

Iris Shimizu, National Center for Health Statistics

### 1. Introduction

The National Nursing Home Survey (NNHS) is conducted periodically by the National Center for Health Statistics (NCHS) to collect data on the characteristics of nursing and related care homes, their residents, and their staffs. The NNHS estimates are for all such homes regardless of their participation in federal programs such as Medicare or Medicaid. The prior survey cycles were conducted in 1973 and 1977. This document describes the 1985 NNHS design which differs from the designs used in prior cycles. The design changes and some of their impact are noted in the discussion.

The survey design objectives are stated in the next section while Section 3 describes the universe and sampling frame. Sections 4 and 5 describe the sampling and data collection strategies. Sections 6 and 7 outline the estimators.

### 2. Survey Objectives

As in prior NNHS cycles, the 1985 cycle was to produce estimates about the nursing and related care homes, their employees, their current residents, and their discharges. The employees targeted in the 1985 cycle were the registered nurses (RN's). In addition the 1985 cycle was to produce estimates about both persons admitted and admission events in a 12 month period. The 1985 survey was also for the first time to produce some estimates by type of location [in a Metropolitan Statistical Area (MSA), not in a MSA].

The design was to produce statistics for current residents with the maximum precision for the available funds. The design was also required to reduce the respondent burden per sampled home from that in the 1977 cycle. In 1977, up to eight each of current residents and discharges were sampled in each facility. In the 1985 NNHS, no more than five current residents and six discharges were to be sampled in each facility so interviewers could complete data collection in each facility in a single day. Also, no more than four RN's were to be selected in any facility.

### 3. Target Population and Sampling Frames

The universe for the 1985 NNHS consists of nursing and related care homes in the conterminous United States, their residents, and the registered nurses working in the homes. To be eligible, a home must both have three or more beds set up and staffed for use by persons not related to the home's owner and routinely provide nursing or personal care services to residents. The facility is also freestanding or a nursing care unit of a hospital, retirement center, or similar institution where the unit maintains financial and resident records separate from those of the parent institution. Places providing only room and board are ineligible. Also ineligible are those places serving only persons with specific health problems (for example, mental retardation or alcoholism).

The 1985 NNHS sampling frame consists of

facilities classified as eligible for the survey from four sources. These sources are (1) the 1982 National Master Facility Inventory (NMFI), (2) homes identified in the Agency Reporting System (ARS) as opening for business by June 1, 1984, (3) places certified for intermediate care by the U. S. Health Care Financing Administration unless they serve only persons who are mentally ill or retarded, and (4) those located in the 1982 NMFI Complement Survey. The NMFI is a census of inpatient health facilities conducted through 1982 every two or three years. In 1982 the NMFI was limited to the nursing and related care homes. The ARS is used to identify and survey homes not included in the last NMFI. Complement Surveys based on probability area samples are conducted by the NCHS to measure the completeness of the NMFI. Inclusion of homes from the Complement Survey permits the representation of homes which are eligible for the survey but which, for various reasons, have never been included in any of the lists from which the NMFI is compiled. Details of these surveys are documented (3, 4, 5, 8, 9).

The NNHS sampling frame was frozen on June 1, 1984, to obtain a sample as current as possible and yet allow time to collect data needed for determining eligibility and sampling stratification. Finally, duplicate entries for facilities were removed from the sampling frame.

The resulting sampling frame contained about 20,500 homes. Based on the 1982 NMFI Complement Survey, about 94 percent of the facilities and about 98 percent of the beds eligible for the 1985 NNHS are in the sampling frame.

### 4. Sampling Design

The sampling design for the NNHS is a stratified two stage probability design. Facilities are selected at the first stage and current residents, discharges, and staff members are sampled at the second stage. In 1985 the staff sample was limited to the RN's.

Twenty primary strata of facilities were formed in the 1985 sampling frame on the basis of certification status, Complement Survey status (added, or not added, to the sampling frame from the Complement Survey), and bed size (3-14 beds, 15-24 beds, 25-49 beds, 50-99 beds, 100-199 beds, 200-399 beds, 400-599 beds, 600 or more beds, and unknown). A home was considered certified if, according to the data in the sampling frame, it was certified as either a skilled nursing facility or an intermediate care facility under Titles XVIII and XIX of the Social Security Act for Medicare and Medicaid. Certification was used as a surrogate for type of service (nursing care home or other) which was a stratification variable used in prior NNHS cycles. Data on type of service were absent for many homes in the 1982 NMFI which was the base of the 1985 NNHS sampling frame while information on certification was available for almost all facilities in the frame.

Within each primary sampling stratum, facilities were arrayed by type of ownership (proprietary or unknown ownership, government, and other nonprofit). Within ownership type, facilities

were sorted by the four Census geographic regions, and within region by Metropolitan Statistical Area (MSA) status (in a MSA, not in a MSA) where MSA's are those defined by the Bureau of the Census on the basis of the 1980 Census. Within each MSA status group, homes were arrayed by State. Within State, the homes in MSA's were sorted first by MSA and then by county within MSA while the homes not in a MSA were next sorted by county. Counties were first arrayed by centroid latitude (in groups of 0.5 degrees each) from south to north. Within latitude group, the counties were arrayed by centroid longitude from east to west. The centroids were obtained from the LOKATE<sub>TM</sub> Geo-File (2). While such an array does not place counties in a serpentine order (the preferred geographic order), the process is computerized and it at least places together counties falling in single strip-shaped areas of each state (in prior NNHS cycles, counties could only be ordered alphabetically). Within county homes were arrayed by zip codes and then alphabetically within zip code.

The facility sample within each of the 20 primary strata was then selected systematically with overall probability proportional to frame bed size (reported in the sampling frame). Before sampling, the number of beds in each facility was multiplied by the reciprocal of the probability that the facility was included in the NNHS frame (that probability was always "1" except for homes in the Complement Survey strata). Cumulative

sums of the weighted beds were then computed for each facility in the stratum's array. Facilities corresponding to a systematic random sample of the weighted beds were selected to the NNHS sample. The overall probability of including a particular facility in the sample is generally the product of that facility's probability of being included in the sampling frame times the probability of its being selected from the frame. That product, for most facilities, is simply that facility's frame bed size divided by the corresponding sampling interval. Table A shows the distribution of facilities in the sampling frame and sample according to response status.

The number of facilities selected from each sampling stratum was based primarily on results of research into the optimum sample design for the 1985 NNHS. The design minimized survey costs while achieving fixed precision levels for current resident statistics.

At the second stage of sampling, systematic random sampling was used to select the current residents, discharges, and RN's within sampled facilities. Interviewers constructed the sampling frames for the within facility samples and selected the samples at the time of the survey in each facility. The sampling frame for current residents consists of all residents on the facility's register for the evening prior to the day of the survey at the facility. Residents away from the facility due to overnight leave or hospitalization were included if the facility

TABLE A: Number of facilities in the 1985 NNHS Universe and Sample, by Sampling Strata (Conterminous U.S.A. 1985)

Sampling Strata	Universe (Sampling Frame)	Sample			
		Total	Out-of- Scope	Refused	Respond- ing
All Types of Certification	20,480	1,220	55	84	1,081
Certified	12,985	1056	29	73	954
Complement Survey	24	19	1	2	16
NMFJ and its updates					
3-14 beds	112	5	2	0	3
15-24 beds	384	9	0	0	9
25-49 beds	1,876	43	4	3	36
50-99 beds	5,000	269	9	21	239
100-199	4,605	478	9	28	441
200-299 beds	861	196	1	13	182
300-599 beds	77	17	1	3	13
600 or more beds	26	10	1	0	9
Unknown beds	20	10	1	3	6
Not Certified	7,495	164	26	11	127
Complement Survey	336	14	3	0	11
NMFJ and its updates					
3-14 beds	2,346	8	3	1	4
15-24 beds	1,087	10	3	1	6
25-49 beds	1,185	15	2	1	12
50-99 beds	1,029	35	1	3	31
100-199	727	39	2	3	34
200-299 beds	132	14	1	1	12
300-599 beds	19	5	1	0	4
600 or more beds	7	4	0	0	4
Unknown beds	627	20	10	1	9

maintained a bed for them.

The sampling frame for discharges consists of all events in which a person was discharged alive or dead during the 12 months ending on the day prior to the facility's survey date. Individuals were listed with every discharge they experienced from the facility in that 12 months. Hence, individuals would be listed more than once in the frame if they had more than one discharge from a sample home during the period. Current residents were also included in the discharge frame if they experienced a discharge during the period.

The sampling frame for RN's includes all RN's who were employed by or scheduled to work at the facility on the day of the survey in the facility. If convenient the RN's were listed by mode of employment (hired by the facility, including those on leave; special contractual arrangement; a temporary service arrangement).

Upon confirming the accuracy of the total count of eligible current residents (discharges, RN's) in each sampling frame, the interviewer determined which residents (discharges, RN's) to select for the sample by consulting a table containing sets of sample line numbers in the frame for each possible total count of residents (discharges, RN's). The residents (discharges, RN's) listed on the sampled lines were selected to the NNHS sample. The sets of line numbers in the sampling tables indicated for each facility, five current residents, six discharges, and four RN's; in homes with fewer than these counts all frame units were selected for the respective sample. Ten sets of tables for selecting the within facility samples were used with the ten sets being assigned to facilities on the basis of terminal digits in the identification numbers given sequentially to facilities for field work purposes. The sample line numbers in the tables were selected using systematic random sampling. Table B shows an example of these tables.

The 1985 second stage sampling technique assured the maximum sample from each sampling frame in each facility while satisfying the survey restrictions on sample sizes within the facility. The technique, thus, also evened out the work load across facilities as compared to the tech-

nique used in the 1977 cycle. In the 1977 cycle, interviewers used random starts and sampling intervals which were predesignated without regard to the actual counts of residents (discharges, staff members) in the facility. In 1977 sample sizes ranged from one up to the allowed maximums of eight each for residents and discharges and up to 23 employees in individual facilities. The sampling tables also relieved the interviewers of having to manually compute the line numbers for the sample cases.

The probability of selection at the second stage does vary inversely with the number of those units within each facility. However for residents, the overall probabilities of selection will still be generally the same across facilities within each primary sampling stratum, since the number of residents tends to be correlated to the number of beds and since facilities were selected with probability proportional to the facility bed size. As in past cycles, variations in overall probability may occur when facility bed size at the time of the survey differs from that recorded in the sampling frame. The sampled discharges and RN's will have unequal probabilities between facilities, as in prior cycles of NNHS, since numbers of discharges and RN's are more independent of the bed size.

#### 5. Data Collection Procedures

The 1985 NNHS used nine instruments to collect data from survey participants. The Facility Questionnaire was completed by personal interview with the facility's administrator or his designee. The interviewer sought authorization from that person for financial data and, if authorized, gave the Expense Questionnaire to the facility's administrator, bookkeeper, or accountant for self completion. Financial statements were accepted in lieu of a completed Expense Questionnaire when the statements contained the requested data. The interviewer(s) then completed the three second stage Sampling Lists (one each for RN's, current residents, and discharges) and selected the respective samples. A Registered Nurse Questionnaire was then left for each sample RN to complete, seal in a postage paid envelope, and either return to the interviewer or mail to the survey contractor. The interviewer attempted to complete a Current Resident Questionnaire for each sample resident and a Discharged Resident Questionnaire for each sample discharge by interviewing a member of the facility staff who referred to the sample individual's medical and financial records. For each current resident, the targeted respondent was the staff member most familiar with care provided to that resident. In facilities with 250 or more beds, two or three interviewers usually conducted the survey to reduce the length of time required in the facility.

Interviewers telephoned facilities about delinquent but authorized Expense Questionnaires one week after the interview. Except for firm refusals, the contractor's home office followed up all subsequent non-response to the Expense Questionnaires by mail four weeks after the interview. Non-response follow-up to the RN Questionnaire consisted of sending a reminder letter and a duplicate questionnaire four weeks after the interview.

TABLE B: Sampled Line Numbers Determined by the Total Count of Residents

Total Count	Sampled Line Numbers				
1	1				
2	1	2			
3	1	2	3		
4	1	2	3	4	
5	1	2	3	4	5
6	1	3	4	5	6
7	1	3	4	5	7
8	1	2	4	5	7
9	1	3	4	6	7
10	2	4	6	8	10

If the total count exceeds 1300, call headquarters for the sample lines.

Computer assisted telephone interviewing (CATI) was used to collect data for the Next-of-Kin Questionnaire from community-based persons or other facilities identified by the facility staff as knowledgeable about the sampled residents. Up to six community contacts were obtained from the facility staff for each sampled resident (current or discharged). The community contacts may be relatives, guardians, facilities to which the residents were discharged, the discharged residents themselves if they were living in the community at the time, or anyone else familiar with the sampled person. The community contacts with phones were prioritized on the basis of the facility staff's opinion about who would be best informed about the sampled resident. Community contacts without phones were not eligible for the survey. Attempts were then made to contact the potential respondents in the order of assigned priority. Each source designated as a primary (highest priority) contact was mailed an introductory letter one week before attempts were made to call that contact. The interviewer attempted up to eight calls to the primary contact if that was the only contact identified by the facility. Otherwise the interviewer attempted up to four calls to the primary contact before attempting to reach an alternate contact.

The next-of-kin component is new to the 1985 survey. The remaining components were included in the prior NNHS cycle.

#### 6. Estimates

Estimates for numbers of facilities and for other facility characteristics not correlated to bed size will be formed by summing data weighted by the reciprocals of the facility selection probabilities adjusted for non-responding facilities within bed size, certification status, and metropolitan status. Estimates for facility characteristics correlated to bed size, and for all resident, discharge, and RN characteristics will also include post-stratification ratio adjustments to sampling frame totals for beds. Estimates for current resident, discharge, and RN characteristics based on the second stage samples will be produced by multiplying the adjusted facility sampling weight by the inverse of the probability of selection within the facility and by an adjustment for questionnaire non-response unique to the facility.

Missing and inconsistent items in completed questionnaires will be imputed using data from a randomly picked similar respondent. A similar respondent facility is defined to be a facility in the same sampling stratum and a similar respondent resident, discharge, or RN is defined to be one from the same sex and age group. Processing errors were kept to a minimum by independently checking 100 percent of the work done on batches of 10 questionnaires each for each editor, coder, and keyer until the person attained a specified error rate with 10 percent samples of questionnaire items checked otherwise. Keying was independently rekeyed twice. Batches failing the quality control check were redone entirely.

Estimates for annual numbers of persons admitted and for annual numbers of admissions to nursing and related care homes will be formed by summing the weighted data from the current resident and discharge samples. A sampled resident or

discharge will have a non-zero contribution to these estimates only if the admission to the facility for the current stay or the stay ending in the sampled discharge occurred in the 12 months ending on the day before the survey date at the facility where the resident or discharge was selected. To estimate numbers of persons admitted, that admission date must also be the first to any nursing or related care home during the 12 months for the sampled individual (1).

#### 7. Variance Estimates

The sampling variances for individual estimates for 1985 NNHS will be approximated using the balanced repeated replication (BRR) procedure as opposed to the SESUDAAN (8) computer software used for the 1977 NNHS and the exact formula method used for the 1973 NNHS. The BRR was chosen since it both requires less resources than use of the exact formula methods and correctly accommodates ratio adjustments such as those in the NNHS estimators whereas, as yet, the SESUDAAN software does not. Also the BRR computes sampling variances that reflect variation in sample size due to ineligibility and non-response for sample units. The SESUDAAN was used in 1977 only when it was realized that the usual methods of producing replicate samples [i.e., one of two primary sampling units (PSU's) from each sampling stratum] yielded only one degree of freedom in variances of statistics for bed size classes coinciding with sampling strata and there was insufficient time to design adjustments to the BRR procedures.

For use of the BRR in the 1985 NNHS, the entire sample of facilities (including out-of-scope and non-respondent facilities) was assigned to 20 replicate samples selected from 20 pseudo strata. To assure all variances in the 1985 NNHS have sufficient degrees of freedom, the sampling strata defined by frame bed sizes were first collapsed to form 17 pseudo strata. These pseudo strata were formed by first arraying the facilities by the four bed size classes (3-49 beds, 50-99 beds, 100-199 beds, 200 or more beds) for which estimates are produced in the survey. Within these bed size classes, the sample homes were arrayed by certification status and MSA status (primary and secondary stratifying variables in the sampling frame) and then in the order of selection to the NNHS sample. The facilities in the array were then assigned on a rotation basis to the 17 pseudo strata with the first facility being randomly assigned. This construction of these pseudo strata allows at least 16 degrees of freedom for variances of statistics for each of the four estimation bed size classes.

The remaining pseudo strata combined sampling strata except for facilities selected with certainty. The certainty facilities were placed in a separate pseudo stratum, since these homes contribute to estimates but not to first stage variances. The 20 pseudo strata for the BRR variances are then:

1. Homes selected to the NNHS with certainty,
2. Non-certainty sample homes from the Complement Survey,
3. Sample homes without recorded bed sizes in the sampling frame, and

4-20. Non-certainty sample homes from the remaining sampling strata.

Except for the pseudo strata formation, the usual procedures for the BRR technique were followed. That is, two pseudo primary sampling units (PSU's) were formed within each pseudo stratum. The method for defining the PSU's differs with stage for the statistics. For the first stage estimates (based on the facility sample), facilities in each pseudo stratum were arrayed in the order of selection to the NNHS sample and then assigned to PSU's on a ABAB... basis with the exception of facilities in the first pseudo stratum. Homes in that stratum were assigned to both PSU's in the stratum, since facilities in that stratum contribute to the estimates but not to sampling variances.

For second stage statistics (based on resident or RN samples), each second stage sample unit in non-certainty pseudo strata was assigned to the same pseudo stratum and PSU that was assigned to the corresponding sample home. In the certainty pseudo stratum the second stage units were first arrayed by facility and then in order of selection within facility. The units were then assigned to PSU's on a ABAB... basis.

The 20 replicate samples then consist of one pseudo PSU from each of the 20 pseudo strata selected by use of an orthogonal design (6, 7). The estimate  $x'_k$  based on the k-th replicate half sample is computed in the same way as the estimate  $x'$  based on the full sample. Then the variance of  $x'$  is approximated by

$$S^2(x') = \frac{1}{20} \sum_{k=1}^{20} (x'_k - x')^2 .$$

Sampling variances for aggregates, means, ratios, and other statistics can be computed by substituting those statistics for  $x'$  and  $x'_k$ .

Due to the large number of statistics from the survey, it is impractical to compute variances for every statistic. Hence, a generalized variance function will be produced for each class of aggregate statistic by fitting curves to points whose coordinates are survey estimates  $x'$  and their corresponding estimated relvariances [ $= S^2(x')/(x')^2$ ]. For each class of statistics, a sample of 100 points will be selected from tables planned for publication. After eliminating from the tables duplicate statistics (statistics with the same value and the same standard error) and statistics based on fewer than 30 observations, the remaining statistics will be arrayed in order of magnitude. The points corresponding to the ten largest values in the array will then be selected with certainty and 90 additional points will be selected using systematic random sampling from the remaining array.

A curve of the form

$$\text{Relvariance } (x') = A + B/x'$$

will then be fit to those points by using a

weighted least squares approach. The fitted curve and the 100 points will be plotted to test for adequacy of fit and to see whether one curve may be used for two or more classes of statistics. These curves and their derivatives will be used to present the sampling errors for the 1985 NNHS statistics.

#### REFERENCES

1. Botman, Steven (1986). "Estimating From a Facility Survey the Annual Number of Individuals Who Were Admitted to and Discharged from a Nursing Home," Presented at 1986 annual meeting of the American Statistical Association.
2. Nabisco Brands, Inc. (July 1983). "Description and Technical Documentation of the LOKATE™ Geo-File." Distribution by Design, 25 The Loch, Roslyn, NY 11576.
3. National Center for Health Statistics (Feb. 1965). Development and Maintenance of a National Inventory of Hospitals and Institutions. Vital and Health Statistics. Series 1, No. 3. PHS Pub. No. 1000. Public Health Service. Washington. U.S. Government Printing Office.
4. National Center for Health Statistics (Jan. 1971). Design and Methodology of the 1967 Master Facility Inventory Survey. Vital and Health Statistics. Series 1, No. 9. PHS Pub. No. 1000. Public Health Service. Washington. U.S. Government Printing Office.
5. National Center for Health Statistics (Apr. 1968). The Agency Reporting System for Maintaining the National Inventory of Hospitals and Institutions. Vital and Health Statistics. Series 1, No.6. PHS Pub. No. 1000. Public Health Service. Washington. U.S. Government Printing Office.
6. National Center for Health Statistics, P. J. McCarthy (Apr. 1966). Pseudoreplication, an approach to the analysis of data from complex survey. Vital and Health Statistics. Series 2, No. 14. PHS Pub. No. 1000. Public Health Service. Washington. U.S. Government Printing Office.
7. Plackett, R. L. and Burman, J. P. (1943-46). "The Design of Optimum Multifactorial Experiments." Biometrika 33. pp. 305-325.
8. Shah, B. V. (April 1981). "SESUDAAN: Standard Errors Program for Computing of Standardized Rates from Sample Survey Data," RTI/5250/00-01S, Research Triangle Institute, Research Triangle Park, N.C.
9. Shimizu, Iris M. (1983). "Identifying and Obtaining the Yellow Pages for a National Area Sample." American Statistical Association, 1983 Proceedings of the Survey Research Section. pp. 558-562.
10. (April 1985). "Forming Survey Frames from the Yellow Pages in a National Area Sample of Businesses." Unpublished document. National Center for Health Statistics, Hyattsville, MD. 14 pp.