

Sample Design of the 1996
Medical Expenditure Panel Survey
Nursing Home Component

Methodology

Report 4



U.S. Department of Health and Human Services
Public Health Service
Agency for Health Care Policy and Research

Health Care Information and Electronic Ordering Through the AHCPR Web Site

Put Prevention Into Practice

[Text Version]

Welcome!

AHCPR
Agency for Health Care Policy and Research

What's New | How To Use This Web Site | Frequently Asked Questions | Información en español

About AHCPR	Agency mission, organization, key contacts, National Advisory Council	Search
News & Announcements	AHCPR profile, press releases, speeches, job vacancies, general announcements	Browse
Funding Opportunities	Research agenda, grant announcements, notices, contracts, research training assistance	
Research Findings	Health services research, primary care, women's & children's health, Research Activities online, User Liaison Program	
Quality Assessment	Quality measurement, CONQUEST, Consumer Assessment of Health Plans (CAHPS)	
Clinical Information	Evidence-based practice, clinical practice guidelines, preventive services, technology assessments, outcomes research	
Consumer Health	Healthfinder, specific health conditions, surgery, prescriptions, health plans, other health care issues	info@ahcpr.gov
Data & Surveys	Medical Expenditure Panel Survey (MEPS), Healthcare Cost & Utilization Project, HIV costs & use, informatics standards	
Information Products	Availability & ordering information, Government Information Locator Service (GILS), Web site links	

Share Your Life. National Initiative on Organ & Tissue Donation

AHCPR: Improving Health Care Quality Through Research & Education

The Agency for Health Care Policy and Research's Web site—<http://www.ahcpr.gov/>—makes practical, science-based health care information available in one convenient place.

Buttons correspond to major categories of Web site information, including funding opportunities, research findings, quality assessments, clinical information, consumer health, and data and surveys.

The Web site features an Electronic Catalog to the more than 450 information products generated by AHCPR, with information on how to obtain these resources. Many information products have an electronic ordering form and are mailed free of charge from the AHCPR Clearinghouse within 5 working days.

<http://www.ahcpr.gov/>

Abstract

The Medical Expenditure Panel Survey (MEPS) is the third in a series of nationally representative surveys of medical care use and expenditures sponsored by the Agency for Health Care Policy and Research (AHCPR). MEPS comprises four component surveys. The Nursing Home Component produces national estimates of insurance coverage and the use of services, expenditures, and sources of payment for persons residing in or admitted to nursing homes. The NHC also gathers information on nursing home characteristics—such as facility type, ownership, chain affiliation, certification, facility size, and location—for a nationally representative

sample of nursing homes. This report documents the implementation of the sample design for the MEPS Nursing Home Component, including the sampling frame, facility selection, and within-facility sample selection through Round 1 of data collection.

Suggested citation

Bethel J, Broene P, Sommers JP. Sample design of the 1996 Medical Expenditure Panel Survey Nursing Home Component. Rockville (MD): Agency for Health Care Policy and Research; 1998. *MEPS Methodology Report No. 4*. AHCPR Pub. No. 98-0042.

Sample Design of the 1996
Medical Expenditure Panel Survey
Nursing Home Component

Methodology

Report 4



U.S. Department of Health and Human Services
Public Health Service
Agency for Health Care Policy and Research
AHCPR Pub. No. 98-0042
September 1998

The Medical Expenditure Panel Survey (MEPS)

Background

The Medical Expenditure Panel Survey (MEPS) is conducted to provide nationally representative estimates of health care use, expenditures, sources of payment, and insurance coverage for the U.S. civilian noninstitutionalized population. MEPS also includes a nationally representative survey of nursing homes and their residents. MEPS is cosponsored by the Agency for Health Care Policy and Research (AHCPR) and the National Center for Health Statistics (NCHS).

MEPS comprises four component surveys: the Household Component (HC), the Medical Provider Component (MPC), the Insurance Component (IC), and the Nursing Home Component (NHC). The HC is the core survey, and it forms the basis for the MPC sample and part of the IC sample. The separate NHC sample supplements the other MEPS components. Together these surveys yield comprehensive data that provide national estimates of the level and distribution of health care use and expenditures, support health services research, and can be used to assess health care policy implications.

MEPS is the third in a series of national probability surveys conducted by AHCPR on the financing and use of medical care in the United States. The National Medical Care Expenditure Survey (NMCES) was conducted in 1977, the National Medical Expenditure Survey (NMES) in 1987. Beginning in 1996, MEPS continues this series with design enhancements and efficiencies that provide a more current data resource to capture the changing dynamics of the health care delivery and insurance system.

The design efficiencies incorporated into MEPS are in accordance with the Department of Health and Human Services (DHHS) Survey Integration Plan of June 1995, which focused on consolidating DHHS surveys, achieving cost efficiencies, reducing respondent burden, and enhancing analytical capacities. To accommodate these goals, new MEPS design features

include linkage with the National Health Interview Survey (NHIS), from which the sample for the MEPS HC is drawn, and enhanced longitudinal data collection for core survey components. The MEPS HC augments NHIS by selecting a sample of NHIS respondents, collecting additional data on their health care expenditures, and linking these data with additional information collected from the respondents' medical providers, employers, and insurance providers.

Household Component

The MEPS HC, a nationally representative survey of the U.S. civilian noninstitutionalized population, collects medical expenditure data at both the person and household levels. The HC collects detailed data on demographic characteristics, health conditions, health status, use of medical care services, charges and payments, access to care, satisfaction with care, health insurance coverage, income, and employment.

The HC uses an overlapping panel design in which data are collected through a preliminary contact followed by a series of five rounds of interviews over a 2½-year period. Using computer-assisted personal interviewing (CAPI) technology, data on medical expenditures and use for 2 calendar years are collected from each household. This series of data collection rounds is launched each subsequent year on a new sample of households to provide overlapping panels of survey data and, when combined with other ongoing panels, will provide continuous and current estimates of health care expenditures.

The sampling frame for the MEPS HC is drawn from respondents to NHIS, conducted by NCHS. NHIS provides a nationally representative sample of the U.S. civilian noninstitutionalized population, with oversampling of Hispanics and blacks.

Medical Provider Component

The MEPS MPC supplements and validates information on medical care events reported in the MEPS HC by contacting medical providers and pharmacies identified by household respondents. The MPC sample includes all hospitals, hospital physicians, home health agencies, and pharmacies reported in the

HC. Also included in the MPC are all office-based physicians:

- Providing care for HC respondents receiving Medicaid.
- Associated with a 75-percent sample of households receiving care through an HMO (health maintenance organization) or managed care plan.
- Associated with a 25-percent sample of the remaining households.

Data are collected on medical and financial characteristics of medical and pharmacy events reported by HC respondents, including:

- Diagnoses coded according to ICD-9 (9th Revision, International Classification of Diseases) and DSM-IV (Fourth Edition, *Diagnostic and Statistical Manual of Mental Disorders*).
- Physician procedure codes classified by CPT-4 (Current Procedural Terminology, Version 4).
- Inpatient stay codes classified by DRG (diagnosis-related group).
- Prescriptions coded by national drug code (NDC), medication names, strength, and quantity dispensed.
- Charges, payments, and the reasons for any difference between charges and payments.

The MPC is conducted through telephone interviews and mailed survey materials.

Insurance Component

The MEPS IC collects data on health insurance plans obtained through employers, unions, and other sources of private health insurance. Data obtained in the IC include the number and types of private insurance plans offered, benefits associated with these plans, premiums, contributions by employers and employees, and employer characteristics.

Establishments participating in the MEPS IC are selected through four sampling frames:

- A list of employers or other insurance providers identified by MEPS HC respondents who report having private health insurance at the Round 1 interview.
- A Bureau of the Census list frame of private-sector business establishments.
- The Census of Governments from the Bureau of the Census.

- An Internal Revenue Service list of the self-employed.

To provide an integrated picture of health insurance, data collected from the first sampling frame (employers and other insurance providers) are linked back to data provided by the MEPS HC respondents. Data from the other three sampling frames are collected to provide annual national and State estimates of the supply of private health insurance available to American workers and to evaluate policy issues pertaining to health insurance.

The MEPS IC is an annual panel survey. Data are collected from the selected organizations through a prescreening telephone interview, a mailed questionnaire, and a telephone followup for nonrespondents.

Nursing Home Component

The 1996 MEPS NHC was a survey of nursing homes and persons residing in or admitted to nursing homes at any time during calendar year 1996. The NHC gathered information on the demographic characteristics, residence history, health and functional status, use of services, use of prescription medications, and health care expenditures of nursing home residents. Nursing home administrators and designated staff also provided information on facility size, ownership, certification status, services provided, revenues and expenses, and other facility characteristics. Data on the income, assets, family relationships, and caregiving services for sampled nursing home residents were obtained from next-of-kin or other knowledgeable persons in the community.

The 1996 MEPS NHC sample was selected using a two-stage stratified probability design. In the first stage, facilities were selected; in the second stage, facility residents were sampled, selecting both persons in residence on January 1, 1996, and those admitted during the period January 1 through December 31.

The sampling frame for facilities was derived from the National Health Provider Inventory, which is updated periodically by NCHS. The MEPS NHC data were collected in person in three rounds of data collection over a 1½-year period using the CAPI system. Community data were collected by telephone using computer-assisted telephone interviewing (CATI) technology. At the end of three rounds of data collection,

the sample consisted of 815 responding facilities, 3,209 residents in the facility on January 1, and 2,690 eligible residents admitted during 1996.

Survey Management

MEPS data are collected under the authority of the Public Health Service Act. They are edited and published in accordance with the confidentiality provisions of this act and the Privacy Act. NCHS provides consultation and technical assistance.

As soon as data collection and editing are completed, the MEPS survey data are released to the public in staged releases of summary reports and microdata files. Summary reports are released as printed documents and electronic files. Microdata files are released on CD-ROM and/or as electronic files.

Printed documents and CD-ROMs are available through the AHCPR Publications Clearinghouse. Write or call:

AHCPR Publications Clearinghouse
Attn: (publication number)
P.O. Box 8547
Silver Spring, MD 20907
800/358-9295
410/381-3150 (callers outside the United States only)
888/586-6340 (toll-free TDD service; hearing impaired only)

Be sure to specify the AHCPR number of the document or CD-ROM you are requesting. Selected electronic files are available through the Internet on the AHCPR Web site:

<http://www.ahcpr.gov/>

On the AHCPR Web site, under Data and Surveys, click the MEPS icon.

Additional information on MEPS is available from the MEPS project manager or the MEPS public use data manager at the Center for Cost and Financing Studies, Agency for Health Care Policy and Research, 2101 East Jefferson Street, Suite 500, Rockville, MD 20852 (301/594-1406).

Table of Contents

Overview of Sample Design.	1
Sampling Frame.	4
Facility Selection	5
Sampling of Persons Within Facilities.	14
References	21

Tables

1. Person-level population estimates and relative standard errors for nursing and personal care homes: 1987 National Medical Expenditure Survey Institutional Population Component.	2
2. Nursing and personal care facilities—number sampled and expected number responding by round: 1996 MEPS Nursing Home Component.	3
3. Optimum allocation to cost strata based on the 1996 MEPS Nursing Home Component sample	8
4. Minimum acceptable response rates for the 1996 MEPS Nursing Home Component	12
5. Rules for facility sampling: 1996 MEPS Nursing Home Component	14
6. Response rates at end of Round 1: 1996 MEPS Nursing Home Component	15

Figures

1. First admissions eligibility determination using a single respondent: 1996 MEPS Nursing Home Component.	18
2. Final determination of eligibility for sampled nursing home residents: 1996 MEPS Nursing Home Component.	20

Sample Design of the 1996 Medical Expenditure Panel Survey Nursing Home Component

by James Bethel, Ph.D., and Pamela Broene, M.S., Westat, Inc., and John Paul Sommers, Ph.D.,
Agency for Health Care Policy and Research

Overview of Sample Design

The goal of the Medical Expenditure Panel Survey Nursing Home Component (MEPS NHC) is to produce national estimates for persons residing in nursing homes during 1996. Information was gathered on nursing home characteristics for a nationally representative sample of nursing homes and on the demographic characteristics, residence history, health status, and long-term care expenditures for a sample of residents in these nursing homes. This report documents the implementation of the sample design, including the sampling frame, facility selection, and within-facility sample selection through Round 1 of data collection.

The target population consists of freestanding nursing homes with at least three beds staffed and set up for nursing care, as well as nursing care units consisting of a distinguishable group of three or more nursing home beds within a larger facility. Either type of facility must be:

- Medicare certified as a skilled nursing facility and/or Medicaid certified as a nursing facility, or
- Licensed as a nursing home with a registered nurse or licensed practical nurse onsite 24 hours a day, 7 days a week.

The sample of nursing home residents was stratified by whether residents resided at the nursing home at the beginning of 1996 (current-residents sample) or were admitted during the calendar year (first-admissions sample). The target population of the current-residents sample consisted of persons who resided in nursing homes as of January 1, 1996. The target population of the first-admissions sample consisted of persons who resided in a nursing home during 1996 but were not current residents as defined above.

The sample was designed with the goal of estimating a population proportion of 0.20 with a coefficient of variation (CV) of 9.8 percent or less for facilities, 5.5 percent or less for current residents, and 6.5 percent or less for first admissions. Table 1 shows the relative standard errors for selected characteristics of current residents and

first admissions obtained in the 1987 National Medical Expenditure Survey Institutional Population Component, a similar previous survey conducted by the Agency for Health Care Policy and Research (AHCPR). These relative standard errors, or CVs, are based on a sample size of approximately 800 responding nursing homes.

The sampling frame for the selection of facilities in the MEPS NHC was an updated version of the 1991 National Health Provider Inventory (NHPI). The 1991 NHPI is a census of approximately 16,000 nursing homes in the United States, collected by the Bureau of the Census for the National Center for Health Statistics and AHCPR. The 1991 NHPI served as the base, and approximately 2,000 new facilities and 275 hospital-based facilities were added to this original list to create the sampling frame.

Facilities were selected as a double, or two-phase, sample. For the first phase, 1,651 facilities were sampled within strata, with probabilities proportional to size. The measure of size was the number of beds in the facility reserved for nursing home use. The first-phase sample was assigned to four strata based on expected survey travel costs, and a second-phase subsample of 1,430 facilities was selected with equal probabilities within these four cost strata.

The second-phase sample was divided into a main sample of 1,150 facilities and a reserve sample of 280 facilities, the latter being divided into four "release groups" of 70 facilities each. The release groups were intended to be sent to the field to supplement the main sample if response and eligibility rates were lower than expected. On the other hand, the main sample was randomly split into 18 recall groups of approximately 64 facilities each. If response and eligibility rates were higher than expected, sampled facilities could be randomly withdrawn from the field by canceling data collection in selected recall groups. In fact, the MEPS NHC Round 1 response and eligibility rates were higher than anticipated. Therefore, at the conclusion of Round 1, the facilities in two randomly selected recall groups were withdrawn from Rounds 2 and 3 of data collection.

Table I. Person-level population estimates and relative standard errors for nursing and personal care homes: 1987 National Medical Expenditure Survey Institutional Population Component

Sample and measure	Estimate	Relative standard error
Current residents		
Average annual expenses	\$16,432	0.018
Per diem expenses	\$52.10	0.016
Average annual Medicaid expenses	\$8,236	0.040
Total days in facility	317.0	0.006
Proportion with dementia	0.133	0.024
Proportion needing help walking	0.700	0.014
Proportion white	0.910	0.008
First admissions		
Average annual expenses	\$6,884	0.029
Per diem expenses	\$68.00	0.027
Average annual Medicaid expenses	\$2,016	0.077
Total days in facility	114.80	0.023
Proportion with dementia	0.063	0.040
Proportion needing help walking	0.754	0.017
Proportion white	0.906	0.011

Source: Agency for Health Care Policy and Research: 1987 National Medical Expenditure Survey.

In most facilities, a fixed sample of four current residents and four new admissions was selected using simple random sampling within the facility. In facilities with measures of size that were poorly correlated with the number of admissions, the first admissions sample size could be increased from two to three per round (to a total

of six). The within-facility sample sizes are intended to yield approximately 3,043 eligible current-resident respondents and 2,218 eligible first-admission respondents, all with complete use and expenditure data. The target sample sizes are summarized in Table 2.

Table 2. Nursing and personal care facilities—number sampled and expected number responding by round: 1996 MEPS Nursing Home Component

Units	Round 1	Round 2	Round 3	Total
Facilities				
Selected at first phase	1,651	—	—	—
Selected at second phase	1,150	—	—	—
Eligible	1,127	—	—	—
Completed FQ	862	—	—	—
Cooperated with sampling	836	811	787	—
Current residents				
Selected and eligible	3,344	—	—	—
Unit response for IUED and RH data ¹	—	3,243	3,144	—
Completed IUED and RH data ²	—	3,177	3,043	—
Completed background data	2,842	—	—	—
Completed baseline health status data	3,210	—	—	—
Completed end-of-year health status data	—	—	2,020	—
First admissions				
Selected	—	1,622	1,574	3,196
Not first admissions	—	357	346	7 03
Eligible first admissions	—	1,265	1,228	2,493
Unit response for IUED and RH data ¹	—	1,189	1,155	2,344
Completed IUED and RH data ²	—	1,125	1,093	2,218
Completed background data	—	1,037	1,007	2,044
Completed baseline health status data	—	1,163	1,130	2,293
Completed end-of-year health status data	—	397	673	1,070

¹At least 1/3 of data completed.

²All data provided.

Note: FQ =Facility Questionnaire. IUED = institutional use and expenditure data. RH = residence history.

Source: Agency for Health Care Policy and Research: 1996 Medical Expenditure Panel Survey Nursing Home Component.

Sampling Frame

Description of the National Health Provider Inventory

The MEPS NHC sampling frame was based on the 1991 NHPI. The U.S. Bureau of the Census collects the NHPI for the National Center for Health Statistics (NCHS) and AHCPR. In 1991, it contained approximately 16,000 nursing homes and 31,000 board-and-care homes. The MEPS NHC sampling frame was created by updating a subset of the 1991 NHPI provided by NCHS to AHCPR. This subset contained 15,811 facilities on the 1991 NHPI that NCHS defined as nursing homes, as well as 1,691 potential new nursing homes that were identified through State lists and directories of nursing homes (National Center for Health Statistics, 1995).

A nursing home, according to the NCHS definition, is a facility having at least three beds and identifying itself on the NHPI questionnaire as one of the following:

- A licensed nursing home;
- A skilled nursing long-term care unit of a hospital;
- A nursing care unit of a retirement center;
- A nursing facility certified under Medicare or Medicaid; or
- Some other type of nursing home.

Among the 17,502 facilities on the basic frame, 205 appeared to be board-and-care homes, which do not meet this definition and were excluded. Another 275 facilities that had been excluded by NCHS—mostly Department of Veterans' Affairs (VA) facilities—were added to the frame. The final updated NHPI contained 17,572 (17,502 – 205 + 275) facilities (Pancholi, 1995).

To be eligible for the MEPS NHC, facilities must have at least three beds and be either Medicare- and/or Medicaid-certified or licensed as nursing homes. Final eligibility for the MEPS NHC was determined during Round 1 of facility data collection; however, the initial sampling frame included all facilities on the updated NHPI that were likely to meet these criteria.

Editing the MEPS NHC Frame

The number of beds reported by the facility on the NHPI questionnaire was edited for hospital-based facilities using information in the AHA Guide (American Hospital Association, 1993). As part of the editing, the number of beds was compared with the number of residents. A large ratio of residents to beds could indicate an inconsistency in reporting, unless the questionnaire shows the presence of a long-term care unit within a larger facility for the elderly. As a result, the number of beds was edited for 209 hospital-based facilities.

Missing values for variables on the NHPI that were needed for sampling were obtained from external sources when possible. Information on license/certification status, type of ownership, and the number of beds was obtained for all but a small percentage of the new facilities. Certification status was not available for the 275 VA facilities and was imputed for an additional 69 facilities. The facility ownership type (profit, nonprofit, government) was unknown for 216 facilities but was not imputed. Missing telephone numbers were supplied for over 1,700 facilities, but 132 still remained missing on the frame at the time of sampling. A Beale code (also known as the Human Resource Profile Code) was placed on the file to indicate the metropolitan statistical area (MSA) status of each facility. These codes were collapsed for use in sampling. The collapsed values were 0 = large metropolitan core, 1 = large metropolitan fringe, 2 = medium metropolitan area, 3 = lesser metropolitan area, 4 = adjacent to an MSA, and 5 = not adjacent to an MSA.

Measure of Size

In the initial survey planning, the number of residents was proposed as the most appropriate measure of size. A careful review of the data fields on the NHPI, however, indicated that the number of “eligible beds” would be preferable. The question on the NHPI concerning the number of residents is somewhat general (“How many residents stayed in this home last night?”). In contrast, the question concerning the number of beds defines certain types of beds that should be excluded

(e.g., beds for day care only and hospital or retirement center beds not associated with the nursing home). Also, the number of residents might be construed to include persons in a board-and-care wing. Therefore, the number of beds as reported on the NHPI questionnaire was used directly as the measure of size except for the 209 cases where the number of beds was edited using the 1993 American Hospital Association Guide (American Hospital Association, 1993).

Facility Selection

Summary

Facilities were selected in two phases. In the first phase, a stratified sample of 1,651 facilities was selected with probability proportional to size. Six of the seven strata were created by crossing three types of Medicaid/Medicare reimbursement with an indicator of whether the facility was hospital based or not. The seventh stratum contained the 20 largest facilities, of which 11 had been chosen by NCHS for inclusion in the NCHS National Nursing Home Survey (NNHS) and the remaining 9 were designated for the MEPS NHC. These nine facilities were then drawn with certainty in the first phase of the MEPS NHC. The stratum sample sizes in the remaining six strata were determined using proportional allocation. The original measure of size was the number of beds, but to minimize overlap with the NCHS NNHS, a Keyfitz procedure was employed to compute new conditional probabilities of selection.

Cost stratification was then performed on the 1,651 facilities in the first-phase sample, with the actual strata being defined in terms of distance from the nearest of the 50 largest U.S. cities and the expected effect on survey travel costs. Next, the optimal sampling rates were determined for these four cost strata. Using the sampling rates, a cost-stratified subsample of 1,430 facilities was selected from the 1,651 facilities in the main sample. Within each cost stratum, the second-phase sample of facilities not drawn with certainty (noncertainty facilities) was subsampled at a rate of .803, yielding a reserve sample of 280 facilities and a remaining main sample of 1,150 facilities. The four release groups were assigned by sorting the reserve sample by order of selection and consecutively

numbering from 1 to 4, repeating until all 280 facilities were assigned. This resulted in four release groups of 70 facilities each. The main sample was randomly divided into 18 subsamples of approximately 64 facilities each by sorting the noncertainty sample facilities in the order of selection and consecutively numbering from 1 to 18, repeating until all 1,139 noncertainty facilities were assigned.

Initial Stratification

The facility sample was a two-phase stratified sample. In the first phase, the 17,572 facilities in the frame were stratified into seven strata. Facilities were selected with probabilities proportional to size (i.e., the number of eligible beds) in each stratum in the first phase. This initial sample was grouped into the four cost strata described above. In the second phase, noncertainty facilities were subsampled with equal probabilities within each cost stratum.

The first-phase strata were formed by grouping facilities according to three types of Medicaid/Medicare reimbursement and whether the facility was hospital based or not. The 20 largest facilities were placed in a separate stratum. Eleven of these had been selected previously for the NNHS, conducted by the NCHS. The remaining nine were designated for the MEPS NHC.

Determining Selection Possibilities

An initial sample of 1,651 facilities was selected using probability proportionate to size (PPS) sampling, with the number of nursing beds as the measure of size. Using this technique, the number of facilities allocated to each of the original strata was proportional to that stratum's number of the total number of nursing beds. For the i th facility in the h th stratum, the initial selection probability was computed as:

$$\begin{aligned}
 P_{hi} &= n_h \frac{\text{Facility } i \text{ MOS}}{\text{Stratum } h \text{ MOS}} \\
 &= (1,651) \frac{\text{Stratum } h \text{ MOS}}{\text{Total MOS}} \frac{\text{Facility } i \text{ MOS}}{\text{Stratum } h \text{ MOS}} \\
 &= (1,651) \frac{\text{Facility } i \text{ MOS}}{\text{Total MOS}}
 \end{aligned}$$

However, before the sample was selected, these selection probabilities were modified to minimize overlap with the 1995 NNHS, which was conducted by NCHS and fielded in late 1995. Because the NNHS and MEPS NHC used similar sampling frames, it was important to prevent (if possible) any nursing homes from being included in both surveys.

A Keyfitz procedure was used to adjust the probabilities of selection to minimize this overlap. This

procedure provides the desired unconditional probabilities of selection for the MEPS NHC sample while at the same time minimizing the overlap. To compute conditional probabilities of selection for the MEPS NHC, the probabilities of selection for a facility in both the MEPS NHC and the 1995 NNHS frame, as well as which nursing homes were selected in the 1995 NNHS, must be known. The following notation describes the procedure.

$P(i \in \text{MEPS NHC})$ = Probability that the i th nursing home in the NHPI is selected for the MEPS NHC

$P(i \in \text{NNHS})$ = Probability that the i th nursing home in the NHPI is selected for the 1995 NNHS

$P(i \in \text{MEPS NHC} \mid i \in \text{NNHS})$ = Conditional probability that the i th nursing home is selected for the MEPS NHC given that it was selected for the 1995 NNHS

The unconditional probability of selection for a facility in the MEPS NHC can be written as:

$$P(i \in \text{MEPS NHC}) = P(i \in \text{MEPS NHC} \mid i \in \text{NNHS}) P(i \in \text{NNHS}) + P(i \in \text{MEPS NHC} \mid i \notin \text{NNHS}) P(i \notin \text{NNHS}).$$

From this statement, expressions for the conditional probabilities of selection $P(i \in \text{MEPS NHC} \mid i \in \text{NNHS})$ and $P(i \in \text{MEPS NHC} \mid i \notin \text{NNHS})$ for the facilities on the MEPS NHC frame can be derived. The actual conditional probabilities of selection for the MEPS NHC depend on which of the two situations described below applies.

Case 1: $P(i \in \text{MEPS NHC}) \geq 1 - P(i \in \text{NNHS})$. For this case, set

$$P(i \in \text{MEPS NHC} \mid i \in \text{NNHS}) = \frac{P(i \in \text{MEPS NHC}) - (1 - P(i \in \text{NNHS}))}{P(i \in \text{NNHS})}$$

and

$$P(i \in \text{MEPS NHC} \mid i \notin \text{NNHS}) = 1$$

Case 2: $P(i \in \text{MEPS NHC}) < 1 - P(i \in \text{NNHS})$. Here, set

$$P(i \in \text{MEPS NHC} \mid i \in \text{NNHS}) = 0$$

and

$$P(i \in \text{MEPS NHC} \mid i \notin \text{NNHS}) = \frac{P(i \in \text{MEPS NHC})}{1 - P(i \in \text{NNHS})}$$

Given the outcome of the 1995 NNHS sampling, it is shown below that selecting the MEPS NHC sample

with these redefined probabilities preserves the original MEPS NHC probabilities of selection.

For Case 1,

$$P(i \in \text{MEPS NHC}) = \frac{P(i \in \text{MEPS NHC}) - (1 - P(i \in \text{NNHS}))}{P(i \in \text{NNHS})} P(i \in \text{NNHS}) + 1 * (1 - P(i \in \text{NNHS})) = P(i \in \text{MEPS NHC})$$

For Case 2,

$$P(i \in \text{MEPS NHC}) = 0 * P(i \in \text{NNHS}) + \frac{P(i \in \text{MEPS NHC})}{1 - P(i \in \text{NNHS})} * (1 - P(i \in \text{NNHS})) = P(i \in \text{MEPS NHC})$$

After these rules were applied in the six noncertainty strata, the facilities were selected for the MEPS NHC using the redefined selection probabilities. Selection was done using a systematic selection process. For this selection the file was sorted by Beale code and ownership to create implicit sampling strata within each of seven explicit strata. In the certainty stratum, the nine facilities not selected for the NNHS sample were sampled with certainty. The remaining 11 facilities were assigned a zero probability of selection. The outcome of using these probabilities of selection was that none of the MEPS NHC sample facilities overlapped with the NNHS sample.

Two additional certainty facilities were selected in two of the noncertainty strata. These two facilities were Case 1 situations and were not selected in the NNHS sample, so their Keyfitz probabilities were set equal to one. The remainder of the sample in the six noncertainty strata were Case 2 facilities.

Cost Stratification

After the first-phase sample was drawn, the sampled facilities were assigned to four cost strata based on the geographic distribution of the sample. The cost strata were approximated by measuring distance in kilometers from the nearest of the 50 largest U.S. cities. Specifically, each facility was assigned to one of four cost strata:

- Stratum 1: Full workload in a single geographic area, such as a city.
- Stratum 2: Partial workload only in a single area, requiring considerable travel.

- Stratum 3: Single facility requiring considerable travel but within the range of other facilities.
- Stratum 4: Single facility at a distance requiring air travel.

The cost stratification process consisted of several steps. The first-phase sample of 1,651 facilities was mapped using computer mapping software. Next, each facility was mapped into the appropriate ZIP Code center point. Then, to approximate the cost strata, a map of the 50 largest U.S. cities and concentric zones around them was overlaid on the facility map. Facilities located within 100 kilometers of a city were assigned to Stratum 1; facilities 100-200 kilometers, to Stratum 2; facilities 200-300 kilometers, to Stratum 3; and facilities beyond 300 kilometers of a city, to Stratum 4.

Minimizing Overlap With the Medicare Current Beneficiary Survey

The Medicare Current Beneficiary Survey (MCBS) is an ongoing survey of Medicare beneficiaries conducted by Westat for the Health Care Financing Administration. As part of this survey, Westat field interviewers visit many nursing homes throughout the United States. As with the NNHS, it was important to minimize the number of nursing homes involved in both surveys. However, an alternative to the Keyfitz procedure was necessary because of the virtual impossibility of calculating the probabilities of selection for the MCBS facilities.

The procedure used was to flag any nursing home that MCBS respondents reported as their current residence as of September 28, 1995, and that was also in

the first-phase MEPS NHC sample. There were 71 such facilities. Of these overlapping facilities, one MEPS NHC noncertainty facility was removed from the first-phase sample prior to sampling in the second phase, thus giving a zero chance of selection. An adjustment factor was applied to the weights within each cost stratum to prevent an undercoverage bias. The nine facilities that were included with certainty in the first phase of MEPS NHC sampling were designated to be selected with certainty in the second phase, regardless of whether they overlapped with the MCBS. Based on opinions of health care analysts at AHCPR, facilities excluded in this way were unlikely to differ in any systematic way from other facilities in the first-phase sample. Thus, this procedure was not expected to cause any sampling bias.

Selecting the Second-Phase Sample

An equal probability subsample of the initial sample was drawn within each cost stratum using systematic sampling. The sample size for each cost stratum was

determined by optimum allocation. The optimum allocation was computed using the formula

$$n_h = n \frac{W_h / \sqrt{c_h}}{\sum_k (W_k / \sqrt{c_k})} \equiv nr_h$$

where W_h and r_h represent the population proportion and sampling rate for the h th stratum. This formula neglects the variance for the analysis variables, since it is expected that they would vary little between cost strata. The MEPS NHC facilities overlapping with MCBS were not removed prior to determining the optimal allocation, since these facilities will be treated in sample weighting as nonrespondents.

The optimum allocation based on the MEPS NHC first-phase sample is shown in Table 3. The optimal subsampling rates ranged from .78 to .89. The proportions (W_h) shown in Table 3 are those obtained in the MEPS NHC sample of 1,651 facilities. The data collection cost estimates include travel costs, interviewer per diem and salary, and data processing costs.

Table 3. Optimum allocation to cost strata based on the 1996 MEPS Nursing Home Component sample

Stratum	Cost per facility	W_h	First-stage sample	Second-stage sample	Optimal sampling rate	Main sample	Reserve sample
1	\$2,216	0.52	855	765	0.894	616	149
2	\$2,583	0.26	439	363	0.827	292	71
3	\$2,335	0.16	255	222	0.872	178	44
4	\$2,949	0.07	102	80	0.781	64	16
Total	—	—	1,651	1,430	—	1,150	280

Note: The stratum proportions are based on the MEPS Nursing Home Component sample. Costs are based on the data collection budget for the MEPS Nursing Home Component and include travel costs, interviewer per diem and salary, and data processing costs. Based on proximity to large U.S. cities, the strata were defined as follows: Stratum 1—Full workload in a single geographic area, such as a city. Stratum 2—Partial workload only in a single area, requiring considerable travel. Stratum 3—Single facility requiring considerable travel but within the range of other facilities. Stratum 4—Single facility at a distance requiring air travel.

Source: Agency for Health Care Policy and Research: 1996 Medical Expenditure Panel Survey Nursing Home Component.

Sampling Algorithms

This section describes in detail the algorithms used to select the main and reserve samples. The following notations are used in this section:

S_0	=	event the facility is selected for the first-phase sample
S_1	=	event the facility is selected for the second-phase sample
n_0	=	overall first-phase sample size
n_{h0}	=	first-phase sample size in stratum h
n_1	=	overall second-phase sample size
$n_{h'1}$	=	second-phase sample size in cost stratum h'
m	=	overall main sample size
$m_{h'}$	=	main sample size in cost stratum h'
r	=	overall reserve sample size
$r_{h'}$	=	reserve sample size in cost stratum h'
c_{h0}	=	number of first-phase certainty selections in stratum h
$c_{h'1}$	=	number of second-phase certainty selections in cost stratum h'
M_{hi}	=	measure of size for i th facility in stratum h
N_h	=	number of facilities on the frame in stratum h
$\pi_{hi}^{(0)}$	=	initial selection probability for i th facility in stratum h
$\pi_{hi}^{(1)}$	=	final selection probability for i th facility in stratum h

First-Phase Sample

Both phases of the facility sampling were accomplished using Westat's macro WESSAMP. Probability proportional to size (PPS) systematic sampling was used in the first phase, and equal probability systematic sampling in the second phase. In the first phase, the unconditional probability of selection for the i th facility was $n_h M_{hi} / M_h$, where M_{hi} is the measure of size for the i th facility in stratum h ($h = 1, 2, \dots, 7$), M_h is the sum of the measures of size in the stratum, and n_h is the number of facilities sampled in the stratum. Any facility with unconditional probability of selection greater than or equal to 1 was classified as a certainty selection and assigned a selection probability equal to 1. Two facilities in the six noncertainty strata met this criterion. In the certainty stratum, there were 20 large facilities, of which 9 were not sampled in NCHS's NNHS. These were taken with certainty for the MEPS NHC. In the large stratum, the remaining 11 facilities had their conditional probabilities set to 0. In the six noncertainty strata, these selection probabilities were modified to minimize overlap with the NNHS, as described previously. The modified probabilities of selection resulted in two additional facilities being selected with certainty.

Thus the sampling algorithm for the first phase consisted of this step:

Step 1. Within each stratum, sort the facilities by Beale code, type of ownership, and ZIP Code. Calculate the conditional (Keyfitz) probabilities of selection. Select n_{h0} facilities with PPS, with the Keyfitz probability of selection as the measure of size. There will be c_{h0} certainties, i.e., facilities that will have $\pi_{hi}^{(0)} = 1$. For the other facilities, the original unconditional selection probabilities will be

$$\pi_{hi}^{(0)} = \frac{(n_{h0} - c_{h0}) M_{hi}}{N_h - c_{h0} \sum_{i=1} M_{hi}}$$

where M_{hi} is the measure of size for the i th facility in the h th stratum.

Second-Phase Sample

The first-phase sample of 1,651 facilities was mapped into four cost strata and subsampled within each cost stratum. The sample size in each cost stratum was determined using optimal allocation. Equal stratum variances were assumed for MEPS NHC variables. Within each cost stratum, the certainty facilities and noncertainty facilities identified as overlapping with MCBS were first removed. The sample was then sorted by the same order of selection used in the first-phase sample, and an equal probability systematic sample of facilities was drawn with the sample sizes shown in Table 3. The resulting second-phase sample of 1,430 facilities was again sorted within cost strata by the order of selection, and the noncertainty facilities were subsampled again at a rate of .803 to create a randomly chosen reserve sample of 280 facilities and a main sample of 1,150 facilities. The reserve sample was split into four release groups of 70 facilities each by sequentially assigning the numbers 1 through 4 to the facilities in their original sort order. Using the same procedure, the noncertainty facilities in the main sample were randomly divided into 18 recall groups consisting of approximately 64 facilities each.

Thus the sampling algorithm for the second phase consisted of these steps:

Step 2. Map the sample of n_0 facilities into the four cost strata using facility ZIP Code and mapping software.

Step 3. Remove noncertainty facilities identified as overlapping with MCBS and certainty facilities from the first-phase sample.

Step 4. To select the second-phase sample of n_1 facilities from the first-phase sample of n_0 , sort the facilities in each cost stratum in the original order of selection. Within each cost stratum, draw an equal probability systematic sample of facilities, where the sample size is determined by optimum allocation. (See Table 3.) Subtract the number of first-phase certainty facilities in each cost stratum from the designated sample size in Table 3 prior to sampling.

Step 5. To select the reserve sample of r facilities from the n_1 second-phase facilities, first sort the noncertainty facilities in the second-phase sample by order of selection in each cost stratum. Within each cost stratum, select an equal probability systematic sample of facilities using the sample sizes in Table 3. Create four release groups by sorting the reserve sample in the order of selection, then consecutively numbering the reserve sample from 1 to 4, repeating until the entire reserve sample has been assigned. There will be $m = n_1 - r$ facilities in the main sample and r facilities in the reserve sample. The reserve sample will consist of four release groups of $r/4$ facilities each.

Step 6. To create the 18 recall groups from the main sample, sort the noncertainty facilities in the main sample in the order of selection, then consecutively number facilities from 1 to 18, repeating until all noncertainty main sample facilities have been assigned. Each recall group thus will represent a random subsample of the main sample.

For a two-phase sampling process like this, the sampling probabilities for the i th facility in the h th stratum can be written as:

$$\pi_{hi}^{(1)} = P(hi \in S_1 | hi \in S_0) P(hi \in S_0).$$

For the “initial certainty/final certainty” facilities—facilities that were selected with certainty in both the first and second phases of sampling—the overall selection probability is 1.00.

For the “initial noncertainty/final noncertainty” facilities, the final selection probability would be:

$$\pi_{hi}^{(1)} = \frac{m_{h'}}{n_{h'0}} \frac{(n_{h0} - c_{h0}) M_{hi}}{\sum_{i=1}^{N_h - c_{h0}} M_{hi}}$$

where $m_{h'}$ is the main sample size in cost stratum h' and $n_{h'0}$ is the number of first-phase sample facilities in cost stratum h' . If release groups are used, the numerator in the first factor is increased by the extra number of facilities released. If no release groups are used but

some recall groups are withdrawn, the numerator is decreased by the number of facilities being withdrawn in cost stratum h' .

Initial Screening of Facilities

An initial screening was carried out by telephone. Only facilities meeting the following requirements were retained in the sample:

- Facilities must have three or more beds that are staffed and set up for residents (or a distinguishable group of three or more beds within a facility).
- Facilities must either be:
 - Medicare certified as a skilled nursing facility and/or Medicaid certified as a nursing facility, or
 - Licensed as a nursing home with a registered nurse or licensed practical nurse onsite 24 hours a day, 7 days a week.

As a result of the screening, 14 facilities were identified as being out of business and 1 facility was determined to be ineligible.

Round 1 Facility Response Rates

Given the response rate assumptions specified in Table 4, the initial sample sizes were intended to result in a final sample of approximately 787 cooperating facilities, with control over the final sample size to be obtained through the use of release and recall groups. At the end of Round 1, 1,124 of the 1,150 facilities fielded for data collection were determined to be eligible. Of these, 951 completed the Facility Questionnaire and sampling of current residents, 158 refused, and 14 broke off the interview. Twelve facilities were ineligible and 15 had gone out of business. The Round 1 response rate to the Facility Questionnaire was 85 percent and the eligibility rate was 98 percent, both exceeding expectations. Based on these data, AHCPR made a decision to recall two groups of facilities, totaling 127 facilities, from Rounds 2 and 3 of data collection. Of these, 108 had cooperated in Round 1.

Table 4. Minimum acceptable response rates for the 1996 MEPS Nursing Home Component

Instrument	Response rate	Method of calculation
Facility Questionnaire (FQ)	77%	<u>Responding facilities</u> Selected eligible facilities
Jan. 1, 1996, sampling list	98%	<u>Facilities completing sampling</u> Facilities completing FQ
Round 2 admissions sampling list	97%	<u>Facilities completing sampling</u> Facilities cooperating on previous contacts
Round 3 admissions sampling list	97%	<u>Facilities completing sampling</u> Facilities cooperating on previous contacts
Institutional use and expenditure data (Jan. 1, 1996, residents)	91%	Sampled persons with complete institutional use and expenditure data for all of 1996/All sampled and eligible Jan. 1, 1996, residents
Institutional use and expenditure data (first admissions)	89%	Sampled persons with complete institutional use and expenditure data for all of 1996/All sampled and eligible first admissions
Residence history data (Jan. 1, 1996, residents)	95%	Sampled persons with complete residence history data for all of 1996/All sampled and eligible Jan. 1, 1996, residents
Residence history data (first admissions)	90%	Sampled persons with complete residence history data for all of 1996/All sampled and eligible first admissions
Background data (Jan. 1, 1996, residents)	85%	Sampled persons with data/All sampled and eligible January 1, 1996, residents
Background data (first admissions)	82%	Sampled persons with data/All sampled and eligible first admissions
Baseline health status data (Jan. 1, 1996, residents)	96%	Sampled persons with data/All sampled and eligible Jan. 1, 1996, residents
Baseline health status data (first admissions)	92%	Sampled persons with data/All sampled and eligible first admissions
End-of-year health status data (Jan. 1, 1996, residents)	91%	Sampled persons with data/All sampled and eligible Jan. 1, 1996, residents residing in an eligible facility on Dec. 31, 1996
End-of-year health status data (first admissions)	89%	Sampled persons with data/All sampled and eligible first admissions residing in an eligible facility on Dec. 31, 1996

Source: Agency for Health Care Policy and Research: 1996 Medical Expenditure Panel Survey Nursing Home Component.

Field Problems

Nursing Homes Associated With Other Facilities

Sampled facilities belonging to nursing home chains were identified before screening to assist both the recruiters and Round 1 interviewers. Facilities that were affiliated with hospitals or retirement centers and facilities with board-and-care wings were also given special attention during the training of field staff. During Round 1 facility data collection, if the facility respondent identified the facility or unit as a hospital-based skilled nursing facility, the hospital name was added to the place roster in the Facility Questionnaire and a flag was set to indicate that the hospital has a skilled nursing facility unit. Interviewers were instructed to carefully identify and list residents only of those parts of the facility that were eligible for the MEPS NHC.

Facilities That Moved or Combined With Other Facilities

During screening it was discovered that some facilities were no longer located at the address given for them in the NHPI. Facilities that had moved were retained in the sample and followed to the new location. If the new location was not learned until fieldwork was underway, the facility was assigned to a new interviewer if necessary to complete data collection.

A sampled facility that combined with another facility was retained in the sample as long as the other facility was not listed in the NHPI. If both of the original facilities were listed separately in the NHPI, the combined facility had an increased chance of selection because it could have been selected through either one of the original facilities. Either this increased chance of selection had to be accounted for in the facility weight or, alternatively, one of the listings had to be considered out of scope. When the combined facilities could be treated as multiple units of one nursing home, the latter approach was used. Otherwise, weighting adjustments were made.

Facilities With Multiple Units

When the Facility Questionnaire was administered, the sampled facility was sometimes discovered to correspond to more than two eligible facilities or to a facility with more than one unit containing eligible nursing home beds. If any of the facilities (or units of one facility) associated with the sampled facility were listed separately in the NHPI frame, they were considered out of scope because they had already had a chance to be selected. Thus, each facility had only one chance of selection, thereby avoiding the need to make an adjustment to the facility base weight for multiple chances of selection. If none of the nursing homes associated with the sampled facility were listed in the frame, the interviewer was instructed to collect data from all of them if time and travel distance permitted. If this was not practical, the plan was to subsample in facilities that contained three or more eligible units or locations where there were too many units to permit data collection from all of them. However, during Round 1 it was not necessary to do this. An alternative plan was to assign some of the units to another interviewer. The rules for deciding which units were eligible are given in Table 5.

Survey Database

A database of the sampled facilities was created and loaded into each computer-assisted personal interviewing (CAPI) machine for the field staff to use in the sampling of residents. Each record contained the following data:

- Facility name, address, and telephone number;
- Numbers of residents and eligible beds from the NHPI;
- Final measure of size;
- The random numbers used for sampling current residents and first admissions.

Table 5. Rules for facility sampling: 1996 MEPS Nursing Home Component

Listed in updated National Health Provider Inventory	Sampled unit	
	Headquarters	One of subunits
Headquarters and all subunits	None are eligible	Only sampled subunit is eligible
Headquarters only	All are eligible	—
All subunits but not headquarters	—	Only sampled subunit is eligible
Subset of subunits is listed; headquarters not listed	—	Sampled unit and unlisted subunits are eligible ¹
Headquarters and subset of subunits are listed	Unlisted subunits are eligible	Only sampled subunit is eligible

¹Either revise computer-assisted personal interviewing (CAPI) to subsample or review in home office for weighting corrections.

Source: Agency for Health Care Policy and Research: 1996 Medical Expenditure Panel Survey Nursing Home Component.

Sampling of Persons Within Facilities

The sample of nursing home residents consisted of samples of persons who resided in institutions on January 1, 1996 (the 1996 current residents sample) and persons who were admitted to institutions at any time from January 1 through December 31, 1996 (the first-admissions sample). The subset of admissions being admitted to a nursing home for the first time in 1996 constituted the eligible first-admissions sample. A more detailed definition of an eligible first admission is given later. The current-residents and first-admissions samples cover the entire population of persons residing in nursing homes during 1996. After all three rounds, the target sample sizes of residents for the 787 cooperating facilities were 3,043 eligible current residents and 2,218 eligible first admissions with complete expenditure and residence history data. These target sample sizes resulted from the number of sampled persons with complete use and expenditure data that were expected after sampling four current residents and four to six first admissions per facility. Two to three first admissions per facility in each of Rounds 2 and 3 were sampled. A fixed sample size per

facility was chosen instead of sampling from each list at a fixed rate because the former method is more reliable for obtaining the desired sample sizes. As a consequence, however, the first-admissions sample weights are not equal across sampling periods, nor are they exactly equal across nursing homes. To lessen the variability of the first-admissions sampling weights, the sample size is permitted to range from 2 to 3.

Checking Facility Data Against Frame Data

The following procedure was implemented by the field interviewers during their visits to the sample institutions in Round 1. During the first visit to the facility, the interviewer made a list of eligible current residents. The interviewer entered the number of current residents on the list into the CAPI system. The computer compared the number of residents listed with the measure of size derived from the NHPI and displayed the message “Call Home Office” if any of the following were true for r_1 , the number of eligible beds listed in the NHPI, or r_2 , the number of current residents listed at the facility:

- $r_2 < 10$ and $|r_1 - r_2| > 5$;
- $10 \leq r_2 \leq 300$ and $r_2/r_1 > 1.5$ or $r_2/r_1 < .5$; or
- $r_2 > 300$ and $r_2/r_1 > 1.33$ or $r_2/r_1 < .67$.

If the nursing home facility existed within a long-term care facility, the interviewer verified that the number of residents listed corresponded to the eligible portion of the facility. The interviewer also verified that no eligible portions of the facility were overlooked.

Current-Residents Sample

The interviewer compiled a list of current residents as of January 1, 1996, in each sampled facility. Within each facility, a systematic random sample of four current residents was drawn within the CAPI system. The within-facility sampling fraction was assigned to be $4/CR_{hi}$, where CR_{hi} is the number of current residents listed at the i th facility in the h th stratum, so that the overall probabilities of selection of current residents within strata were as close to equal as possible. The probabilities of selection were not exactly equal because

the measure of size used to select facilities was the number of beds; however, to the extent that the number of current residents was correlated with the number of beds at the facility, the selection probabilities were approximately equal. In facilities with fewer than four residents, the sampling fraction was set to one and all residents were included.

The interviewer entered the size of the list of current residents in the CAPI system, which then determined the random start, the skip interval, and the sample of line numbers. The selected line numbers were displayed on the computer screen and stored in memory for later validation. The order of selection for the sampled current residents was stored for inclusion in the final database. At the end of Round 1, the response rates shown in Table 6 were obtained for current residents.

The overall response rate for the current-residents sample is 98.8 percent. Sampled residents needed to have 75 percent of their baseline health status items complete and age, sex, and race reported in order to be considered respondents. Forty-four eligible current residents did not meet this requirement. Of these, four met the baseline health status criteria but were missing at least one of the demographic variables. In addition, 17 sampled persons were ineligible.

Table 6. Response rates at end of Round 1: 1996 MEPS Nursing Home Component

Questionnaire module	Completed		Partial response		Nonresponse	
	Number	Percent	Number	Percent	Number	Percent
Residence history	3,725	98	22	1	44	1
Background	3,621	96	13	0	157	4
Income/assets	—	—	—	—	—	—
Health insurance	3,545	94	38	1	208	5
Baseline health status	3,751	99	10	0	30	1
Prescribed medicines	3,584	95	145	4	62	2

Source: Agency for Health Care Policy and Research: 1996 Medical Expenditure Panel Survey Nursing Home Component.

First-Admissions Sample

First-Admissions Sample Size

Lists of residents were obtained from the sampled facilities and screened to determine who had been newly admitted since the last round of data collection. Listing, sampling, and data collection for first admissions took place in Rounds 2 and 3. The reference period for Round 2 was from January 1 to June 30, 1996, and the reference period for Round 3 was from July 1 to December 31, 1996. The first admissions were systematically sampled in the same manner as the

current-residents sample, except that the sample size was determined in the CAPI program. The order of selection for each sampled first admission was stored within the program. If the measure of size differed substantially from the number of current residents listed, then the first-admissions sample probabilities of selection would have led to excessive variability in the first-admissions sampling weights if not corrected.

Thus, the sample size for the first-admissions sample at a given facility may be revised based on the relationship between the current residents and the number of first admissions listed. The revised sample size was based on the selection probability:

$$\pi = (\text{Facility sample size}) \frac{\text{Facility MOS}}{\text{Total MOS}} \frac{\text{FA sample size}}{\text{Number of FAs listed}}$$

where

Facility sample size = the number of facilities sampled;
 FA sample size = the number of first admissions sampled at the given facility;
 Number of FAs = the number of first admissions listed at the given facility;
 Facility MOS = the number of nursing home beds on the frame for the facility; and
 Total MOS = the total number of nursing home beds for facilities on the frame.

In order to have an approximately self-weighting sample:

$$\pi = \frac{\text{Overall FA sample size}}{\text{Total FAs}}$$

Thus, in order to have equal selection probabilities, the FA sample size should be:

$$\begin{aligned} \text{FA sample size} &= \frac{\pi (\text{Total MOS})}{\text{Facility sample size}} \frac{\text{Number of FAs listed}}{\text{Facility MOS}} \\ &= \frac{\text{Overall FA sample size}}{\text{Total FAs}} \frac{\text{Total MOS}}{\text{Facility sample size}} \frac{\text{Number of FAs listed}}{\text{Facility MOS}} \\ &= \frac{\text{Overall FA sample size}}{\text{Facility sample size}} \frac{\text{Total MOS}}{\text{Total FAs}} \frac{\text{Number of FAs listed}}{\text{Facility MOS}} \\ &= (\text{Average FA sample size}) \frac{\text{Number of FAs listed}}{\rho(\text{Facility MOS})} \end{aligned}$$

where

$$\rho = \frac{\text{Total FAs}}{\text{Total MOS}} = \frac{\text{Average FAs}}{\text{Average MOS}}$$

Thus, the first-admissions sample sizes were adjusted upward or downward according to whether more or fewer were listed, based on the measure of size adjusted by the factor ρ to reflect the average number of first admissions to residents. However, the within-facility first-admissions sample size was not permitted to exceed three per round, and was less than two only when there were fewer than two first admissions in the facility for the round. Although ρ is unknown, it can be approximated using 1987 NMES data on the ratio of nursing home admissions to residents. The value of ρ using 1987 NMES data turns out to be $718,670/1,523,540 = .472$ (Agency for Health Care Policy and Research, 1990).

Eligibility Determination

Since residents could be admitted to a facility more than once during the course of the reference period, more than one record may have existed for some persons on the facility's list. Interviewers deleted duplicates so that no individual appeared on the list more than once. The interviewer then selected two or three first admissions per facility per round of data collection, using the CAPI software in the same manner as for current residents.

An eligible first admission is defined as a person with no admissions or stays on or after January 1, 1996, in MEPS NHC-eligible facilities prior to the admission for which the person was sampled at the primary sampled facility. Information about where the person lived from January 1, 1996, until the date of admission to the sampled facility, referred to as the pre-stay period, was collected from facility respondents. Using CAPI, data were collected on the beginning and ending dates for each separate period of residence during the pre-stay period, the name and type of each place where the sampled person stayed, and whether the person stayed at that place the whole time between the beginning and ending dates. Place types include the sampled facility, community residence, acute care or long-term care hospitals, and other long-term facilities. All places of residence provisionally identified as long-term care facilities were searched on the NHPI file for a determination of nursing home eligibility status. Since this would include hospitals with long-term care skilled

nursing units, the AHA file also was searched during residence history data collection to determine first-admission eligibility.

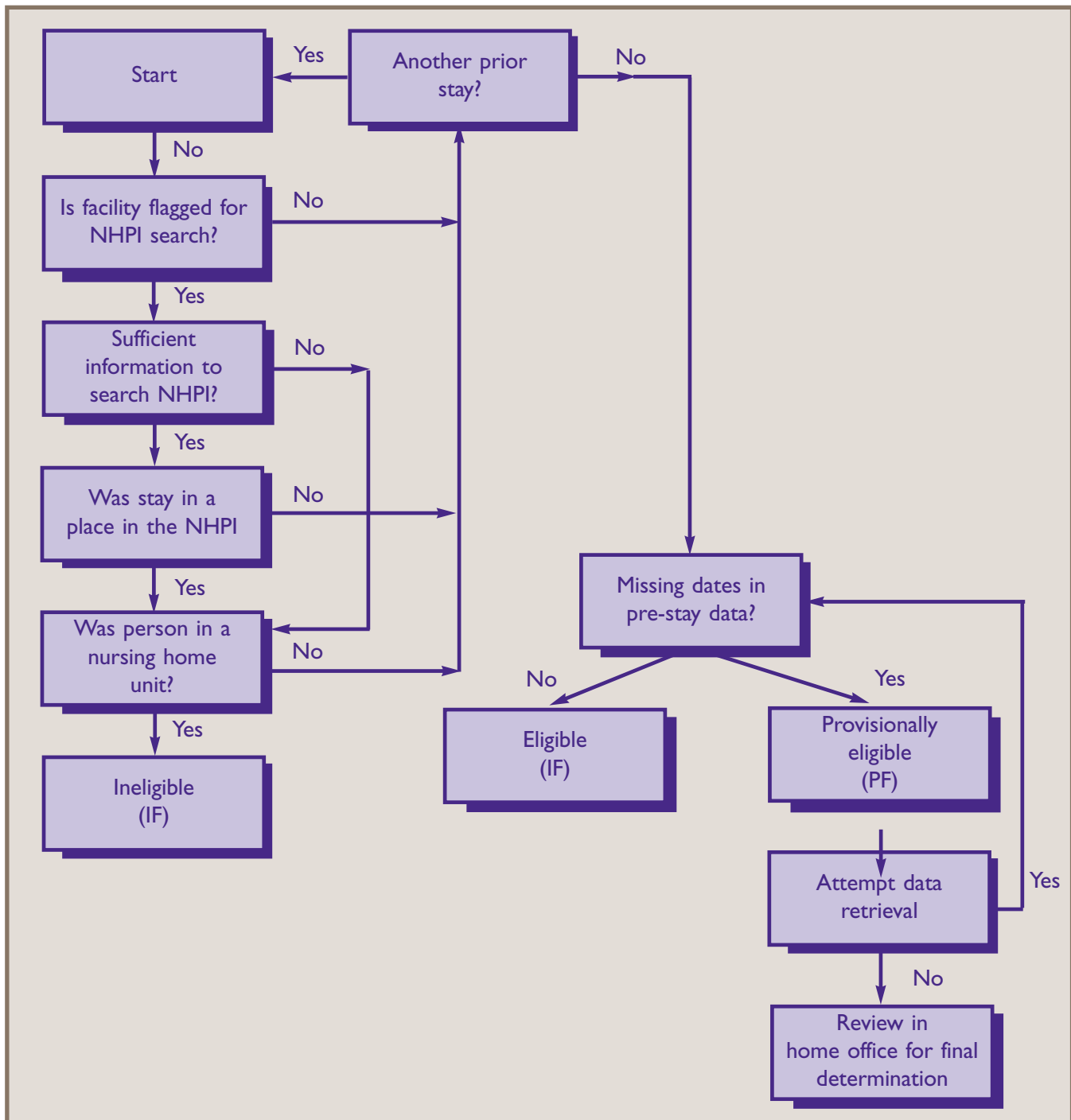
As an aid in determining eligibility, the NHPI and AHA list (American Hospital Association, 1993) were loaded into the interviewers' laptop computers and incorporated into their CAPI software. A search software program allowed the field interviewers to search for an identified long-term care facility on the NHPI or AHA files in different ways, including name, address, State, and telephone number. Interviewers were able to conduct searches based on portions of the information to maximize the likelihood of finding matches. At the conclusion of the pre-stay residence history data collection, the CAPI system automatically brought the interviewer to the NHPI and AHA search functions to search for matches to reported long-term care facilities. Interviewers were trained to search for the facility name and, if that failed, to use the facility address and telephone number. Statisticians could verify NHPI and AHA searches at any time in the home office.

Based on information collected from the facility about prior admissions to other facilities, the sampled admissions were classified as eligible (with no prior stay in an eligible facility during the reference period), ineligible (with one or more prior stays identified), or indeterminate (with some period of time within the reference period for which the facility could not report whether the resident was in an eligible facility).

Figure 1 shows the data collection process and the flowchart for determining eligibility. There were four possible outcomes, each having a different protocol for data collection:

- Eligible first admission: No admissions prior to sampling.
 - All data collection continues.
- Ineligible first admission: One or more admissions prior to sampling.
 - All data collection stops.
- Provisionally eligible first admission: Eligibility cannot be determined because either the facility has a gap in the pre-stay data or there was an admission to a facility but the name of the facility is unknown or it did not match the NHPI listing.

Figure 1. First admissions eligibility determination using a single respondent: 1996 MEPS Nursing Home Component.



Note: EF=eligible first admission. IF=ineligible first admission. NHPI=National Health Provider Inventory. PF=provisionally eligible first admission.

Source: Agency for Health Care Policy and Research: 1996 Medical Expenditure Panel Survey Nursing Home Component.

- All data collection continues.
- Sampling error: The sampled admission is listed twice and the entry sampled is the second one.
- All data collection stops.

For persons who were eligible or indeterminate, interviewers attempted to complete a community residence history for the pre-stay period by telephoning a knowledgeable community respondent (usually a relative). Information from the community residence history questionnaires was consulted to make an eligibility determination for persons in the indeterminate group. Persons found to be ineligible on the basis of either the facility data or the community data were dropped from further data collection.

The process to determine eligibility was the same for both the facility and the community pre-stay data. Figure 2 shows the rules used to make a final eligibility determination. This process was implemented if determinations of eligibility were made using data from both the facility and community respondent. Table 2 shows the number of sampled first admissions expected to be eligible and the expected final first-admissions sample sizes.

Resolution of Sampling Errors

A number of types of sampling errors can occur. In most cases, the interviewer should have notified the home office of the situation, continuing data collection until a contact with the home office resulted in instructions to proceed otherwise. These sampling errors and their resolution were, for the most part, handled in the CAPI software. Some errors that might occur, and their resolution, are listed below. The first three types could be resolved in the CAPI software.

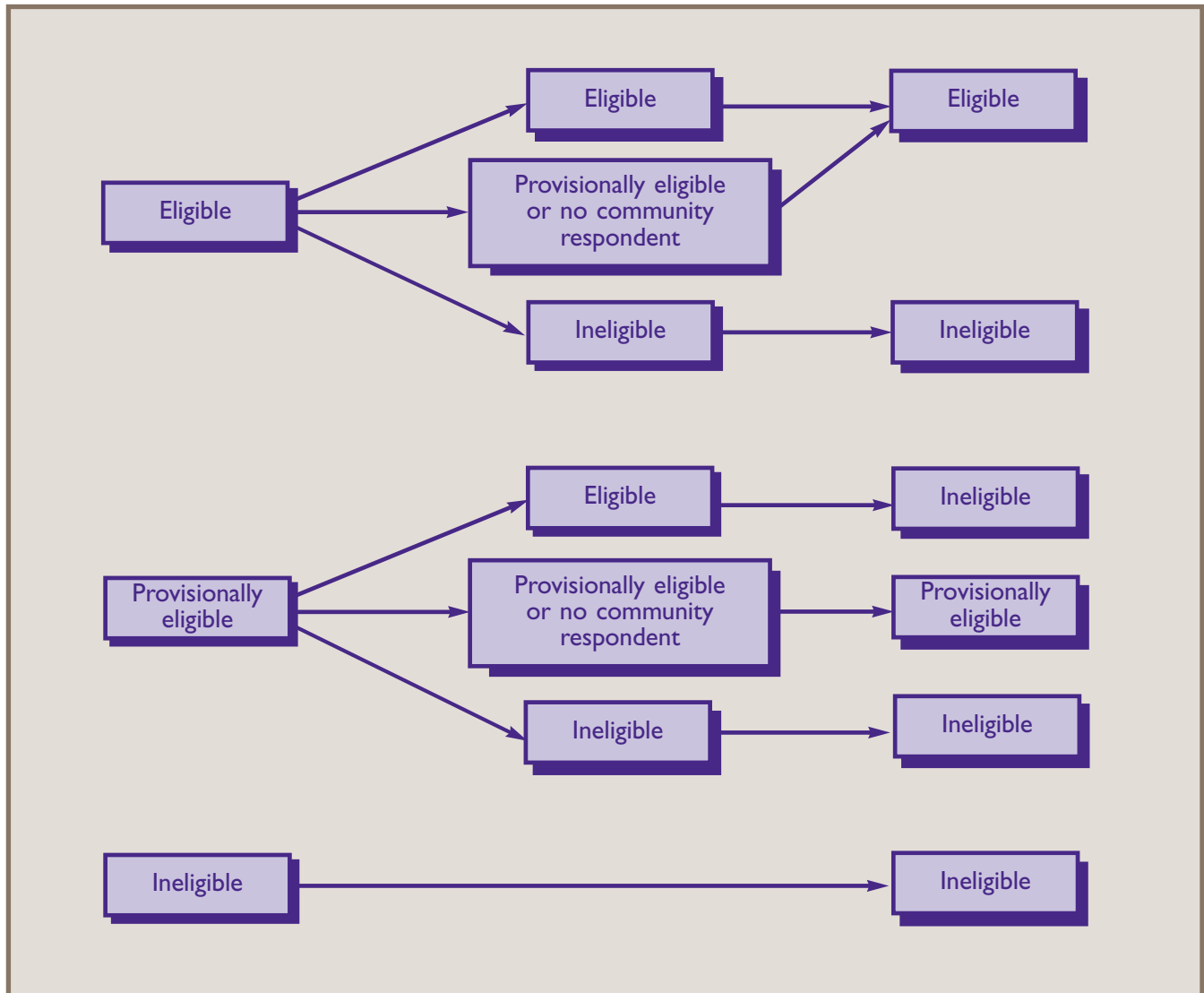
- Person sampled as a first admission was a resident on January 1, 1996.
Resolution:
 - First admission was not listed for current-residents sample: Drop from the first-admissions sample and code as a sampling error; adjust current-residents sample weights.
 - First admission was listed for both current-residents and first-admissions samples: Drop

from the first-admissions sample and code as a sampling error.

- Person sampled as a current resident was not a resident on January 1, 1996, but was admitted later.
Resolution:
 - Current resident was not listed for first-admissions sample: Drop from the current-residents sample and code as a sampling error; add to first-admissions list before first-admissions sampling to ensure a chance of selection.
 - Current resident was listed for first-admissions sample: Drop from the current-residents sample.
- Person sampled as a first admission was admitted and listed twice or more.
Resolution:
 - First admission was sampled on first admission: Retain first admission in the sample.
 - First admission was sampled on later admission: Drop first admission from the sample.
- Eligible persons were omitted from listing.
Resolution: Call home office; adjust sampling weights.
- Ineligible persons (e.g., residents of an assisted living wing) were listed or sampled.
Resolution: Call home office; clean list and resample, if possible. May require CAPI intervention from home office to allow resampling. If resampling is not possible, CAPI software detects ineligible sampled persons in the residence history questionnaire and they are dropped from the sample and coded as out of scope.

These resolutions are not perfect. While they were intended to preserve the rule of a single chance of selection, they do not preserve the clear stratification of the current-residents versus first-admissions samples. In each case, the sampled person being dropped could instead be retained if proper adjustments were made to the sampling weights. It should be noted, however, that this latter resolution would not preserve the stratification of the two samples either.

Figure 2. Final determination of eligibility for sampled nursing home residents: 1996 MEPS Nursing Home Component.



Source: Agency for Health Care Policy and Research: 1996 Medical Expenditure Panel Survey Nursing Home Component.

References

Agency for Health Care Policy and Research. NMES Public Use File 8: Institutional Population Component, baseline questionnaire. Rockville, MD; 1990.

American Hospital Association. American Hospital Association guide to the health field. Chicago; 1993.

National Center for Health Statistics. Sampling frame for the 1995 National Nursing Home Survey (facility file) [microdata tape documentation]. Hyattsville, MD; 1995.

Pancholi, M. Message to P. Broene in electronic mail system, 1995 Oct 18; Pancholi, M. Unpublished memo to P. Broene, "The Updated Health Provider Inventory," 1995 July 17.

U.S. Department of Health and Human Services
Public Health Service
Agency for Health Care Policy and Research
2101 East Jefferson Street, Suite 501
Rockville, MD 20852



AHCPR Pub. No. 98-0042
September 1998