, the expected number of patient visits and the number of days that they expected to see patients at in-scope locations.

The sampling rates were so designed that each physician would complete about 30 patient record forms during a reporting week. Physicians expecting 10 or fewer visits per day completed a patient record for all visits. Those expecting more than 10 visits per day completed a form for every second, third, or fifth visit, based on their assigned sampling interval. For log forms with sampling intervals less than unity, a random start was provided on the first page of the log so that predesignated sample visits recorded on each succeeding page of the log provided a systematic random sample of patient visits during the reporting period. A facsimile of the patient record and patient log is shown in appendix I.

The information needed by interviewers to determine the sampling frame for physician visit sampling rates, and to assure complete coverage of the sample physician's office practice, was collected in an induction interview with the physician. The interview form is shown in appendix I. Also, a chart is shown on the form which relates the physicians' expected work loads during their assigned week and the appropriate patient log form.

1985 redesign of NAMCS

Overview of redesign

As mentioned before, NAMCS was conducted as a continuing survey between 1973 and 1981. Due to fiscal constraints, the survey was discontinued in 1982. It was next conducted in 1985 with, at that time, a planned periodicity of every 3 years. This lull in the survey provided an opportunity to conduct research to improve the survey and to redesign the sample to make it more efficient.

The first stage of the sample was designed jointly by the Survey Research Center (SRC) of the University of Michigan and NORC as part of their 1980 national sample (Heeringa, Connor, and Darrah, 1986). This sample was used rather than the one designed specifically for NAMCS because NORC, as in previous years, was selected as the contractor to conduct the 1985 NAMCS. It was more economical to use the SRC-NORC design because a trained field staff was already employed in the PSU's that were to be used for NAMCS.

The second-stage sample of physicians and the overall sample design of NAMCS were the responsibility of NCHS staff (Tompkins and Shimizu, 1985).

The principal changes made in the new design were in the way that the PSU's and physicians were selected, the sample size, and the ability to produce detailed statistics for a number of medical and surgical specialists. In the 1985 design, the sample consisted of 84 PSU's with 1 PSU per stratum selected with a probability proportional to the number of occupied housing units in the PSU and stratum. The 1973–81 design included 87 PSU's based on a modified probability proportional to size (PPS) design in which 1 PSU was selected independently from each "paper stratum" of 1 million people.

The 1985 first-stage sample employed a controlled selection technique which helped ensure a balanced distribution of the sample to census geographic divisions, or similar geographic divisions of census regions. The 1973 design did not include this feature.

In 1985, the second-stage sample was stratified into 14 physician specialty groups and optimally allocated to minimize cost of the survey for a fixed precision in estimates. Precise estimates were required for each physician specialty. Thus the sample was significantly increased from about 1,925 responding physicians and 51,000 patient visits annually for the period 1974–81 to targeted sample sizes of about 5,000 total physicians, 3,200 responding physicians, and 80,000 patient visits in 1985. The earlier sample was not designed to make estimates for physician specialties, based on 1 year of data collection. However, because the earlier survey was continuous, it was possible to make estimates by combining data for more than 1 year of collection. Because in 1985 the plan was to conduct the survey only every 3 years, combining samples would not be feasible.

Increasing the sample to 5,000 physicians assured, on the basis of the prior design, about the same analytic detail as for earlier surveys, with about the same precision obtained from combining 2 years of data collection.

Optimum sampling plan for 1985 survey

The sample was specified as a three-stage design with sampling of PSU's, physicians, and ambulatory patient visits. Also, the number of PSU's was fixed at 84, because that was the number of PSU's in the then-current master sample of NORC, the data collection contractor for NAMCS in 1985. The number of sample physicians was also fixed at 5,000 because that is the number that would have been needed in a single year to obtain the desired precision for the specialties under the prior sample design and, hence, the number included in the contract when survey funds had to be committed before sample design research was completed. Even so, it is useful to assume that the numbers of PSU's, physicians, and patients are variables and were determined in a way to minimize the cost of the survey to achieve specified precision in estimates.

In research for the optimum design, separate estimates were assumed for each of the 14 largest physician specialties or specialty groupings. These specialties are general and family practice, internal medicine, pediatrics, general surgery, obstetrics and gynecology, orthopedic surgery, cardiovascular disease, dermatology, urology, psychiatry, neurology, ophthalmology, otorhinolaryngology, and allergy. It was desired that estimates for 1 percent of the visits to any specialty have relative standard errors (RSE's) no greater than 30 percent and that the precision of larger proportions of visits be reasonable. Reasonable precision was interpreted as including 10 percent RSE's for estimates of 10 percent or more of the visits and 15 percent RSE's for estimates of 5–9 percent of the visits in each specialty.

Formulas for the required sample sizes at each stage were derived using the method of Lagrange multipliers. The goal was to allocate resources to the various stages of sampling to minimize cost for specified levels of precision. For the NAMCS research, precision was fixed at four levels of relative errors (0.30, 0.15, 0.10, and 0.05). Also, except for PSU's outside of SMSA's when PSU's were stratified by SMSA status, costs of adding a PSU to the sample were assumed to be roughly 5.7 times the cost of adding a sample physician; the cost of adding a sample physician was assumed to be 77.6 times the cost of a sample visit. For the outside-SMSA PSU strata, the corresponding cost ratios were assumed to be 4.0 and 90.0, respectively. The cost assumptions and research are based on the 1980 NAMCS when data were collected from 2,284 physicians and 46,081 patient visits. The results of that research have been published (Tompkins and Shimizu, 1985) and are summarized here.

Sample sizes of PSU's, physicians, and patient visits were determined for each of 11 selected visit characteristics and 14 physician specialties. The visit characteristics used in the research are as follows:

- Patient is given drug treatment.
- Patient is 15–44 years of age.
- Clinical laboratory test is done for visit.
- Disposition is to return to physician if needed.
- Diagnostic services are general history and examination.
- Principal diagnosis is disease of the respiratory tract.
- No diagnostic services are performed during visit.
- Principal diagnosis is disease of the circulatory system.
- Psychotherapy or therapeutic listening is employed.
- Diagnostic service is mental status examination.
- Endoscopy is performed.

Table 1 gives the optimum sample sizes needed to produce visit statistics with the specified RSE's. For each specialty, four groups of visit characteristics were defined on the basis of the frequency with which they occurred in the specialty, that is, characteristics occurring in 0–4, 5–9, 10–49, and 50 or more percent of the visits.

For the analysis, a stratified design was used for the six largest physician specialties where PSU's were stratified by SMSA status. An unstratified design was used for the remaining specialties due to insufficient clustering among sample doctors in the 1980 NAMCS to produce better estimates for the components of variances. Based on other computations done for a design in which it is assumed that independent estimates are to be produced for SMSA and outside-SMSA areas, it appears that most of the remaining eight physician specialties are clustered in SMSA's and that a stratified design will likely require smaller sample sizes than those shown for the unstratified design.

Table 2 condenses table 1 by presenting the largest sample sizes required across frequency groups to produce desired precision levels.

The figures in table 2 were used as a guide in designing the 1985 NAMCS sample. The number of PSU's range from 24 for pediatricians to 113 for cardiovascular disease specialists. All but 5 specialty groups require fewer PSU's than the 84 in the SRC-NORC national sample. The range of the optimum is usually quite broad, and it is likely that precision in estimates based on 84 PSU's would not be very different from precision based on 100 PSU's.

For the physician sample, the numbers of physicians given in table 2 were inflated by the specialty response rates experienced in 1981 to yield the minimum sample sizes to be selected from each specialty. These minimum sizes were then used to determine the portion of the total sample to be selected from each specialty.

After sample design research was completed, some changes were made in the survey specifications regarding allergists and osteopaths. The requirement to make separate estimates for allergists was dropped because of small numbers in the universe. On the other hand, separate estimates for osteopaths were added because of their relatively large numbers and their importance in the medical care system. To accommodate this new requirement in the sample design, it was assumed that the variation in osteopath practices is similar to that for general practitioners; hence, the research results on sample size for the general practitioners were used to design the sample of osteopaths.

For the "visit" sample, the average number of visits per physician varies across specialties from 16 to the maximum allowed of 30. To simplify operations, the average number of visits per physician was set at 30 for each stratum.

Stratification and selection of PSU's

The universe for the SRC-NORC national sample consists of all of the households in the 48 conterminous States of

the United States and the District of Columbia. It excludes the military population and people in institutions. The sampling frame was composed of all of the counties, parishes, townships, and independent cities in the United States, which will be referred to as counties for convenience. From these counties, primary sampling units (PSU's) were formed, consisting of SMSA's, single counties outside SMSA's, and groups of counties outside SMSA's. PSU's outside SMSA's were formed from single counties outside SMSA's of more than 4,000 people. Counties outside SMSA's with less than 4,000 population (about 60 in all) were linked when possible to a geographically contiguous county to form a multicounty, non-SMSA PSU of the required minimum size. The sampling frame was developed from data collected in the 1980 decennial census, with county level records the principal elements of the data base, including total population, population counts for race and ethnic subgroups, total housing, population of the largest city, counts of cities and towns in several size categories, and occupied housing unit counts. This information was useful for stratification of PSU's and for determining the probabilities of selecting the sample of PSU's.

PSU's were selected, utilizing sampling with probability proportional to size, the measure of size being the number of occupied housing units in a PSU.

Based on their measure of size, 16 SMSA's were designated as self-representing (SR) because each constitutes an entire stratum. The New York SMSA is the largest of the SR PSU's with a 1980 total population of 9,119,737 and a 1980 occupied housing unit count of 3,498,354. The smallest SR PSU is Atlanta, with a 1980 population of 2,029,618 and 719,799 occupied housing units. Table E provides population and housing data for each of the 16 SR PSU's.

Table F shows how the 16 self-representing PSU's and 68 non-self-representing PSU's are distributed by geographic region and type of PSU. Because 16 areas were designated as self-representing, their allocation to census regions is fixed. The allocation of the 68 non-self-representing selections is based on the total measure of size (in occupied housing units) for each cell of the SMSA status by region grid.

Across regions, the rounding of fractional sample size expectations for the numbers of SMSA and outside-SMSA PSU's was done subjectively, favoring somewhat the SMSA's of the South region, which can be expected to experience continued population growth into the 1980's. After rounding fractional strata expectations to integers, the average stratum size for individual region by SMSA status cells varies somewhat about the expected mean value of 857,500 occupied housing units (table G).

Because the basic design called for the selection of one PSU from each stratum, the allocation of primary sampling

Table E. Self-representing primary sampling units for SRC-NORC (Survey Research Center-National Opinion Research Center) 1980 national sample and their population and 1980 occupied housing units

Standard Metropolitan Statistical Area	*1980 census decennial counts in sample PSU's*	
	Population	*Occupied housing units*
All self-representing areas	61,268,728	22,066,249
New York, NY-NJ	9,119,737	3,498,354
Los Angeles-Long Beach, CA	7,477,657	2,730,674
Chicago, IL	7,102,328	2,486,302
Philadelphia, PA-NJ	4,716,818	1,639,330
Detroit, MI	4,352,762	1,508,832
San Francisco-Oakland, CA	3,252,721	1,281,316
Washington, DC-MD-VA	3,060,240	1,112,579
Dallas-Ft. Worth, TX	2,974,878	1,076,320
Houston, TX	2,905,350	1,027,067
Boston, MA	2,763,357	990,660
Nassau-Suffolk, NY	2,605,813	809,120
St. Louis, MO-IL	2,355,276	837,797
Pittsburgh, PA	2,263,894	828,504
Baltimore, MD	2,174,023	756,980
Minneapolis-St. Paul, MN-WI	2,114,256	762,615
Atlanta, GA	2,029,618	719,799

Table F. Allocation of SRC-NORC (Survey Research Center-National Opinion Research Center) 1980 national primary sampling units (PSU's) and their population and occupied housing units, by region and type of PSU

Region and type of PSU	*Sample PSU's*	*1980 census decennial counts in sample PSU's*	
		Population	*Occupied housing units*
All regions	*Number*	*Number in thousands*	
All PSU's	84	226,505	80,377
Self-representing	16	61,269	22,066
Non-self-representing	68	165,236	58,310
SMSA's	45	108,136	38,424
Outside SMSA's	23	57,100	19,886
Northeast			
All PSU's	16	49,137	17,470
Self-representing	5	21,470	7,766
Non-self-representing	11	27,667	9,704
SMSA's	8	20,272	7,121
Outside SMSA's	3	7,395	2,584
North Central			
All PSU's	22	58,854	20,856
Self-representing	4	15,925	5,596
Non-self-representing	18	42,929	15,261
SMSA's	11	25,783	9,174
Outside SMSA's	7	17,147	6,087
South			
All PSU's	31	75,349	26,479
Self-representing	5	13,144	4,693
Non-self-representing	26	62,205	21,786
SMSA's	16	37,209	13,204
Outside SMSA's	10	24,997	8,582
West			
All PSU's	15	43,165	15,571
Self-representing	2	10,730	4,012
Non-self-representing	13	32,435	11,559
SMSA's	10	24,874	8,926
Outside SMSA's	3	7,561	2,633

NOTES: Population values may not add exactly to totals due to independent rounding.

SMSA is Standard Metropolitan Statistical Area.

Table G. Numbers of expected and allocated non-self-representing (NSR) primary sampling units (PSU's) and average stratum size, by major domains of the SRC-NORC (Survey Research Center-National Opinion Research Center) 1980 national sample

Domain	Expected NSR PSU's	Allocated NSR PSU's	Average stratum size[1]
All domains	68.0	68	857,500
SMSA's	44.8	45	853,873
Northeast	8.3	8	890,117
North Central	10.7	11	883,988
South	15.4	16	825,235
West	10.4	10	892,573
Outside SMSA's	23.2	23	864,609
Northeast	3.0	3	861,176
North Central	7.1	7	869,577
South	10.0	10	858,215
West	3.1	3	877,762

[1] Number of occupied housing units.

NOTE: SMSA is Standard Metropolitan Statistical Area.

units by region was fixed by the strata assignments shown in table G. Also, once the allocation was fixed, stratification and selection of PSU's proceeded independently within each region.

Stratification of PSU's—The principal criterion for stratifying PSU's was the census region in which the PSU is located.

Some PSU's such as Wilmington, DE, and a number of SMSA's along the Ohio and Mississippi Rivers span regional boundaries. In these instances, the entire SMSA was assigned to the region containing the principal city of the SMSA. (The four census regions are shown in figure 1.)

The second major stratification variable is the SMSA status of the PSU. Within census regions, separate strata of SMSA and non-SMSA PSU's were formed. Table G shows the number of strata formed in each cell for the two-way primary stratification matrix.

The South region was divided into two subregions: The "Deep South" which includes South Carolina, Georgia, Alabama, Mississippi, and Louisiana; and the remaining States in the region. Thus stratification of PSU's was done independently within five "regions." (The South was originally subdivided in the 1970's at the request of the Center for Political Studies (CPS), a companion unit to the Survey Research Center of the University of Michigan (SRC) in the Institute for Social Research. The subdivision allows the inclusion of a proportionate sample of cases from the Deep South States, an outcome important in CPS analysis of National Election Study data. Because the subdivision does not cause "real" inefficiencies, the subdivision was retained in the SRC-NORC 1980 National Sample.)

In general, strata were about the same size, that is, they contained about the same number of occupied housing units according to the 1980 census. Sometimes this was difficult to achieve, especially for very large SMSA's that were too small to be self-representative. In these instances two large strata were collapsed to form a "double stratum." From these double strata, two PSU's were selected with PPS and without replacement.

Controlled selection of PSU's—Controlled selection of PSU's was used in the sampling process to assure a proportionate distribution of the sample of PSU's to census divisions or similar subdivisions of census regions. This objective, alone, should justify the use of controlled selection. However, the procedure is expected to achieve modest reductions in sampling errors because the control variables have the effect of stratification.

The procedure used by SRC-NORC is described in appendix III.

Sampling of individual PSU's—The controlled selection steps described in appendix III do not in general identify actual PSU selections, but they do produce a set of restrictions that must be followed in choosing a sample PSU from each non-self-representing stratum. For example, in the West census region the controlled selection outcome specifies that the stratum 83 sample PSU will be from the Mountain division and further it must be located in the State of Colorado. Given these constraints, the choice of a sample PSU simply involves listing, or otherwise identifying, the full set of Colorado PSU's in stratum 83 and choosing one with probability proportionate to its occupied housing unit measure of size (MOS). With the appropriate controlled selection constraints on the selection process, the same PPS procedure was used to identify the sample PSU in each of the remaining 67 non-self-representing strata.

For the special case of a double stratum, two systematic PPS selections were made. Because the measure of size for each PSU in the double stratum was less than one-half the MOS value for the stratum itself, systematic PPS selections constitute a sampling without replacement.

A PSU's probability of selection under this PPS procedure is calculated as the ratio of its MOS to that of the complete stratum from which it was chosen. It is important to note that the controlled selection process does not in any way influence a PSU's overall probability of being selected from a stratum, that is, selection probabilities are calculated just as they are under any stratified PPS sampling procedure. In other words, for this sample, the probability of a PSU's selection is that PSU's number of occupied housing units divided by the corresponding number of housing units in the stratum. Probabilities at the primary stage of this multistage design span a considerable range, from a minimum value of 0.003028 to a maximum value of 1.0 for the 16 self-representing PSU's.

Second-stage design for physician sample

The procedure for selecting the sample of physicians for the 1985 design was similar to that described earlier for the 1973 NAMCS design. Again the current listings of physicians and osteopaths maintained by AMA and AOA served as the sampling frame for the 1985 NAMCS. As before, anesthesiologists, pathologists, and radiologists were excluded from the sampling frame, as were any physicians "not in office-based practice."

A major difference was that the 1985 design stratified physicians into specialty groups and utilized differential sampling rates to oversample certain specialists. In 1973, the number of specialists selected in the sample was proportional to the number of physicians in each specialty; that is, a constant sampling rate was applied for all specialties in each PSU.

The strata were formed as indicated in table H. First, the physicians were stratified into 15 specialty classes and 2 SMSA status classes, or 30 groups. Because of the small number of physicians in areas outside of SMSA's, all physicians in those PSU's were selected in the sample.

Next, within specialty classes in SMSA PSU's, physicians were ordered by PSU and (sub)region. Within PSU's, physicians were ordered by subspecialty. Then the residual group of "all other subspecialties" was divided into medical, surgical, and other specialties and deep stratified within these groups' subspecialty.

To minimize the number of sampling rates, the 15 physician specialty strata were collapsed into 5 groups so that the preliminary sampling rates for individual strata in the group were approximately the same. The preliminary sampling rates were the ratio of sample physicians needed to the total number of physicians in the 15 individual specialties in the sampled PSU's according to the AMA and AOA files.

Physicians were selected at the second stage with a systematic sampling scheme and with probabilities that were inversely related to the probability of PSU selection. Table H shows the proportions of physicians that were selected for the 1985 NAMCS in the sampled SMSA PSU's.

The sample of physicians was first arrayed by the 15 specialties. Within specialty, the physicians were arrayed by strata with the SMSA stratum first. Then, within each stratum the physicians were arrayed in the order that they were selected in the sample and numbered serially from 1 to n, the size of the sample.

The sampled physicians were then assigned to two panels according to the last digit of their serial number. One panel was formed from even terminal digits, and the other from odd terminal digits.

Physicians were assigned to reporting weeks independently by panel, with each panel randomly distributed over all weeks of 1985. Within each panel, 52 groups of physicians were formed systematically, that is, the first group consists of the first and every 52d physician thereafter in the panel. The second group consists of the second and every 52d thereafter, and so forth. Lastly, the 52 groups of physicians were randomly assigned to weeks of the year.

Physicians were assigned to panels and reporting weeks in the same way as in the 1973 design. The procedure is described in more detail in the section "Second-stage selection and assignment of physicians."

Table H. Physician specialty strata and sampling rates within sampled Standard Metropolitan Statistical Areas for the 1985 National Ambulatory Medical Care Survey

Strata group and specialty	*Sampling rate*[1]
Group 1	
All group strata	1 in 44
General and family practice	1 in 42
General surgery	1 in 44
Obstetrics and gynecology	1 in 45
All other specialties	1 in 45
Group 2	
All group strata	1 in 13
Osteopathy	1 in 13
Ophthalmology	1 in 13
Urology	1 in 13
Group 3	
All group strata	1 in 72
Internal medicine	1 in 76
Pediatric and adolescent medicine	1 in 72
Group 4	
All group strata	1 in 59
Psychiatry	1 in 59
Group 5	
All group strata	1 in 23
Cardiovascular diseases	1 in 24
Dermatology	1 in 23
Orthopedic surgery	1 in 22
Otorhinolaryngology	1 in 22
Neurology	1 in 23

[1] Rates for individual specialties are preliminary rates. Rates for groups of specialties are the final rates used in the 1985 design.

1989 NAMCS design

Beginning in 1989 NAMCS will be continuous again. Also, NAMCS will represent all 50 States and the District of Columbia instead of only the conterminous United States, as was done in prior NAMCS cycles.

Beginning in 1989 NAMCS will also be primarily conducted in a subset of the PSU's selected for the National Health Interview Survey (NHIS). In 1989–94, the first-stage sample for NAMCS will be 112 PSU's in two panels of the 1985–94 NHIS PSU sample. Each of the NHIS PSU panels is a probability sample representative of the entire United States. The sample for NHIS is described in detail by Massey (NCHS, in preparation). As in prior cycles, the PSU's for the 1989 NAMCS design consist of counties (or parishes or independent cities) or groups of counties except for some PSU's in New England and Hawaii that are formed from groups of townships. The primary stratifying variables for PSU's are census region and PSU size (largest, medium, small). The sample PSU's were selected using probability proportional to size (1980 census population) with the 12 largest PSU's being selected with certainty.

The second-stage sample of physicians in the 1989 NAMCS will be selected using the methodology described earlier in the section "Second-stage design for physician sample" for the 1985 design. The proportions of physicians in the sample, however, will be adjusted to yield a total sample of 2,500 physicians. Subject to the total of 2,500, the number of sample physicians from each specialty will be based on the precision levels and response rates obtained in the most recent year of NAMCS and on differing priorities, if any, placed on the estimates for the individual specialties.

The sample of physicians will be assigned to two physician panels and to reporting weeks with each physician panel randomly distributed over all weeks in the year. The procedure for making the assignments will be the same as that used to assign physicians to reporting weeks in the 1973–81 NAMCS (see "Second-stage selection and assignment of physicians").

NAMCS estimation procedure

Estimator for visit statistics

The NAMCS estimator for patient visit characteristics is an inflation estimator poststratified by the number of physicians in a specialty class.

The basic estimator of an aggregate parameter Y for all visits to physicians who are primarily in office-based practice is

$$\hat{Y} = \sum_{\alpha}^{L} R_{\alpha} \sum_{h} \hat{Y}_{\alpha h} \qquad (1)$$

where α = physician specialty group (for the 1973 design these groups were General Practice, Internal Medicine, Pediatrics, Other Medical Specialties, General Surgery, Obstetrics and Gynecology, Other Surgical Specialties, Psychiatry, and All Other; for the 1985 design the groups are General and Family Practice, Osteopathy, Internal Medicine, Pediatric and Adolescent Medicine, General Surgery, Obstetrics and Gynecology, Orthopedic Surgery, Cardiovascular Disease, Dermatology, Urology, Psychiatry, Neurology, Ophthalmology, Otorhinolaryngology, and All Other Specialties),

h = PSU stratum (for the 1973 design $h = 1$; for the 1985 design $h = 1$ or 2 with strata defined by SMSA status (SMSA or non SMSA),

L = total number of specialty groups,

$\hat{Y}_{\alpha h}$ = weighted nonresponse adjusted estimator of $Y_{\alpha h}$,

$Y_{\alpha h}$ = aggregate of visits with selected characteristic to physicians in the αhth specialty PSU stratum,

$$R_{\alpha} = \frac{M_{\alpha}}{\hat{M}_{\alpha}},$$

M_{α} = total number of office-based non-Federal physicians in the αth specialty group in the AMA and AOA files,

$$\hat{M}_{\alpha} = \sum_{i} W_{1i} M_{\alpha i},$$

the weighted estimate for the number of office-based physicians in the αth specialty based on sampled PSU's,

i = a sample PSU,

$M_{\alpha i}$ = the total number of office-based physicians in the αth specialty and ith PSU in the AMA and AOA files,

W_{1i} = the reciprocal of the probability of selecting the ith PSU.

The weighted nonresponse adjusted estimator for $Y_{\alpha h}$ is

$$\hat{Y}_{\alpha h} = A_{2\alpha h} \sum_{i \varepsilon h} \Lambda_{\alpha i} A_{1\alpha i} C_{2\alpha i} \sum_{j} 52 W_{\chi \alpha i} \lambda_{\alpha ij} C_{1\alpha ij} \sum_{k} Y_{\alpha ijk} \qquad (2)$$

where ε = membership and, hence, $i \varepsilon h$ is the set of PSU's in the hth PSU stratum,

j = a sample physician or physician listing,

k = a sample patient visit,

$Y_{\alpha ijk}$ = measure of interest for the αijkth sample visit,

$$C_{1\alpha ij} = \begin{cases} N_{\alpha ij}/n_{\alpha ij} & \text{for the 1973 design,} \\ C^{*}_{1\alpha ij} & \text{for the 1985 design,} \end{cases}$$

(adjustment for visit nonresponse within the αijth physician's practice),

$N_{\alpha ij}$ = the total visits made to the αijth physician in the physician's reporting week,

$n_{\alpha ij}$ = the number of sample visits with records in the data file for the αijth physician,

$$C^{*}_{1\alpha ij} = \begin{cases} N_{\alpha ij}/n_{\alpha ij} & \text{if } N_{\alpha ij}/n_{\alpha ij} \leq 11, \\ 11 & \text{otherwise} \end{cases}$$

(this adjustment assumes the maximum sampling rate of 1 visit per 10 is rare),

$$\lambda_{\alpha ij} = \begin{cases} 1 & \text{if } n_{\alpha ij} \text{ is nonzero,} \\ 0 & \text{otherwise,} \end{cases}$$

$W_{\chi \alpha i} = W_{1i} W_{2\alpha i}$
(the overall sampling weight of the αijth physician),

$W_{2\alpha i} = W_{2\alpha ij}$
(the reciprocal of the probability for selecting the jth physician from the αth specialty and ith PSU; that is, the second-stage weight is constant across physicians within the αth specialty class and ith PSU),

$$C_{2\alpha ij} = \begin{cases} 1 & \text{for the 1973 design,} \\ \dfrac{\sum_j \lambda_{\alpha ij} N_{\alpha ij}}{\sum_j \lambda_{\alpha ij} C_{1\alpha ij} n_{\alpha ij}} & \text{for the 1985 design,} \end{cases}$$

(the adjustment across physicians within the ith PSU and αth specialty group for excess visit nonresponse within physician practices),

$$A_{1\alpha i} = \begin{cases} m'_{\alpha i}/\dot{m}_{\alpha i} & \text{for the 1973 design,} \\ A^*_{1\alpha i} & \text{for the 1985 design,} \end{cases}$$

(the adjustment within the ith PSU for physician nonresponse in the αth specialty group),

$$m'_{\alpha i} = \sum_j \gamma_{\alpha ij}\omega_{\alpha ij}$$

(the number of in-scope sample physicians seeing patients in their selected reporting week),

$$\dot{m}_{\alpha i} = \sum_j \lambda_{\alpha ij}$$

(the number of αth specialty group physicians from the ith PSU with nonzero $n_{\alpha ij}$),

$$A^*_{1\alpha i} = \begin{cases} m'_{\alpha i}/\dot{m}_{\alpha i} & \text{if } m'_{\alpha i}/\dot{m}_{\alpha i} \le 2, \\ 2 & \text{otherwise} \end{cases}$$

(the adjustment within the ith PSU for physician nonresponse in the αth specialty group),

$$\gamma_{\alpha ij} = \begin{cases} 1 & \text{if the } \alpha ij\text{th physician is in scope for the NAMCS,} \\ 0 & \text{otherwise,} \end{cases}$$

$$\omega_{\alpha ij} = \begin{cases} 1 & \text{if the } \alpha ij\text{th physician saw patients during the physician's reporting week,} \\ 0 & \text{otherwise,} \end{cases}$$

$$\Lambda_{\alpha i} = \begin{cases} 1 & \dot{m}_{\alpha i} \text{ is nonzero,} \\ 0 & \text{otherwise,} \end{cases}$$

$$A_{2\alpha h} = \begin{cases} \sum_i m'_{\alpha i} \Big/ \sum_i \Lambda_{\alpha i} m'_{\alpha i} & \text{for the 1973 design,} \\ \hat{M}'_{\alpha h}/\dot{M}_{\alpha h} & \text{for the 1985 design} \end{cases}$$

(the adjustment across PSU's for excess physician nonresponse within PSU's for the αhth specialty PSU stratum),

$$\hat{M}'_{\alpha h} = \sum_{i \varepsilon h} W_{\chi\alpha i} m'_{\alpha i}$$

(an estimate of in-scope physicians in the αhth specialty PSU stratum, based on physicians who saw patients in their reporting week),

$$\dot{M}_{\alpha h} = \sum_{i \varepsilon h} W_{\chi\alpha i} A_{1\alpha i} \dot{m}_{\alpha i}$$

(the initial nonresponse adjusted estimate for in-scope physicians seeing patients in the αhth specialty PSU stratum).

Estimator for physician aggregate statistics

The estimator of an aggregate parameter X for all physicians who are primarily in office-based practice during the year is the post-ratio-adjusted estimator

$$\hat{X} = \sum_{\alpha}^{L} \dot{R}_{\alpha} \sum_{h} \hat{X}_{\alpha h} . \qquad (3)$$

The weighted nonresponse adjusted estimator of X for physicians in the αth specialty group and the hth PSU stratum is

$$\hat{X}_{\alpha h} = B_{2\alpha h} \sum_{i \varepsilon h} \tau_{\alpha i} B_{1\alpha i} W_{\chi\alpha i} \sum_j X_{\alpha ij} . \qquad (4)$$

The terms and symbols not earlier defined are as follows:

$X_{\alpha ij}$ = the measure of the characteristic of interest for the αijth sample physician,

$$B_{1\alpha i} = \begin{cases} m''_{\alpha i}/\mathring{m}_{\alpha i} & \text{for the 1973 design,} \\ B^*_{1\alpha i} & \text{for the 1985 design,} \end{cases}$$

$$m''_{\alpha i} = \sum_j \gamma_{\alpha ij}$$

(the number of in-scope physicians in the sample, including those who saw no patients in their selected reporting week),

$$\mathring{m}_{\alpha i} = \sum_j \zeta_{\alpha ij}$$

(the number of sampled in-scope physicians completing the physician interview, including those who may not have completed visit records),

$$B^*_{1\alpha i} = \begin{cases} m''_{\alpha i}/\mathring{m}_{\alpha i} & \text{if } m''_{\alpha i}/\mathring{m}_{\alpha i} \le 2, \\ 2 & \text{otherwise,} \end{cases}$$

$$\zeta_{\alpha ij} = \begin{cases} 1 & \text{if the } \alpha ij\text{th sample physician completed the physician interview,} \\ 0 & \text{otherwise,} \end{cases}$$

$$\tau_{\alpha i} = \begin{cases} 1 & \mathring{m}_{\alpha i} \text{ is nonzero,} \\ 0 & \text{otherwise,} \end{cases}$$

$$B_{2\alpha h} = \begin{cases} 1 & \text{for the 1973 design,} \\ M''_{\alpha h}/\mathring{M}_{\alpha h} & \text{for the 1985 design} \end{cases}$$

(the adjustment across PSU's within strata for excess physician nonresponse within PSU's),

$$M''_{\alpha h} = \sum_{i \varepsilon h} W_{\chi\alpha i} m''_{\alpha i}$$

(the estimated number of in-scope physicians of αhth specialty PSU stratum based on the full physician sample),

$$\mathring{M}_{\alpha h} = \sum_{i \varepsilon h} B_{1\alpha i} W_{\chi \alpha i} \mathring{m}_{\alpha i}$$

(the initial nonresponse adjusted estimate for number of physicians in the αhth specialty PSU stratum based on respondents to the physician interview only).

Variance estimation and presentation procedures

Estimation of variances

The balanced half-sample replication procedure was used to compute sampling variances for NAMCS. This procedure was developed by Professor Philip J. McCarthy (NCHS, 1966).

The NAMCS application of the procedure employs a 48×48 orthogonal matrix, consisting of 48 pseudo PSU's and 48 balanced half-samples.

For the procedure to work, each stratum must contain at least two PSU's. Because only one PSU was selected from each stratum by NORC (except for a few "double strata"), it was necessary to collapse strata to form pseudo strata of multiple PSU's.

Each pseudo stratum was defined to consist of the following:

- An individual self-representing PSU, or
- Two PSU's selected from "double strata," or
- The successive two or three PSU's in an array of the remaining PSU's within each region and SMSA status group.

In SMSA status groups within a region, the PSU's were arrayed by stratum size. Three PSU's were assigned to a single pseudo stratum only when there was an odd PSU in an SMSA status group within a region. Assignment of the odd PSU to a stratum with two PSU's eliminates that part of the replicate variance that would result from pairing the odd PSU with no PSU's in the replicate samples. PSU's outside of SMSA's were not paired with SMSA PSU's because SMSA status was a stratification variable at both the first and the second stages of sampling.

Within each self-representing stratum two PSU's were formed by dividing the sample physicians in the SMSA into two random groups. This was done at the time the physician sample was divided into two independent panels as described in the section "Second-stage selection and assignment of physicians," for 1973–81 design.

In the non-self-representing strata, pseudo strata were formed by pairing PSU's. Pairing of PSU's was a subjective process to make the strata that they represent as similar as possible. There were a few pseudo strata that contained three PSU's. In these situations, two of the PSU's were randomly paired to form a single pseudo PSU. This pseudo PSU and the remaining actual PSU represented the pseudo stratum.

The final weights for each replicate half-sample were computed using the equations shown in the section on the NAMCS estimation procedure, adjusted by the factors shown in table J. The factors in table J adjust for the varying degrees of representation from various pseudo strata in the half-samples, a variability that could bias the variance estimates if the adjustments were not made. When these adjustments are applied, weighted totals for both the pseudo PSU's in each replication stratum properly represent the entire pseudo stratum. Adjustments for nonresponse were computed for each replicate.

Table J. Adjustments to inflation weight for pseudo primary sampling units (PSU's) in replicate half-samples for variance computation in the National Ambulatory Medical Care Survey design

Pseudo stratum and pseudo PSU	*Inflation weight*	
	W_{1i}	$W_{2\alpha i}$
Self-representing pseudo strata, all pseudo PSU's...	1	2
Non-self-representing pseudo strata, 3 PSU's:		
Pseudo PSU combines 2 PSU's.....	3/2	1
Pseudo PSU is single PSU.....	3	1
Non-self-representing pseudo strata, 2 PSU's:		
All pseudo PSU's.....	2	1

To construct a variance estimate for a statistic $\hat{Y}$, it is necessary to compute 48 half-sample estimates $\hat{Y}_r$. The estimates $\hat{Y}_r$ are analogous to $\hat{Y}$, except that the weight specifications are applied to the sample of respondents in the rth half-sample and the inflation weight adjusted according to table J. The variance of $\hat{Y}$ is then estimated by

$$S_{\hat{Y}}^2 = \frac{1}{48} \sum (\hat{Y}_r - \hat{Y})^2 . \qquad (5)$$

This procedure is also used when estimating the variance of estimates for physician characteristics, proportions, ratios, or other estimates. The desired estimate is computed for each replicate sample, and their mean square deviation is taken about the estimate from the whole sample.

When the sampling errors for statistics based on NAMCS are computed by these procedures, the sampling distribution for the statistics is approximated by a noncentral "t" distribution with at most 47 degrees of freedom.

Presentation of variances

The sampling variance is a function of the sample design, the variability and prevalence of the statistic in the population,

and the estimator employed. Generally, each statistic derived from a survey has its own unique sampling error. Thus, because a publication may contain thousands of different estimates, computing the sampling error for every statistic would be time consuming and very costly.

Fortunately, many of the sampling errors will be of similar size because they are influenced by similar levels of clustering, stratification, and other design features.

Instead of presenting variances for every statistic, the variances can be grouped and averages obtained for the group. Bean (NCHS, 1974a) described procedures for grouping the variances and obtaining the group averages. In forming the groups, two basic principles should be kept in mind: Survey characteristics such as prevalences of any diseases represented in a group should have similar design effects, and the groups should cover the possible range of variation of the data.

Empirically, it has been shown that there is a relationship between the size of the estimate and the estimate's relative variance (the ratio of the sampling variance and the square of the estimate). The relationship is expressed by the formula

$$V_X^2 = a + \frac{b}{X'} .$$

Thus, using the relative variances of the selected statistics of a group, values for a and b are calculated. Then a smooth relative standard error curve can be drawn. From this curve the reliability of any estimate falling into the group can be approximated.

The standard method of estimating a and b is the method of least squares. The least squares estimators give values of a and b that minimize the sum of squares of deviations between the observed values $\hat{V}_X^2$, and the predictions V_X^2. Thus,

$$S = \sum_i \left(V_{X_i}^2 - a - \frac{b}{X_i'} \right)^2$$

is minimized. The exact estimators that will minimize S are found by differentiating the sum with respect to a and b and equating to zero. However, the method used here was to minimize the squared relative residuals of V_X^2. Here the quantity

$$S' = \sum_i \left(\frac{V_{X_i}^2 - a - b/X_i'}{V_{X_i}^2} \right)^2$$

is the one minimized.

These formulas have the unknown value term $V_{X_i}^2$ in them. Hence, an iterative procedure was employed. Substituting $V_{X_i}^2$ for $V_{X_i}^2$, values a_1 and b_1 were computed and then used to calculate

$$\hat{V}_{X_i}^2 = a_1 + \frac{b_1}{X_i'} .$$

Next, new estimates a_2 and b_2 were computed, using $\hat{V}_{X_i}^2$ for $V_{X_i}^2$.

$$\text{If } \left| \frac{a_2 - a_1}{a_2} \right| \geq 2 \text{ percent} \quad \text{or} \quad \left| \frac{b_2 - b_1}{b_2} \right| \geq 2 \text{ percent,}$$

the process was repeated with

$$\hat{V}_{X_i}^2 = a_2 + \frac{b_2}{X_i'}$$

replacing $V_{X_i}^2$. Iterations were run until

$$\left| \frac{a_j - a_{j-1}}{a_j} \right| \leq 2 \text{ percent} \quad \text{and} \quad \left| \frac{b_j - b_{j-1}}{b_j} \right| \leq 2 \text{ percent.}$$

This curve-fitting methodology is illustrated using 1980 NAMCS estimates of the number of office visits to physicians. First, relative sampling errors (RSE's) were computed for 260 different statistics, the estimates ranging from as many as 576 million visits to as few as 110,000 visits. These 260 point estimates of the number of visits were relatively evenly distributed all along this range.

Curve A in figure 3 was fitted to these 260 point estimates using the methodology just described.

The equation on which the curve in figure 3 is based is

$$Y = 0.001650 + \frac{36.364326}{x}$$

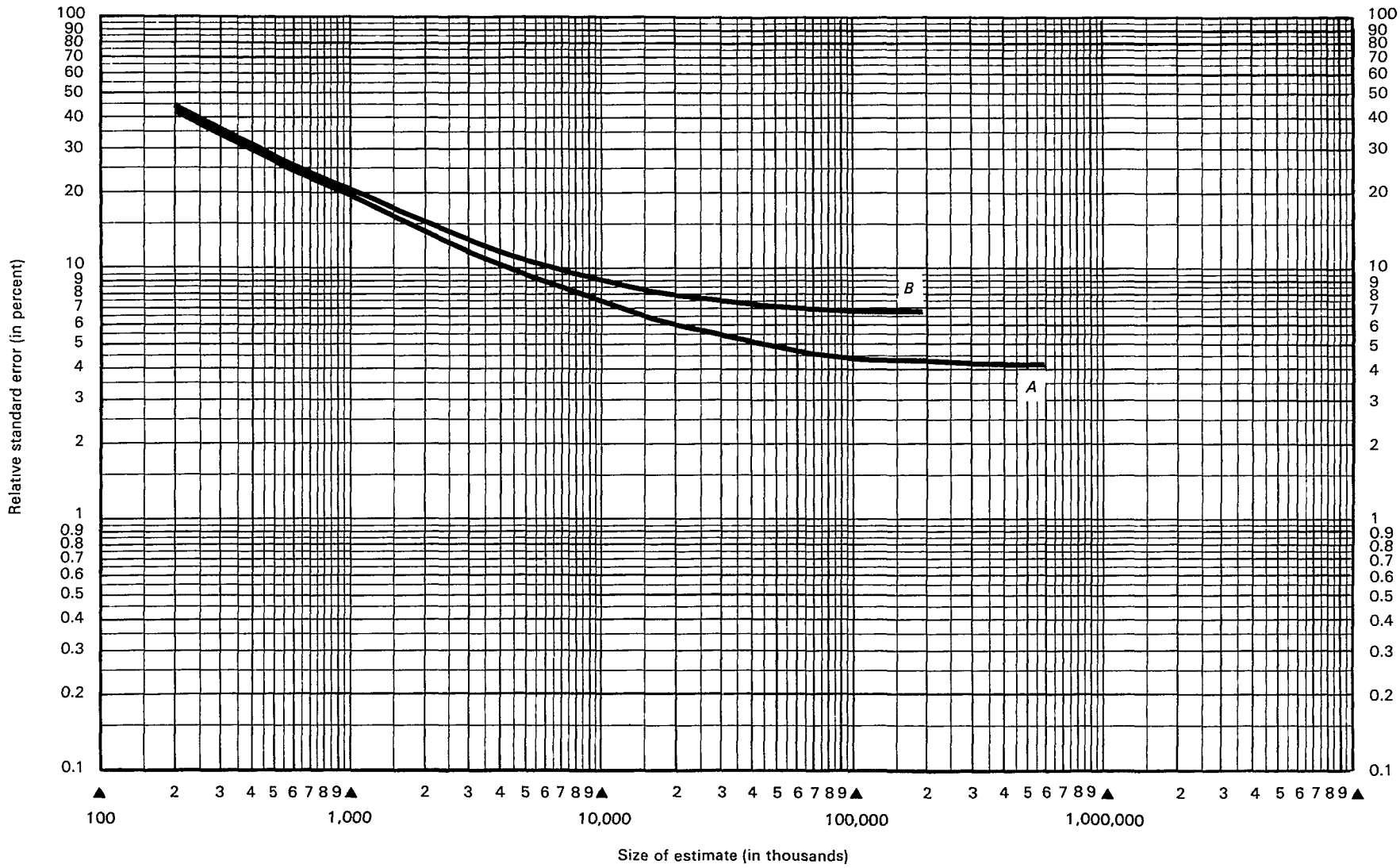
where Y = "Predicted" relvariance of Y,

$a = 0.001650$,

$b = 36.364326$, and

x = estimated number of visits.

In this particular application, four iterations were required to minimize the function S'. The ratio of the final a to the previous one was 1.005 and the ratio of the final b to the previous one was 0.997.



Example of use of chart: An estimate of 20 million office visits to general surgeons (read from scale at bottom of chart) has a relative standard error of 7.9 percent (read from curve *B* on scale at left of chart) or a standard error of 1,580,000 office visits (7.9 percent of 20 million office visits).

Figure 3. Approximate relative standard errors for estimated numbers of office visits based on all physician specialties (*A*) and individual specialties (*B*), 1980 National Ambulatory Medical Care Survey

References

Cherkin, D., and D. Lawrence. 1977. An evaluation of the AMA physician masterfile as a data source—one State's experience. *Med. Care* IV (9):767–79.

Committee to Evaluate the National Center for Health Statistics, Subcommittee on Statistical Methodology. 1972. Report on AMA Physician File Study (SR–16). Unpublished document. Washington.

Goodman, R., and L. Kish. 1950. Controlled selection—a technique in probability sampling. *J. Am. Stat. Assoc.* 45:350–72.

Heeringa, S. G., J. H. Connor, and D. C. Darrah. 1986. The 1980 SRC/NORC National Sample Design and Development. Unpublished document. Sampling Section, Institute for Social Research, University of Michigan, Ann Arbor.

King, B. F., and C. Richards. 1972. The 1972 NORC Probability Sample. Unpublished document. National Opinion Research Center, Chicago.

Medical Marketing and Media. 1975. How good is the American Medical Association masterfile of physicians. *Medical Marketing and Media* 10(11):15–7.

National Center for Health Statistics, P. J. McCarthy. 1966. Replication, an approach to the analysis of data from complex surveys. *Vital and Health Statistics.* Series 2, No. 14. PHS Pub. No. 1000. Public Health Service. Washington: U.S. Government Printing Office.

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

National Center for Health Statistics, J. B. Tenney, K. L. White, and J. W. Williamson. 1974b. National Ambulatory Medical Care Survey, background and methodology, United States, 1967–72. *Vital and Health Statistics.* Series 2, No. 61. DHEW Pub. No. (HRA) 76–1335. Health Resources Administration. Washington: U.S. Government Printing Office.

National Center for Health Statistics, D. K. French. 1979. National Survey of Family Growth, Cycle I, sample design, estimation procedures, and variance estimation. *Vital and Health Statistics.* Series 2, No. 76. DHEW Pub. No. (PHS) 78–1350. Public Health Service. Washington: U.S. Government Printing Office.

National Center for Health Statistics, G. J. Gardocki, T. McLemore, and J. E. DeLozier. 1984. The National Ambulatory Medical Care Complement Survey, United States, 1980. *Vital and Health Statistics.* Series 13, No. 77. DHHS Pub. No. (PHS) 84–1738. Public Health Service. Washington: U.S. Government Printing Office.

National Center for Health Statistics, J. T. Massey. In preparation. Sample design and estimation procedures for the National Health Interview Survey, 1985–94. *Vital and Health Statistics.* Series 2. Public Health Service, DHHS, Hyattsville, Md.

Roback, G., L. Randolph, D. Mead, and T. Pasko. 1984. *Physician Characteristics and Distribution in the U.S., 1984 Edition.* Chicago: Survey & Data Resources, American Medical Association.

Tompkins, L., and I. M. Shimizu. 1985. Sample design research for the 1985 NAMCS. *1985 Proceedings of the Survey Research Section of the American Statistical Association,* pp. 213–18.

List of detailed tables

Table 1. Sample size requirements for optimum stratified design for the 1985 National Ambulatory Medical Care Survey, by percent range of estimates, physician specialty, and sampling stage

Physician specialty stratum and sampling stage	*Percent range of estimate*			
	0–4[1]	*5–9*[2]	*10–49*[3]	*50+*[4]
General practitioners	Sample size if stratified by SMSA status			
PSU's	43	37	55	22
Physicians	195	255	264	28
Visits	5,833	4,321	4,660	927
Internal medicine				
PSU's	25	34	60	25
Physicians	88	67	205	104
Visits	2,619	1,991	3,519	1,595
Pediatrics				
PSU's	3	24	19	21
Physicians	39	90	25	21
Visits	748	1,456	746	650
General surgery				
PSU's	71	46	84	---
Physicians	75	87	209	---
Visits	2,272	2,360	3,712	---
Obstetrics and gynecology				
PSU's	21	---	54	25
Physicians	32	---	187	40
Visits	961	---	3,340	852
Psychiatry				
PSU's	77	68	50	35
Physicians	114	77	126	32
Visits	3,403	1,358	1,724	813
Cardiovascular diseases	Sample size if not stratified by SMSA status			
PSU's	39	66	193	271
Physicians	49	66	193	271
Visits	1,478	1,980	5,790	8,130
Dermatology				
PSU's	64	42	111	209
Physicians	64	42	170	209
Visits	1,920	1,260	5,109	6,270
Orthopedic surgery				
PSU's	35	40	156	---
Physicians	94	119	236	---
Visits	2,821	2,416	6,041	---
Ophthalmology				
PSU's	76	---	126	---
Physicians	216	---	181	---
Visits	4,749	---	4,414	---
Otorhinolaryngology				
PSU's	19	56	80	250
Physicians	19	56	117	250
Visits	573	1,680	3,516	7,500
Neurology				
PSU's	88	45	169	311
Physicians	57	29	110	201
Visits	1,706	872	3,274	6,029

NOTE—See footnotes and notes at end of table.

Table 1. Sample size requirements for optimum stratified design for the 1985 National Ambulatory Medical Care Survey by percent range of estimates, physician specialty, and sampling stage—Con.

Physician specialty stratum and sampling stage	Percent range of estimate			
	0–4[1]	5–9[2]	10–49[3]	50+[4]
Urology	Sample size if not stratified by SMSA status			
PSU's	95	166	169	254
Physicians	138	241	199	254
Visits	4,145	7,243	5,983	7,620
Allergy				
PSU's	---	74	189	308
Physicians	---	29	74	121
Visits	---	870	2,222	3,620

[1]Relative standard error of the estimates = 0.30.
[2]Relative standard error of the estimates = 0.15.
[3]Relative standard error of the estimates = 0.10.
[4]Relative standard error of the estimates = 0.05.

NOTES: PSU is primary sampling unit. SMSA is Standard Metropolitan Statistical Area.

Table 2. Largest sample size requirements for estimates for each sampling stage by physician specialty included in research for 1985 National Ambulatory Medical Care Survey

Physician specialty	PSU's	Physicians	Visits	Average number of visits per physician
All specialties	...	1,696	40,143	24
	Sample size if stratified by SMSA status			
General practice	43	255	5,833	23
Internal medicine	38	142	2,619	18
Pediatrics	24	90	1,456	16
General surgery	53	124	2,385	19
Obstetrics and gynecology	31	154	2,708	18
Psychiatry	77	114	3,403	30
	Sample size if not stratified by SMSA status			
Cardiovascular diseases	113	113	3,390	30
Dermatology	64	64	1,920	30
Orthopedic surgery	92	172	4,121	24
Ophthalmology	76	216	4,749	22
Otorhinolaryngology	56	56	1,680	30
Neurology	86	56	1,667	30
Urology	102	102	3,060	30
Allergy	98	38	1,152	30

NOTES: PSU is primary sampling unit. SMSA is Standard Metropolitan Statistical Area.

Appendixes

Contents

List of appendix tables

Appendix I
Principal forms used in survey

CONFIDENTIAL*
NORC-4413

BEGIN DECK 3
Form Approved
OMB No. 0937-0141
Expires: 9/30/86

FOR OFFICE USE ONLY:
(BATCH NO.)
5-6/
(LOG NO.)
7-10/

NATIONAL AMBULATORY MEDICAL CARE SURVEY

INDUCTION INTERVIEW

(Phys. ID Number)
1-4/

BEFORE STARTING INTERVIEW

1. ENTER PHYSICIAN I.D. NUMBER IN BOX TO RIGHT.

2. ENTER DATES OF ASSIGNED REPORTING WEEK IN Q.2, P.2.

TIME BEGAN: AM PM

Doctor, before I begin, let me take a minute to give you a little background about this survey.

Although ambulatory medical care accounts for nearly 90 percent of all medical care received in the United States, there is no systematic information about the characteristics and problems of people who consult physicians in their offices. This kind of information has been badly needed by medical educators and others concerned with the medical manpower situation.

In response to increasing demands for this kind of information, the National Center for Health Statistics, in close consultation with representatives of the medical profession, has developed the National Ambulatory Medical Care Survey.

Your own task in the survey is simple, carefully designed, and should not take much of your time. Essentially, it consists of your participation during a specified 7-day period. During this period, you simply check off a minimal amount of information concerning patients that you see.

Now, before we get into the actual procedures, I have a few questions to ask about your practice. The answers you give me will be used only for classification and analysis, and of course all information you provide is held in strict confidence.*

1. First you are a _____.
(ENTER SPECIALTY FROM CODE ON FACE SHEET LABEL.)

Is that right? Yes.....X
No.....(ASK A).....Y

A IF NO: What is your specialty (including general practice)?

_____
(Name of Specialty)

11-13/

*
The National Ambulatory Medical Care Survey is authorized by Congress in Title 42, United States Code, section 242K. It is a voluntary study and there are no penalties for refusing to answer any question. All information collected is confidential and will be used only to prepare statistical summaries. No information which will identify an individual or a physician's practice will be released

2. Now, doctor, this study will be concerned with the ambulatory patients you will see in your office during this week of (READ REPORTING DATES ENTERED BELOW).

_____/____ (that's a Monday) through _____/____ (that's a Sunday)
month date month date

Are you likely to see any ambulatory patients in your office during that week?

Yes.....(GO TO Q.3).....X

No.....(ASK A).....Y

A. IF NO: Why is that? RECORD VERBATIM, THEN READ PARAGRAPH BELOW.

Since it's very important, doctor, that we include any ambulatory patients that you do happen to see in your office during that week, I'd like to leave these forms with you anyway--just in case your plans change. I'll plan to check back with your office just before (STARTING DATE) to make sure, and I can explain them in detail then, if necessary.

GIVE DOCTOR THE A PATIENT RECORD FORMS AND GO TO Q.9, P.6.

3. A. At what office location will you be seeing ambulatory patients during that 7-day period? RECORD UNDER A BELOW AND THEN CODE B.

B. FOR EACH OFFICE LOCATION ENTERED IN A, CODE YES OR NO TO "IN SCOPE."

IN SCOPE (Yes)	OUT OF SCOPE (No)
Private offices	Hospital emergency rooms
Free-standing clinics (non-hospital based)	Hospital outpatient departments
Groups, partnerships	College or university infirmaries
Kaiser, HIP, Mayo Clinic	Industrial outpatient facilities
Neighborhood Health Centers	Family planning clinics
Privately operated clinics (except family planning)	Government-operated clinics (VD, maternal & child health, etc.)

IN CASE OF DOUBT, ASK: Is that (clinic/facility/institution) hospital based?

Is that (clinic/facility/institution) government operated?

C. Is that all of the office locations at which you expect to see ambulatory patients during that week?

Yes.....X

No.....Y

IF NO: OBTAIN ADDITIONAL OFFICE LOCATION(S), ENTER IN "A" BELOW, AND REPEAT.

	A. Office Location	B. In Scope? Yes	No
(1)	_____	1	0
(2)	_____	1	0
(3)	_____	1	0
(4)	_____	1	0

TOTAL IN-SCOPE LOCATIONS: [ ] 14/

IF ALL LOCATIONS ARE OUT OF SCOPE, THANK THE DOCTOR AND LEAVE.

4. A. During that week (REPEAT DATES), how many ambulatory patients do you expect to see in your office practice? (DO NOT COUNT PATIENTS SEEN AT OUT-OF-SCOPE LOCATIONS CODED IN 3-B.)

ENTER TOTAL UNDER "A" BELOW AND CIRCLE NUMBER CATEGORY ON APPROPRIATE LINE.

B. And during those seven days (REPEAT DATES IF NECESSARY), on how many days do you expect to see any ambulatory patients? COUNT EACH DAY IN WHICH DOCTOR EXPECTS TO SEE ANY PATIENTS AT AN IN-SCOPE OFFICE LOCATION.

CIRCLE NUMBER OF DAYS IN APPROPRIATE COLUMN UNDER "B" BELOW.

DETERMINE PROPER PATIENT LOG FORM FROM CHART BELOW. READ ACROSS ON "TOTAL PATIENTS" LINE UNDER "A" AND CIRCLE LETTER IN APPROPRIATE "DAYS" COLUMN UNDER "B."

THIS LETTER TELLS YOU WHICH OF THE FOUR PATIENT LOG FORMS (A, B, C, D) SHOULD BE USED BY THIS DOCTOR.

LOG FORM DESCRIPTION	A. Expected total patients during survey week.	B. Total days in practice during week.						
	ENTER TOTAL FROM Q. 4-A	18/						
A--Patient Record is to be completed for ALL patients listed on Log. 15-17/	[ ][ ][ ]	1	2	3	4	5	6	7
	1- 12 PATIENTS	A	A	A	A	A	A	A
	13- 25 "	B	A	A	A	A	A	A
B--Patient Record is to be completed for every SECOND patient listed on Log.	26- 39 "	C	B	A	A	A	A	A
	40- 52 "	C	B	B	A	A	A	A
	53- 65 "	D	C	B	B	A	A	A
	66- 79 "	D	C	B	B	B	A	A
C--Patient Record is to be completed for every THIRD patient listed on Log.	80- 92 "	D	D	C	B	B	B	B
	93-105 "	D	D	C	B	B	B	B
	106-118 "	D	D	C	C	B	B	B
	119-131 "	D	D	C	C	B	B	B
	132-145 "	D	D	D	C	C	B	B
*D--Patient Record is to be completed for every FIFTH patient listed on Log.	146-158 "	D	D	D	C	C	B	B
	159-171 "	D	D	D	C	C	C	C
	172-184 "	D	D	D	C	C	C	C
	185-197 "	D	D	D	D	D	D	D
	198-210 "	D	D	D	D	D	D	D
	211+ "	D	D	D	D	D	D	D

*In the rare instance the physician will see more than 500 patients during his assigned reporting week, give him two D Patient Log Folios and instruct him to complete a patient record form for only every tenth patient. Then you are to draw an X through the Patient Record on every other page of the two folio pads, starting with Page 1 of the pad. The physician then completes the Patient Log on every page, but completes the Patient Record on every second page.

5. FIND LOG FOLIO WITH APPROPRIATE LETTER AND CIRCLE LETTER, ENTER FIRST FOUR NUMBERS OF THE FORM AND NUMBER OF LINES STAMPED "BEGIN ON NEXT LINE" FOR THE B-C-D LOG FORMS (if no lines are stamped, enter "0") BELOW.

FOLIO		No. Lines Stamped "BEGIN ON NEXT LINE"	FOR OFFICE USE ONLY Number patient record forms completed.	
Letter	Number			
A				19-23/
B				24-26/
C				
D				

6. HAND DOCTOR HIS FOLIO AND EXPLAIN HOW FORMS ARE TO BE FILLED OUT. SHOW DOCTOR INSTRUCTIONS ON THE POCKET OF FOLIO, ITEMS 6,10 AND 13 ON CARDS IN POCKET OF FOLIO AND ITEM DEFINITIONS ON THE BACK OF FOLIO, TO WHICH HE CAN REFER AFTER YOU LEAVE.

EMPHASIZE THAT EVERY PATIENT VISIT EXCEPT ADMINISTRATIVE PURPOSE ONLY IS TO BE RECORDED ON THE LOG FOR ENTIRE REPORTING PERIOD. FOR EXAMPLE, IF A MEDICAL ASSISTANT GAVE THE PATIENT AN INOCULATION OR A TECHNICIAN ADMINISTERED AN ELECTROCARDIOGRAM AND THE PATIENT DID NOT SEE THE DOCTOR, THIS VISIT MUST STILL BE LISTED ON THE LOG.

RECORD VERBATIM BELOW ANY CONCERN, PROBLEMS OR QUESTIONS THE DOCTOR RAISES.

7. IF DOCTOR EXPECTS TO SEE AMBULATORY PATIENTS AT MORE THAN ONE IN-SCOPE LOCATION DURING ASSIGNED WEEK, TELL HIM YOU WILL DELIVER THE FORMS TO THE OTHER LOCATION(S). ENTER THE FORM LETTER AND NUMBER(S) AND NUMBER OF LINES STAMPED "BEGIN ON NEXT LINE" FOR THE B-C-D LOG FOR THOSE LOCATIONS BELOW, BEFORE DELIVERING FORM(S).

Location	FOLIO		No. Lines Stamped "BEGIN ON NEXT LINE"	FOR OFFICE USE ONLY Number patient record forms completed.	
	Letter	Number			
					27-31/
					32-34/
					35-39/
					40-42/
					43-47/
					48-50/

8. During the survey week (REPEAT EXACT DATES), will anyone be available to help you in filling out these records (at each IN-SCOPE location)?

Yes..... (ASK A) 1 51/
No.....2

A. IF YES: Who would that be?

RECORD NAME, POSITION AND LOCATION.

NAME	POSITION	LOCATION

PERSONALLY BRIEF EACH PERSON LISTED ABOVE.

EMPHASIZE THAT EVERY PATIENT VISIT DURING THE ENTIRE WEEK IS TO BE RECORDED ON THE LOG EXCEPT "ADMINISTRATIVE PURPOSE ONLY."

---

9. Do you have a solo practice, or are you associated with other physicians in a partnership, in a group practice, or in some other way?

Solo.....(GO TO Q.10).....1 52/
Partnership...(ASK A-C).....2
Group.....(ASK A-C).....3
_____ <--- Other...(SPECIFY AND ASK A-C).....4

IF PARTNERSHIP, GROUP, OR OTHER:

A. Is this a prepaid group practice? Yes...(ASK [1].....1 53/
No.....2

[1] IF YES TO A: What percent of patients are prepaid? _____ percent 54-56/

B. How many other physicians are associated with you?

NUMBER OF PHYSICIANS: _____ 57-59/

C. What are the specialties of the other physicians associated with you? (How many of these are there?)

	Specialty	Number of Physicians
(1)		
(2)		
(3)		
(4)		
(5)		

D. CIRCLE ONE:

All physicians in this partnership/group practice have the same specialty.....1 60/

More than one specialty in this partnership/ group practice.....2

10. Are you currently participating in any prepaid plans, such as HMO (Health Maintenance Organization), IPA (Independent Practice Association), and PPO (Preferred Provider Organization), or other prepaid plan?

		Yes	No	
(1)	HMO	1	0	61/
(2)	IPA	1	0	62/
(3)	PPO	1	0	63/
(4)	OTHER (SPECIFY)_____			
	_____	1	0	64/

BEGIN DECK 4

11. Now I have just one more question about your practice. (NOTE: IF DOCTOR PRACTICES IN LARGE GROUP, THE FOLLOWING INFORMATION CAN BE OBTAINED FROM SOMEONE ELSE.)

A. What is the total number of full-time (35 hours or more per week) employees of your (partnership/group) practice? Include persons regularly employed who are now on vacation, temporarily ill, etc. Do not include other physicians. RECORD ON BOTTOM LINE OF COLUMN A BELOW.

1. How many of these full-time employees are a . . . (READ CATEGORIES BELOW AS NECESSARY AND RECORD NUMBER OF EACH IN COLUMN A.)

B. And what is the total number of part-time (less than 35 hours per week) employees of your (partnership/group) practice? Again, include persons regularly employed who are now on vacation, ill, etc. Do not include other physicians. RECORD ON BOTTOM LINE OF COLUMN B BELOW.

1. How many of these part-time employees are a . . . (READ CATEGORIES BELOW AS NECESSARY AND RECORD NUMBER OF EACH IN COLUMN B.)

Employees	A. Full-time (35 or more hours/week)		B. Part-time (Less than 35 hours/week)	
(1) Registered Nurse.....	_____	11-13/	_____	35-37/
(2) Licensed Practical Nurse....	_____	14-16/	_____	38-40/
(3) Nursing Aide.....	_____	17-19/	_____	41-43/
(4) Physician Assistant*.....	_____	20-22/	_____	44-46/
(5) Technician.....	_____	23-25/	_____	47-49/
(6) Secretary or Receptionist...	_____	26-28/	_____	50-52/
(7) Other (SPECIFY) _____	_____	29-31/	_____	53-55/
TOTAL:	[ ]	32-34/	TOTAL: [ ]	56-58/

*Physician Assistant must be a graduate of an accredited training program for Physician Assistants (Physician Extenders, Medex, etc.) or certified by the National Board of Medical Examiners through the Certification Exam for Assistant to the Primary Care Physician.

BEFORE YOU LEAVE, AGAIN STRESS THAT EACH AND EVERY AMBULATORY PATIENT SEEN BY THE DOCTOR OR HIS STAFF DURING THE 7-DAY PERIOD AT ALL IN-SCOPE OFFICE LOCATIONS (REPEAT THEM) IS TO BE INCLUDED IN THE SURVEY, THAT EACH PATIENT IS TO BE RECORDED ON THE LOG, AND ONLY THE APPROPRIATE NUMBER OF PATIENT RECORDS COMPLETED.

Thank you for your time, Dr. _____. If you have any (more) questions, please feel free to call me. My phone number is written in the folio. I'll call you on Monday morning of your survey week just to remind you.

12. TIME INTERVIEW ENDED ..... _____ AM / PM

13. DATE OF INTERVIEW ..... [ ][ ][ ][ ][ ][ ]
(Month) (Day) (Year)

COMMENTS:

INTERVIEWER NUMBER

[ ][ ][ ][ ][ ]

INTERVIEWER'S SIGNATURE

_____

FOR OFFICE USE ONLY:

No. of Patients Seen: [ ][ ][ ] 59-61/

Total Days in Practice during Week: [ ] 62/

No. of Patient Record Forms [ ][ ][ ] 63-65/

C 550428

Assurance of Confidentiality—All information which would permit identification of an individual, a practice, or an establishment will be held confidential, will be used only by persons engaged in and for the purposes of the survey and will not be disclosed or released to other persons or used for any other purpose

Department of Health and Human Services
Public Health Service
National Center for Health Statistics

C 550428

PATIENT LOG

As each patient arrives, record name and time of visit on the log below. For the patient entered on line #3, also complete the patient record to the right.

	PATIENT'S NAME	TIME OF VISIT
1		a.m. / p.m.
2		a.m. / p.m.
3		a m / p m

Record items 1-16 for this patient

CONTINUE LISTING PATIENTS ON NEXT PAGE

PATIENT RECORD
NATIONAL AMBULATORY MEDICAL CARE SURVEY

OMB No. 0937-0141
Expires 9/30/86
(PHS) 6105-C
456-232

1. DATE OF VISIT — Month / Day / Year

2. DATE OF BIRTH — Month / Day / Year

3. SEX
1 ☐ FEMALE
2 ☐ MALE

4. COLOR OR RACE
1 ☐ WHITE
2 ☐ BLACK
3 ☐ ASIAN/PACIFIC ISLANDER
4 ☐ AMERICAN INDIAN/ ALASKAN NATIVE

5. ETHNICITY
1 ☐ HISPANIC ORIGIN
2 ☐ NOT HISPANIC

6. EXPECTED SOURCE(S) OF PAYMENT *[Check all that apply]*
1 ☐ SELF-PAY
2 ☐ MEDICARE
3 ☐ MEDICAID
4 ☐ BLUE CROSS/ BLUE SHIELD
5 ☐ OTHER COMMERCIAL INSURANCE
6 ☐ HMO/PRE-PAID PLAN
7 ☐ NO CHARGE
8 ☐ OTHER *[Specify]*

7. WAS PATIENT REFERRED FOR THIS VISIT BY ANOTHER PHYSICIAN?
1 ☐ YES 2 ☐ NO

8. PATIENT'S COMPLAINT(S), SYMPTOM(S), OR OTHER REASON(S) FOR THIS VISIT *[In patient's own words]*
a MOST IMPORTANT
b OTHER

9. GLUCOSE TESTS THIS VISIT *[Check all ordered or provided]*
1 ☐ NONE
2 ☐ BLOOD
3 ☐ URINE
4 ☐ ORAL

10. OTHER DIAGNOSTIC SERVICES THIS VISIT *[Check all ordered or provided]*
1 ☐ NONE
2 ☐ BREAST EXAM
3 ☐ PELVIC EXAM
4 ☐ RECTAL EXAM
5 ☐ VISUAL ACUITY
6 ☐ URINALYSIS
7 ☐ HEMATOLOGY
8 ☐ BLOOD CHEMISTRY
9 ☐ PAP TEST
10 ☐ OTHER LAB TEST
11 ☐ BLOOD PRESSURE CHECK
12 ☐ EKG
13 ☐ CHEST X-RAY
14 ☐ OTHER RADIOLOGY
15 ☐ ULTRASOUND
16 ☐ OTHER SERVICE *[Specify]*

11. PHYSICIAN'S DIAGNOSES
a. PRINCIPAL DIAGNOSIS/PROBLEM ASSOCIATED WITH ITEM 8a.
b OTHER SIGNIFICANT CURRENT DIAGNOSES

12. HAVE YOU SEEN PATIENT BEFORE?
1 ☐ YES 2 ☐ NO
IF YES, FOR THE CONDITION IN ITEM 11a?
1 ☐ YES 2 ☐ NO

13. NON-MEDICATION THERAPY *[Check all services ordered or provided this visit]*
1 ☐ NONE
2 ☐ PHYSIOTHERAPY
3 ☐ AMBULATORY SURGERY
4 ☐ RADIATION THERAPY
5 ☐ PSYCHOTHERAPY
6 ☐ FAMILY PLANNING
7 ☐ DIET COUNSELING
8 ☐ OTHER COUNSELING
9 ☐ CORRECTIVE LENSES
10 ☐ OTHER *[Specify]* _____

14. MEDICATION THERAPY *[Record all new or continued medications ordered or provided at this visit. Use the same brand name or generic name entered on any Rx or office medical record.]*
IF NONE, CHECK HERE ☐

	a. NEW MEDICATION? YES	a. NEW MEDICATION? NO	b. FOR DX IN ITEM 11a? YES	b. FOR DX IN ITEM 11a? NO
1 _____	1 ☐	2 ☐	1 ☐	2 ☐
2 _____	1 ☐	2 ☐	1 ☐	2 ☐
3 _____	1 ☐	2 ☐	1 ☐	2 ☐
4 _____	1 ☐	2 ☐	1 ☐	2 ☐
5 _____	1 ☐	2 ☐	1 ☐	2 ☐

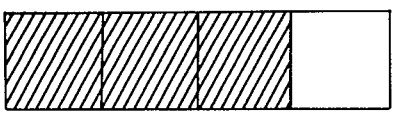
15. DISPOSITION THIS VISIT *[Check all that apply]*
1 ☐ NO FOLLOW-UP PLANNED
2 ☐ RETURN AT SPECIFIED TIME
3 ☐ RETURN IF NEEDED, PRN
4 ☐ TELEPHONE FOLLOW-UP PLANNED
5 ☐ REFERRED TO OTHER PHYSICIAN
6 ☐ RETURNED TO REFERRING PHYSICIAN
7 ☐ ADMIT TO HOSPITAL
8 ☐ OTHER *[Specify]* _____

16. DURATION OF THIS VISIT *[Time actually spent with physician]*
_____ Minutes

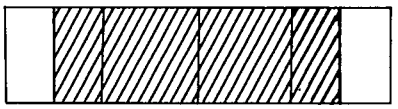
Appendix II
PSU selection probabilities for 1973 design

The methods used to assign zone selections to replicated subsamples (systematic selection of narrow zones in the case of SMSA's (Standard Metropolitan Statistical Areas), and random selection of narrow zones within broad zones for nonmetropolitan counties) are ad hoc procedures based on "best judgment" and designed to maximize the efficiency of sample estimation by ensuring adequate geographic and demographic spread for the PSU's (primary sampling units). For example, even after the preliminary stratification by region, population size, race, and urban status characteristics, there was apprehension about the chance that a simple random selection of four PSU's (one for each replicated subsample) within broad zones of 4 million would result in excessive geographic clustering; hence, the more systematic allocation scheme described in the section "Stratification and selection of PSU's, 1973 design," was chosen.

Because of the possibility that a large SMSA may fall across two zones or more, and not all of the zones will necessarily have been included in the finally selected subsample, the method of selection employed by the National Opinion Research Center of the University of Chicago (NORC) is not exactly equivalent to other methods of probability proportional to size sampling at the first stage. An alternative method of selecting 70 SMSA's, for example, would have been to break the SMSA frame into 70 zones of size 2 million each and, with a random start in the first zone, systematically select through the frame. In that method, any SMSA with a population greater than or equal to 2 million would be selected with certainty. On the other hand, if replication were desired, the zone width could have been made equal to 4 million and two random starts used to begin systematic sampling. To be selected with probability one in this case, an SMSA would have to be at least of size 4 million. Under the scheme of selection that was actually used, the probability of selection of an SMSA or county at the first stage depends on the placement of the geographic area in the frame with respect to the zone boundaries, as well as the size of the PSU. To illustrate, consider an SMSA of size 3 million placed in such a way that it completely covers three successive zones:



Because it is certain that two out of four zones will be selected for the main sample, the SMSA will be represented with probability one. On the other hand, consider the same SMSA placed as follows:



Because only one-half of zones 1 and 4 is covered, if these two zones are the final selections, it is possible that the SMSA will not be struck in either of the two. The probability of selection is as follows:

$$1-\frac{1}{6}\cdot\frac{1}{2}\cdot\frac{1}{2}=\frac{23}{24}.$$

Note that this example yields the minimum possible probability for an SMSA of 3 million. The factor of 1/6 is the probability that zones 1 and 4 are selected together, and the allocation of 500,000 to each zone maximizes the probability that the SMSA is missed in both zones.

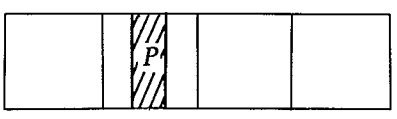
It is possible that an SMSA as large as 3,999,998 could be placed in such a way (999,999 in each of zones 1 and 4) that it might not be selected (the same reasoning applies here as in the preceding example). The selection probability is very close to one, but only if the SMSA is 4 million or greater will the selection probability be independent of zone placement.

A peculiarity of this selection method is that the probability of choosing a geographic area at the first stage is proportional to its size only in certain frame placement situations.

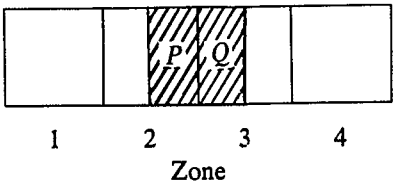
First-stage selection probabilities were calculated for each of the SMSA's and counties outside SMSA's represented in the main NORC national sample. The method for computing the probabilities is described in the following paragraphs.

There are only four types of configuration of area placement on the zone grid that will result in a first-stage selection probability less than one. They are as follows:

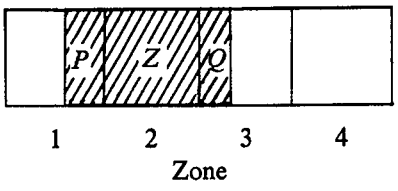
- Case a:



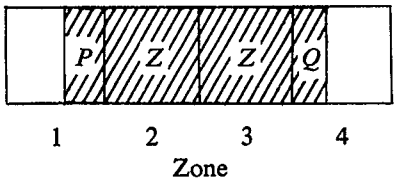
- Case b:



- Case c:



- Case d:



Let Z equal the width of a zone. In case a the first-stage area is of size $P < Z$. Recall that for the main NORC sample, two zones were randomly selected out of every four. Hence, the selection probability for the area in case a is $P/2Z$.

In case b, a first-stage area of size $P + Q < 2Z$ is placed on the grid in a way that causes it to straddle a zone boundary, with areas of size P and Q, respectively, on each side of the boundary. In the choice of two zones out of four, the selection probability for any pair of zones is 1/6. The probability that the geographic area shown in case b will be selected at the first stage is

$$\frac{1}{3}\left(\frac{P}{Z}+\frac{Q}{Z}\right)+\frac{1}{6}\left[\frac{P}{Z}\left(1-\frac{Q}{Z}\right)+\frac{Q}{Z}\left(1-\frac{P}{Z}\right)+\frac{PQ}{Z^2}\right]$$

$$=\frac{P+Q}{2Z}-\frac{PQ}{6Z^2}.$$

In case c, the geographic area covers one zone completely and parts of two end zones; hence, the total size is $P + Q + Z$ with P and Q both less than Z. Considering all of the possible pairings of zones in the selection of two out of four, one obtains the following probability:

$$\frac{1}{6}\left(1+\frac{P}{Z}+\frac{Q}{Z}\right)+\frac{1}{6}\left[\frac{P}{Z}\left(1-\frac{Q}{Z}\right)+\frac{Q}{Z}\left(1-\frac{P}{Z}\right)+\frac{PQ}{Z^2}\right]$$

$$+\frac{1}{6}\left(1-\frac{P}{Z}+\frac{P}{Z}\right)+\frac{1}{6}\left(1-\frac{Q}{Z}+\frac{Q}{Z}\right)$$

$$=\frac{1}{2}+\frac{P+Q}{3Z}-\frac{PQ}{6Z^2}.$$

Newark is an example of an SMSA of the form of case c, with $P = 319{,}476$ and $Q = 537{,}080$. The first-stage selection probability is 0.756921.

Case d covers the situation already illustrated earlier in this appendix. The general probability formula is

$$1-\frac{1}{6}\left(1-\frac{P}{Z}\right)\left(1-\frac{Q}{Z}\right).$$

Appendix III
Controlled selection of PSU's for 1985 design, SRC-NORC 1980 master sample

Controlled selection (Goodman and Kish, 1960) is a probability sampling technique that incorporates distributional controls on sample elements beyond those explicitly built into the stratification of the sample design. Controlled selection was used extensively in selecting the primary sampling units (PSU's) for the 1980 national sample designed jointly by the Survey Research Center of the University of Michigan and the National Opinion Research Center of the University of Chicago. The exact nature of the sample controls varies depending on the census region, Standard Metropolitan Statistical Area (SMSA) status (SMSA or not), and the sample strata being considered.

In general a two-step procedure was used: A first controlled selection step to establish the sample's distribution to SMSA's and areas outside of SMSA's of census divisions or similar geographic subdivisions of census regions; and a second step, constrained by the outcome of the first, that determined more specific characteristics of the set of PSU's from which each stratum's selection was to be drawn. Table I provides the exact specifications for each controlled selection used in the primary stage of the 1980 national sample.

To describe the several steps in the selection procedure, an example drawn from the West region sample is useful. Once the sample design strata are formed, the controlled selection problem is defined. In the case of the West region, further controls on the geographic distribution of the sample were desired. A number of other control variables might have been used, but the small number of strata involved limits the effectiveness of a more complex set of control specifications. The geographic controls of interest in the West were: (1) A proportionate distribution of the sample to the Mountain and Pacific census divisions; and (2) within divisions, a further control on the sample's proportionate distribution to States and selected State groups.

For the West's Mountain and Pacific census divisions, SMSA status, and sample design strata, table II gives the expected PSU sample sizes under the probability proportional to size sampling design. (Measures of size are in proportion to occupied housing units.) Note that the sample size expectations for the stratum margin are exactly 1.0 (or 2.0 in the special case of a double stratum). This is a constraint imposed by the sample design. In contrast to the strata marginals, the division and SMSA status marginals may have nonintegral values for sample size expectations. Looking at the marginal cell representing the sample size expectation for SMSA's of the Mountain division, table II shows a value of 2.76. Under controlled selection, there is a 0.76 probability that three Mountain division SMSA's will be selected. The probability of two selections is $1.0 - 0.76 = 0.24$.

Moving to the interior cells of the sample expectation matrix, it is clear that the sample outcome for Mountain division SMSA's is only in question in stratum 55 ($p = 0.61$ of a Mountain division selection) and stratum 59 ($p = 0.15$ of a selection). Note that SMSA strata 52, 53 and 54, 56 and 57, and 58 contain no Mountain division PSU's. Consequently, the controlled selection outcome for these strata is fixed. The selections must come from the Pacific division. Likewise, strata 60 and 61 selections can only be Mountain division SMSA's.

Among the strata outside of SMSA's, the allocation of the sample PSU to either the Mountain or Pacific division is predetermined except for the case of stratum 83; the selected PSU in stratum 82 is in the Pacific division and that from stratum 84 is in the Mountain division by default. Given the probability (sample expectation) matrix shown in table II, the actual solution to the controlled selection program was obtained using a computer program that incorporates the controlled selection algorithm.

The output of the controlled selection program is a series of sample patterns. Each sample pattern is a matrix that defines one particular sample allocation scheme that conforms, under controlled rounding, to the sample expectations calculated for the cells and margins of the controlled selection input matrix. Individually, patterns in the series are formed with varying probabilities, and each is assigned a weight that indicates the likelihood (on a relative basis) of its being included in the solution set. Collectively, when weighted by the respective probabilities, the set of patterns will exactly satisfy the sampling probabilities of the controlled selection matrix. For subsequent sampling steps, a single pattern is selected from the set with probability proportionate to the assigned pattern weight.

Again using the West region as an example, table III presents the pattern that was chosen for the 1980 national sample. By reviewing the outcome for the strata on which the controlled selection could operate, it may be seen that strata 55 and 83 selections were to be Mountain division PSU's. At the same time, the stratum 59 sample PSU was designated to come from the Pacific States. The selected pattern merely specified the census division from which each stratum's sample PSU will be drawn. Another controlled selection step occurs before the sample PSU's are finally identified.

NOTE: A list of references follows the text.

Table I. Schema for the controlled selection of 1980 SRC-NORC (Survey Research Center-National Opinion Research Center) national sample primary sampling units

	Selection step 1		*Selection step 2*[1]		
Region	*Control variable*	*Category*	*Separately for—*	*Control variable*[2]	*Category*[3]
Northeast	Census division	Mid-Atlantic; New England	Mid-Atlantic division	State	NJ; NY; PA
	Degree urban	SMSA; outside SMSA's		Sample replicate	Basic; reserve
	Strata	16 strata	New England division	State group	NH, VT, ME; CT, RI; MA
				Sample replicate	Basic; reserve
South—"Deep South"	Geographic location	Atlantic; Gulf	SMSA's—Atlantic	State	GA; SC
	Degree urban	SMSA; outside SMSA's	SMSA's—Gulf	State	AL; LA; MS
	Strata	8 strata	All outside SMSA's	State	AL; GA; LA; MS; SC
				Largest city size	0–2,500; 2,501–10,000; 10,001–25,000; greater than 25,000
				Percent black	Less than 30 percent; 30–49 percent; 50 percent or greater
South—balance of States	Geographic location	AR, TX, OK; KY, TN, WV; DE, MD, VA, NC; FL	SMSA's—AR, TX, OK	State	AR; TX; OK
				Sample replicate	Basic; reserve
	Degree urban	SMSA; outside SMSA's	SMSA's—KY, TN, WV	State	KY; TN; WV
	Strata	23 strata		Sample replicate	Basic; reserve
			SMSA's—DE, MD, VA, NC	State	DE; MD; VA; NC
				Sample replicate	Basic; reserve
			Non-SMSA's—AR, TX, OK	State	AR; TX; OK
				Largest city size	0–2,500; 2,501–10,000; 10,001–25,000; greater than 25,000
			Non-SMSA's—KY, TN, WV	State	KY; TN; WV
				Largest city size	0–2,500; 2,501–10,000; 10,001–25,000; greater than 25,000
			Non-SMSA's—DE, MD, VA, NC	State	DE; MD; VA; NC
				Largest city size	0–2,500; 2,501–10,000; 10,001–25,000; greater than 25,000
North Central	Geographic location	East; West	East	State	OH; IN; IL; MI
	Degree urban	SMSA; outside SMSA's		Sample replicate	Basic; reserve
	Strata	22 strata	West	State	WI; MN; IA; MO; KS; NE; ND; SD
				Sample replicate	Basic; reserve
West	Census division	Mountain; Pacific	Mountain division	State group	NM, AZ, NV, CO; UT, ID, MT, WY
	Degree urban	SMSA; outside SMSA's		Sample replicate	Basic; reserve
	Strata	15 strata	Pacific division	State	CA; OR; WA
				Sample replicate	Basic; reserve

[1]Selection step 2 is not applicable for FL.
[2]Control to strata sample size allocation continues in force.
[3]Postal Service abbreviations are used for States. City sizes are given in numbers of persons.

NOTE: SMSA is Standard Metropolitan Statistical Area.

Table II. Sample size expectations for first step in controlled selection of 1980 primary sampling units in West census region, Mountain division, and Pacific division

Stratum	West region	Mountain division	Pacific division
All strata	13.00	4.45	8.56
NSR, SMSA			
All strata	10.00	2.76	7.24
52	1.00	0.00	1.00
53 and 54[1]	2.00	0.00	2.00
55	1.00	0.61	0.39
56 and 57[1]	2.00	0.00	2.00
58	1.00	0.00	1.00
59	1.00	0.15	0.85
60	1.00	1.00	0.00
61	1.00	1.00	0.00
NSR, not SMSA			
All strata	3.00	1.68	1.32
82	1.00	0.00	1.00
83	1.00	0.68	0.32
84	1.00	1.00	0.00

[1]Double stratum: 2 selections are made without replacement.

NOTES: NSR is non-self-representing; SMSA is Standard Metropolitan Statistical Area.

Table III. Sampled pattern from 1st step controlled selection of primary sampling units (PSU's) in West census region, Mountain division, and Pacific division: Step 2 selections

Stratum	West region	Mountain division	Pacific division
	Number of PSU's		
All strata	13	5	8
NSR, SMSA			
All strata	10	3	7
52	1	-	1
53 and 54[1]	2	-	2
55	1	1	-
56 and 57[1]	2	-	2
58	1	-	1
59	1	-	1
60	1	1	-
61	1	1	-
NSR, not SMSA			
All strata	3	2	1
82	1	-	1
83	1	1	-
84	1	1	-

[1]Double stratum: 2 selections are made without replacement.

NOTES: NSR is non-self-representing; SMSA is Standard Metropolitan Statistical Area.

Having exercised a control on the distribution of PSU's to census divisions, additional constraints were imposed in a second controlled selection step. The right-hand columns of table I provide the details of the second controlled selection for each of the major census regions. In the West, the principal variable in the second controlled selection was geographic location by State (within the Pacific division) or State group (within the Mountain division). The control operated across all strata, ensuring that each of the designated States or State groups receives (in expectation) a proportionate representation in the 1980 national sample.

Except in the South region, allocation to sample replicate—basic or reserve—was also incorporated in the step 2 controlled selection. The example used in the preceding discussion contains a simplification in that it deals only with the selection of the basic sample replicate. The actual primary-stage selection was more complex. Instead of selecting the basic and reserve samples independently, the two replicates were chosen simultaneously. Through the first controlled selection step, the problem was treated as one in which two selections were to be drawn from each stratum, and only in the second controlled selection step were the two sample replicates actually distinguished. As a result, a sample design that combines the two replicates will benefit from the same geographic and other controls that are present in the individual basic and reserve half-sample replicates.

* U.S. GOVERNMENT PRINTING OFFICE 1988; 201-855/80008

Vital and Health Statistics series descriptions

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

301–436–8500

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

OFFICIAL BUSINESS
PENALTY FOR PRIVATE USE, $300

THIRD CLASS MAIL
BULK RATE
POSTAGE & FEES PAID
PHS/NCHS
PERMIT No. G–281

DHHS Publication No. (PHS) 88–1382, Series 2, No. 108