

## **2004 National Nursing Home Survey and National Nursing Assistant Survey**

Link to each section in the document:

	Page
1. Sample Design	
▪ National Nursing Home Survey (NNHS)	2
▪ National Nursing Assistant Survey (NNAS)	2
2. Sampling Frame	3
3. Scope of Surveys	3
4. Data Collection and Procedures	
▪ National Nursing Home Survey (NNHS)	4
▪ National Nursing Assistant Survey (NNAS)	5
5. Reliability of Estimates	6
6. Presentation of Estimates	9

(Text begins on next page with Sample Design.)

## **Sample Design for the 2004 National Nursing Home Survey (NNHS)**

The sampling was a stratified two-stage probability design. The first stage was the selection of facilities and the second stage was the selection of residents. The primary sampling strata of facilities were defined by sampling bed size category and metropolitan area status. Within primary strata, facilities were sorted by the following factors: certification status, hospital-based and nonhospital-based; ownership; geographic region; State, county, and zip code. Nursing homes were then selected using systematic sampling with probability proportional to their bed size. For the 2004 NNHS, 1,500 nursing homes were selected.

The number of nursing homes estimated by the survey is less than the universe figure (16,600) for several reasons. Some facilities went out of business or became ineligible for the scope of the survey between the time the universe figure was frozen and the survey was conducted. A facility was considered out of scope if it did not provide nursing care services.

The second-stage sampling of current residents was carried out by the interviewers at the time of their visits to the facilities. The sample frame for current residents was the total number of residents on the register of the facility as of midnight of the day prior to the day of the survey. Residents who were physically absent from the facility due to overnight leave or a hospital visit, but had a bed maintained for them at the facility, were included in the sample frame. A sample of up to twelve current residents per facility was selected.

## **Sample Design for the 2004 National Nursing Assistant Survey (NNAS)**

The sample design for the NNAS is a stratified, multistage probability design. The first stage consists of the selection of a stratified probability sample of the nursing facilities in the sampling frame. Each facility was placed into a stratum comprised of sampling bed-size category and Core Based Statistical Area (CBSA) status. To permit implicit stratification within these broader sampling strata, nursing facilities were arrayed by certification status, hospital based and non-hospital based status, ownership, geographic region, State, county, and zip code. Facilities were then selected using systematic sampling with probability proportional to their bed size. For the 2004 NNHS, 1,500 nursing facilities were selected from the universe consisting of about 16,600 nursing homes in the United States.

Next, a sub-sample (n=790) of these selected nursing homes was then selected to participate in the NNAS. A sample of nursing assistants, stratified by tenure in the facility, was then selected from each of the participating and eligible facilities in the sub-sample. Nursing assistants were selected from a list the facility provided of the NAs employed by the facility as of midnight the day before the interview. The NAs were divided into two tenure groups: (1) employed by the facility for less than one year and (2) employed by facility for one year or more. A computer-assisted personal interviewing (CAPI) system program randomly selected up to four nursing assistants from each tenure group. A total sample of 4,542 nursing assistants was selected for the NNAS; 4,274 were eligible and 3,017 participated in the survey. Interviewers conducted telephone interviews with those nursing assistants who they were able to contact and found willing to participate in the survey.

## **Sampling Frame**

For the 2004 National Nursing Home Survey, 1,500 nursing home facilities were selected from a sampling frame of U.S. nursing homes. The sampling frame was drawn from two sources: (1) the Centers for Medicare and Medicaid Services Provider of Services file of U.S. nursing homes, and (2) State licensing lists compiled by a private organization. These two files contained approximately 17,000 nursing homes. The combined files were matched and unduplicated, resulting in a sampling frame of 16,628 nursing homes.

## **Scope of Surveys**

For the 2004 National Nursing Home Survey (NNHS), a sample of 1,500 nursing homes was selected. Of these, 283 refused to participate and 43 were considered out of scope for one or more of the following reasons: the nursing home had gone out of business, it failed to meet the definition used in this survey, or it was a duplicate of another facility in the sample. A total of 1,174 nursing homes participated at the first stage by providing facility information, resulting in a first stage response rate of 81 percent. A total of 14,017 residents were sampled from the responding facilities. Of these, 8 were out of scope and 502 refused, yielding a second stage response rate of 96 percent, and an overall response rate for the resident component of the NNHS of 78 percent.

For the 2004 National Nursing Assistant Survey, a subsample of 790 facilities was selected from the NNHS facility sample. Of these, 21 facilities were determined to be out of scope. Of the 769 eligible facilities, 164 did not participate in any aspect of the NNHS and 23 others elected not to participate in the NNAS portion of the survey, for a first stage response rate of 76 percent. Of the 582 eligible facilities that agreed to participate in the NNAS, 4,542 nursing assistants were selected to participate in the NNAS. NAs were considered eligible to participate in the survey if they 1) provided assistance with activities of daily living (ADLs); 2) were paid to provide those services; 3) were certified (or in the process of certification) to provide Medicare/Medicaid reimbursable services; 4) worked at least 16 hours per week; and 5) were employees of the nursing home and not contract employees. Among the 4,542 NAs selected, 4,274 were eligible to participate and 3,017 completed an interview, yielding a second stage response rate of 71 percent and an overall NNAS response rate of 53 percent.

## **2004 National Nursing Home Survey (NNHS) Data Collection Procedures**

The 2004 NNHS was administered in sampled nursing home facilities using a computer-assisted personal interviewing (CAPI) system that was loaded on the interviewers' laptops. A self-administered staffing questionnaire was mailed to the nursing home administrator, along with the appointment confirmation letter, to be completed by the time of the facility interview. The Facility Component of the CAPI included facility qualifications and facility characteristics data items. The interviewers were instructed to complete the facility qualifications data items first to ensure that the nursing home was eligible to participate in the 2004 NNHS. The interviewers were then free to administer the facility characteristics and Resident components in any order based upon the availability of the facility respondents. The resident data items were organized into four modules: Health Status (HA), Non-Minimum Data Set (HN), Prescribed Medications (PM), and Payment Sources (PA). The Health Status module collected data about a resident's health status as documented in the Minimum Data Set (MDS); the MDS is a federally mandated clinical assessment of all residents in a Medicare or Medicaid certified nursing home. MDS information is transmitted electronically by nursing homes to the MDS database in their respective states; the MDS information from each state database is captured into the national MDS database at CMS. The Non-MDS module collected data about a resident's health status and medical care that were not available from the MDS. The NNHS also included a first-time supplemental survey of nursing assistants employed by nursing homes.

Data were collected according to the following procedures: (1) An advance package of information, including a letter, was mailed to the administrator of sample facilities, informing them of the survey and the fact that they would be contacted for an appointment. Endorsement letters from the American College of Health Care Administrators, American Association of Homes and Services for the Aging, and the American Health Care Association were sent with the advance package. Also included in this package was one of the reports from the 1999 survey to illustrate how the data would be displayed; (2) After the mailing of the packages, the interviewer telephoned the sample facility and made an appointment with the administrator; (3) After the interviewer successfully scheduled an appointment with the administrator, a confirmation package was mailed to provide details about the interview. This package included a confirmation letter with details about what the interviewer would need for the interview, in addition to the self-administered staffing questionnaire that the administrator was expected to complete by the time of the in-person facility interview; (4) At the in-person interview with the administrator, the interviewer collected the completed staffing questionnaire and administered the Facility component of the CAPI. Provided the facility was eligible to participate in the survey, the interviewer then sampled up to 12 current nursing home residents using the Sampling module of the CAPI; (5) The interviewer then interviewed designated staff, familiar with the residents and their care, to answer the data items in the Resident component. The respondents were asked to use the residents' medical records to answer the data items. No resident was interviewed directly; and (6) If the facility was also selected to participate in the supplemental National Nursing Assistant Survey (about half of the nursing homes were selected to participate in the NNAS), then the interviewer constructed a sampling list of currently employed nursing assistants, selecting up to 8 nursing assistants, and requested contact information for each sampled nursing assistant.

After the data were collected, they were transmitted from the interviewer's laptop to the contractor's office. Extensive data checking, editing and coding were then performed to ensure that all responses were accurate, consistent, logical, and complete.

## **2004 National Nursing Assistant Survey (NNAS) Data Collection Procedures**

The 2004 NNAS was administered by telephone using a computer-assisted telephone interviewing (CATI) system. The questionnaire included 11 modules, the first of which was a screening section to determine eligibility. In addition to the screening module, the questionnaire included modules on recruitment; education, training and licensure; job history; family life; management and supervision; client relations; organizational commitment and job satisfaction; workplace environment; work-related injuries; and socio-demographics. An additional module was only asked of nursing assistants who were sampled for the survey but were no longer working at the facility at the time of the NNAS survey.

Data were collected according to the following procedures:

(1) Initial contacts with the nursing facilities were as described for the National Nursing Home Survey; (2) After the appointments were scheduled and confirmed, the appointment confirmation letter in the packet sent to the facilities chosen to participate in the NNAS included additional information unique to the NNAS. The letter explained that up to eight nursing assistants (NAs) would be selected from the facility to participate in the survey. Administrators were requested to provide a list of nursing assistants employed by the facility, indicating their tenure—less than one year or one or more years—and contact information for selected nursing assistants. The confirmation packet also included letters from three professional NA organizations endorsing the NNAS (National Association of Geriatric Nursing Assistants, the National Network of Career Nursing Assistants, and the Paraprofessional Healthcare Institute), two copies of a NNAS flyer for administrators to post in a common area, and an advance letter about the survey for the administrator to present to all nursing assistants employed by the facility; (3) At the time of the in-person interview with the facility administrator, the facility provided a list, for sampling purposes, of nursing assistants employed by the facility as of midnight the day before the interview. The interviewer cleaned and numbered the list so that the nursing assistants were divided into two tenure groups: those employed by the facility for less than one year and those employed by the facility for one or more years. A computer-assisted personal interviewing system (CAPI) program randomly selected up to four nursing assistants from each tenure group. After the eight nursing assistants were selected, the NNHS field interviewers recorded contact information obtained from the nursing home for each of the selected nursing assistants; (4) Information packets were left at the facility to be distributed to the selected NAs. The packets included letters from the director of NCHS and the NCHS project officer, a survey fact sheet, a DVD about the survey, a pen, a \$5 bill clipped to the informed consent letter, and a return postcard (and postage paid return envelope) for NAs to indicate their willingness to participate in the survey and their contact information. These materials also included a toll-free number allowing NAs to contact Westat, who conducted the telephone interviews, if they were interested in participating in the survey; and (5) The Telephone Research Center (TRC) staff used the contact information provided by the facility to contact the NAs and to solicit their participation in the survey.

Data were entered by TRC staff directly into a computer-assisted telephone interview (CATI) instrument. After data were extracted from the instrument database, extensive data checking, editing and coding were performed to ensure that the responses were accurate, consistent, logical, and complete.

## Reliability of Estimates

Because the data presented on this website are based on a sample, they will differ somewhat from data that would have been obtained if a complete census had been taken, and the same instructions and procedures used. The standard error (SE) is primarily a measure of the variability that occurs by chance because only a sample, rather than the entire universe, is surveyed. The standard error also reflects part of the measurement error, but it does not measure any systematic biases in the data or other nonsampling error. The chances are about 95 in 100 that an estimate from the sample differs from the value that would be obtained from a complete census by less than twice the standard error.

The standard errors in all tables presented here were approximated with SUDAAN software. SUDAAN computes standard errors by using a first-order Taylor approximation of the deviation of estimates from their expected values. It should be noted that some estimates are presented but cannot be assumed reliable, and are flagged with an asterisk. Estimates are flagged if they are based on between 30 and 59 cases; or if they are based on more than 59 cases but have a relative standard error exceeding 30 percent. NCHS does not present any estimate based on fewer than 30 cases.

Standard errors can be calculated for facility, resident and nursing assistant estimates using any statistical software package as long as clustering within facilities and other aspects of the complex design are taken into account. Software products such as SAS, Stata, and SPSS all have these capabilities.

All three of the NNHS public use files (the facility, resident and nursing assistant files) include design variables that designate each record's stratum marker and the first-stage unit (or cluster) to which the record belongs. Examples follow for using these design variables with SUDAAN, STATA, and SAS survey procedures. Examples for using these design variables with SPSS are being tested and will be included in an update.

Table I. Computations using SUDAAN

<b>FILE</b>	<b>PROC STATEMENT</b>	<b>NEST STATEMENT</b>	<b>TOTCNT STATEMENT</b>	<b>WEIGHT STATEMENT</b>
<b>FACILITY FILE</b>	PROC x FILE=y DESIGN=WOR;	NEST STRATA;	TOTCNT POPFAC;	WEIGHT SBEDWT or SFACWT;
<b>RESIDENT FILE</b>	PROC x FILE=y DESIGN=WOR;	NEST RSTRATA FACNUM / MISSUNIT;	TOTCNT POPFAC NPOPRES;	WEIGHT SAMWT;
<b>NURSING ASST. FILE</b>	PROC x FILE=y DESIGN=WOR;	NEST STRATUM NAFACNUM NASSTRAT / MISSUNIT;	TOTCNT POPFAC _ZERO_ NPOPNAS;	WEIGHT SAMWT;

2004 National Nursing Home Survey and National Nursing Assistant Survey

Table II. Computations using STATA

FILE	DESIGN DESCRIPTION IN STATA
<b>FACILITY FILE</b>	svyset[pw=SFACWT], strata(STRATA) fpc(POPFAC) or svyset[pw=SBEDWT], strata(STRATA) fpc(POPFAC)
<b>RESIDENT FILE</b>	svyset[pw=SAMWT], strata(RSTRATA) psu(FACNUM) fpc(POPFAC)    _n
<b>NURSING ASST. FILE</b>	svyset[pw=SAMWT], strata(STRATUM) psu(NAFACNUM) fpc(POPFAC)    _n

NOTE on RESIDENT FILE: Prior to STATA runs, three facilities which have only one resident record (singletons) each must be resolved. Moving each of the 3 single resident records into another FACNUM in the same RSTRATA is suggested.

NOTE on NURSING ASST. FILE: Prior to STATA runs, 13 nursing assistants that are singletons in their NAFACNUM must be resolved. Moving each of the 13 single nursing assistant records into another NAFACNUM in the same STRATUM is suggested.

Table III. Computations using SAS survey procedures

STATEMENT	PROC	STRATA	CLUSTER	WEIGHT
<b>FACILITY FILE</b>	PROC SURVEY__ DATA = Y TOTAL= SECONDFILE;	STRATA STRATA;	CLUSTER CASENUM;	WEIGHT SBEDWT OR SFACWT;
<b>RESIDENT FILE</b>	PROC SURVEY__ DATA = Y;	STRATA RSTRATA;	CLUSTER FACNUM;	WEIGHT SAMWT;
<b>NURSING ASST. FILE</b>	PROC SURVEY__ DATA = Y;	STRATA STRATUM;	CLUSTER NAFACNUM;	WEIGHT SAMWT;

NOTE ON FACILITY FILE: Use of the finite population correction is recommended for analysis of the Facility File only. Prior to runs on the Facility File, create a two variable data set named SECONDFILE containing one record for each unique value of STRATA, and the corresponding value of \_TOTAL\_ (\_TOTAL\_ = POPFAC).

NOTE ON ALL FILES: Blank values for variables in any of the files may understate variances, and results should be interpreted accordingly.

In addition to using software to account for the complex design, the user may approximate standard errors by generalized variance curves, although this method is far less accurate. The curve parameters, A and B, as well as the lowest reliable estimates for various types of estimates are presented in Table IV below. Standard errors for aggregate estimates may be approximated by using the appropriate parameter(s) and the general formula:

$$SE(X) = \frac{X \cdot RSE(X)}{100}$$

where X is the estimate (in thousands) and RSE(X) is the relative standard error (RSE) of the estimate as a percent. The relative standard error for X is the SE for the estimate, X, divided by X, times 100. Throughout this documentation, relative standard errors are taken as percents, between 1 and 100. The relative standard error of an estimate may be approximated using the following general formula:

$$RSE(X) = 100 \cdot \sqrt{A + \frac{B}{X}}$$

where X is the estimate (in thousands) and A and B are the appropriate parameters, as shown in Table IV.

A different formula is needed to approximate the reliability of percent estimates, expressed as a proportion p (where 0 < p < 1). The relative standard error and the standard error of percents are designated as RSE(p) and SE(p) respectively. The appropriate value of the same parameter B as given in Table IV is used in the following equations:

$$RSE(p) = 100 \cdot \sqrt{\frac{B \cdot (1 - p)}{p \cdot Y}}$$

and

$$SE(p) = \frac{p \cdot RSE(p)}{100}$$

where p = X/Y, X = the numerator of the estimated percent, and Y = the denominator of the estimated percent, both in thousands.



Table IV. Parameters used to compute relative standard errors by type of estimate

Type of estimate	A	B	Least Reliable Estimate (in 1000's)
Facility estimate correlated with bedsize	-.00005	.01277	.141
Facility estimate uncorrelated with bedsize	.00169	.00954	.108
Current resident	.00806	.06541	.798
Nursing assistant	.00784	.33170	4.037

The approximations of the relative standard error and the standard error of a percent are valid only when one of the following conditions is satisfied: the relative standard error of the denominator is 5 percent or less, or the relative standard errors of the numerator and the denominator are both 10 percent or less.

### **Presentation of Estimates**

Presentation of estimates for the NNHS is based on the relative standard error of the estimate and the number of sample records on which the estimate is based (referred to as the sample size). Estimates are not presented in NCHS reports unless a reasonable assumption regarding the probability distribution of the sampling error is possible.

Based on consideration of the complex sample design of the NNHS, the following guidelines are used for presenting the NNHS estimates:

If the sample size is less than 30, the value of the estimate is not reported.

If the sample size is 30-59 or if the sample is 60 or more and the RSE is more than 30 percent, the estimate is reported, but should not be assumed reliable. This is indicated by an asterisk (\*) in the tables.

If the sample size is 60 or more and the relative standard error is less than or equal to 30 percent, the estimate is reported and is considered reliable.