Appendix B

Survey Methodology

Five national sample surveys of registered nurses (RNs) have been carried out: September 1977; November 1980; November 1984; March 1988; and March 1992. These studies all utilized a design initially developed in 1975-76 under a contract the Division of Nursing, Bureau of Health Professions, Health Resources and Services Administration, had with Westat, Inc. In designing the approach, Westat, Inc., took into account the lack of single listing of all individuals who have licenses to practice in the United States and the fact that nurses may be licensed in more than one State at a time. The description of the survey methodology included here has been abstracted from a technical report of the adaption of the survey design to this current study prepared by Research Triangle Institute (RTI), the contractor for the 1992 study.

Sample Design

Because State-level estimates were desired, different sampling rates were set for the States based on considerations of the statistical precision of the estimates and the costs involved. States that had smaller numbers of RNs currently licensed were assigned higher sampling rates than were larger States, to yield a sample large enough to provide State estimates of reasonable precision. It was deemed too expensive to set sampling rates that would yield estimates of equal precision for all States while also achieving good precision for national estimates. Sample sizes and the precision of sample estimates thus typically were lower for smaller States than they were for larger States.

The use of differential State sampling rates substantially reduced variations in State sample sizes and thus permitted more precise State-level estimation than if a uniform sampling rate had been used. For example, States such as Wyoming and Alaska would have had samples of fewer than 200 RNs if a uniform sampling rate (near the overall sampling rate) had been employed across all States. The cost of this disproportionate sampling (i.e., using higher sampling rates in smaller States) was reduced precision of national estimates due to unequal weighting effects.

When the survey was originally designed, the State Boards of Nursing in the 50 States and in the District of Columbia (hereafter also referred to as a State) were determined to be the only sources for a sampling frame for the RN survey. Each

State's Board of Nursing maintains a file containing the name, address, and license number of every RN who is currently licensed in that State. These 51 files formed the basis of the sampling frame.

Approximately equal-sized alpha-segments were constructed by partitioning an alphabetically ordered list of all RN names nationwide into 250 segments with equal (or as nearly equal as possible) numbers of RNs. An alpha-segment consisted of all alphabetically adjacent names falling between set boundaries.

Registered nurses were in the sample on the basis of name, with an RN falling into the sample if her/his name of licensure fell within a specific portion of the alpha-segments included in the sample from her/his State. Specifically, the lower boundary of an alpha-segment was the name lowest in alphabetical order of all names included in that segment. The membership of the segment consisted of all names, beginning with the lower boundary, <u>up to but not including</u> a name that defined the upper boundary. (The latter name fell into the next alpha-segment.)

There was a planned variation in the size of the portions of segments used to accommodate the differing State sampling rates. The largest portions used were full alpha-segments, while other sizes were 1/2-, 1/4-, 1/8-, 1/16-, and 1/32-portions. The fractions indicated the size of the specified alpha-segment portion relative to the size of the basic alpha-segment. The sampling rate required for a given State was achieved using a combination of these portions of alpha-segments. Each State's sample consisted of 40 primary sampling units (portions of alpha-segments). Although each State had the same number (40) of sample segments (i.e., portions of alpha-segments), the segments differed in size depending on the State's sampling rate.

The fact that some RNs maintain active licenses to practice in more than one State complicated how selection probabilities were determined. An RN is represented on the licensure file of each State in which she/he has an active license at the time a sample is selected. To identify and account for such multiple licenses, the alpha-segment portions associated with larger States were "nested" (or included) within those associated with smaller States. Under this scheme, an RN who was licensed under the same name in two States with identical sampling rates was selected (or not selected) for both States, because the alpha-segments and portions of alpha-segments that defined sample membership were identical for both States. However, for two States that were sampled at different rates, the alpha-segment portions for the lower sampling rate (the larger State) were nested within those of the higher sampling rate (the smaller State).

The nesting was based on how the 40 basic alpha-segment selections were used to define the sample for each State. Each of these alpha-segments, or one of the fractional portions of it, constituted one of the 40 sample clusters for each State. Accordingly, each of the basic alpha-segments had associated with it a 1/2-portion selection, a 1/4-portion selection from the selected 1/2-portion, a 1/8-portion selection from the selected 1/4-portion, a 1/16-portion selection from the selected 1/8-portion, and a 1/32-portion selected from the 1/16 portion. The sample 1/2-portion for a particular alpha-segment was selected at random and remained in the sample for all States for which half-portions were sampled from this alpha-segment. Similarly, each of the smaller sample portions was randomly selected from the halves of the next largest sample portion; e.g., a sample 1/8-portion was a randomly selected half of the corresponding sample 1/4-portion.

The sampling rate for a particular State was obtained from some combination of the alpha-segments and portions. For example, the 40 alpha-segments constituted the sample for States with a 16-percent sampling rate. (Because each segment contained an expected 0.4 percent of the State's RN names, taken together they contained an expected 40 x 0.4 percent, or 16 percent of those names.) The sample for a State with an 8-percent sampling rate consisted of the 40 1/2-portion selections. A 5-percent sampling rate was achieved by first randomly dividing the 40 alpha-segments into two groups, the first containing 30 alpha-segments and the other containing 10, and using the 1/4-portions from the first group and 1/2-portions from the second group $(0.4 \times [(30 \times 1/4) + (10 \times 1/2)] = 5)$.

The survey design specified, at the time of implementation, precisely which alpha-segments and portions would correspond to each of the different sampling rates used. This task resulted in the specification of 40 pairs of names for each of the sampling rates. Each pair consisted of the names defining the lower and upper boundary for one of the alpha-segments or alpha-segment portions corresponding to the sampling rate. Thus, the alpha-segment (portion) was defined by all names, beginning with its lower boundary, up to but not including its upper boundary.

To ensure that current information about RNs could be obtained, the survey design called for periodic implementation. A panel structure for the RN survey allowed for the systematic replacement of several of the sample alpha-segments in the periodic surveys. Under the design, the 40 sample alpha-segments were randomly assigned to five panels of 8 alpha-segments each. For each successive survey, a new panel (consisting of eight new alpha-segments) was entered into the sample, thus replacing one of the five panels that was in the previous survey. Under this scheme, a nurse whose name did not change could be retained in the sample for up to five surveys. The alpha-segments were reconstructed in the fourth survey (1988) and, therefore, exact correspondence of the current segments to those established initially may no longer exist.

Each of the 51 State Boards of Nursing provided one or more files that contained the names of currently licensed RNs. The files formed the basis of the sampling frame from which the RNs for that State were selected. The licensure files provided by the States were submitted on computer tape, on diskettes, or on a printed list. Essentially the same procedure was followed for sample selection for all States regardless of which form was submitted. Once a licensure file provided by a State was complete (i.e., contained all appropriate names of individuals with active RN licenses) and met all specifications, the required sample names in that file were selected.

The sample for a given State consisted of all RN names falling into any one of the State's designated 40 alphabetic portions (one portion from each of the 40 alpha-segments associated with the current sample). The sample alphabetic portions varied among States, depending on the State's sampling rate. The sample for a particular State was defined by the alphabetic portion boundaries, associated with its sampling rate. Regardless of the way a State alphabetized and standardized the names on its files, the sample names were selected according to the standards established by the survey design. That is, sample selections ignored blanks and punctuation in the last names (except a dash in hyphenated names) and ignored titles (e.g., "Sister").

Table B-1 shows the sampling rates and sample sizes that were planned and actually obtained for the 51 States in the survey. Both the sampling rate planned for each State and the actual sampling rate are shown. States are priority-ordered by frame size (smaller to larger) so that sampling rates are decreasing down the table.

The percentage difference shows the State-specific variation caused by the nurse names in each State. The percentage difference averaged less than 0.50 percent over the full sample. The State frame sizes shown in Table B-1 were adjusted to account for ineligible licenses (i.e., frame errors) found in the sample. The ineligible licenses were identified in the process of reconciling the State and nurse reported licenses.

Table B-1 State Sampling Rates and Sample Sizes (Priority-Ordered)

State	Frame[1/] Size	Percent Sampling Rate[2/] Planned	Actual	Actual Sample Size
Total	2,579,208			45,130
Wyoming	4,274	16.0	14.7	629
Alaska	5,128	12.0	12.0	616
North Dakota	6,853	9.0	8.1	557
Vermont	7,212	9.0	8.9	641
South Dakota	8,306	8.0	7.8	644
Idaho	8,705	8.0	8.3	723
Nevada	9,139	8.0	7.6	695
Montana	9,154	8.0	7.8	709
Delaware	9,528	7.0	7.4	709
Hawaii	10,986	7.0	7.0	768
New Mexico	11,250	6.0	5.8	657
Utah	12,458	6.0	6.5	805
Maine	14,472	5.0	5.3	773
Rhode Island	14,544	5.0	5.4	792
New Hampshire	15,556	5.0	5.2	816
Nebraska	17,095	4.0	4.0	688
District of Columbia	17,173	4.0	3.9	673
Mississippi	18,075	4.0	4.1	749
West Virginia	18,643	4.0	3.9	725
Arkansas	20,841	3.5	3.5	736
Oklahoma	23,748	3.5	3.5	841
Kansas	25,457	3.0	3.2	808
South Carolina	27,098	3.0	2.9	787
Oregon	30,846	2.5	2.3	706
Kentucky	31,048	2.5	2.5	774
Louisiana	31,416	2.5	2.4	766
Iowa	32,090	2.5	2.4	774
Alabama	34,189	2.5	2.4	834
Colorado	35,359	2.5	2.6	915
Arizona	35,835	2.5	2.3	811
Tennessee	46,480	1.8	1.9	874
Minnesota	49,096	1.8	1.7	818
Connecticut	49,318	1.8	1.7	826
Washington	49,688	1.8	1.7	831
Maryland	52,215	1.5	1.6	842
Wisconsin	55,551	1.5	1.4	758
Missouri	55,884	1.5	1.5	836
Georgia	57,514	1.5	1.8	1,041
Indiana	58,102	1.5	1.5	887
North Carolina	60,274	1.5	1.7	993
Virginia	65,480	1.25	1.2	786
Massachusetts	96,497	1.00	0.93	921
Michigan	102,554	1.00	0.93	950
New Jersey	102,236	1.00	1.00	1,036
Ohio	121,788	1.00	1.02	1,245
Illinois	123,472	1.00	1.00	1,229
Texas	124,410	1.00	0.98	1,220
Florida	127,919	1.00	0.99	1,264
Pennsylvania	183,004	0.90	0.89	1,629
California	223,463	0.90	0.86	1,927
New York	227,785	0.90	0.92	2,096

1/ Adjusted frame size

2/ Since the actual distribution of names differs for each State from the distribution derived from the merged States used for the development of the 250 alpha-segments some variation occurs between the planned and actual sampling rate.

Cases that could not be reconciled by RTI were sent to the State Boards of Nursing for resolution. Based on the States' responses, some nurses were found to have been incorrectly included in the sample frame; that is, they were not licensed to practice as an RN at the time of frame construction.

An adjustment to the State frame size was made by computing an estimated number of ineligibles in the State based on the proportion of verified ineligibles in the sample, and subtracting this estimated number from the original frame total. This adjustment results in frame sizes more closely reflecting the population numbers of eligible nurses and thus improves the accuracy of all survey estimates. Priority changed only for the States of New Jersey and Michigan and both States had the same sampling rate. Therefore, the sample alpha-segments were the same for the States involved.

Some RNs who had more than one active license were selected more than once. Steps were taken, in accordance with the sample design, to ensure that each sampled RN was retained in the national sample exactly once to avoid multiple questionnaires being sent to nurses. Specifically, after all State samples were selected, they were combined on computer into a single national sample file. This file, referred to as the master file, was sorted by last name, ZIP code, address, and first name. A complete listing of the file was printed out and reviewed visually. Special attention was paid to all groups of names that sorted together. Two names were taken to represent the same individual if entire names appeared to be the same and addresses were the same. Allowances were made for obvious unimportant differences between representations of the same name and/or address. When there was any doubt, both names were retained in the sample and questionnaires were mailed to both addresses.

Statistical Techniques

A. Weighting

The probability sample design for the Fifth National Sample Survey of Registered Nurses, (RN V) permits the computation of unbiased estimates of characteristics of the target population. These estimates are based on weights that reflect the complex design and compensate for the potential risk of nonresponse bias to the extent feasible. The weights that are assigned to each sample nurse may be interpreted as the number of nurses in the target population that the sample nurse represents. The weight for an RN is the reciprocal of the nurse's probability of selection, adjusted to account for nonresponding nurses and for multiple licenses.

A nurse is uniquely linked on the national sampling frame with his/her "priority State," i.e., the State with the lowest number of currently licensed RNs in which she or he was licensed. All nurses with the same priority State had an equal probability of being selected within that State and, consequently, all sampled nurses with that priority State had equal weights. The sum of the weights for all respondents assigned a specific priority State equals, approximately, the total number of active licenses in the State at the time the sample was drawn less the number of those licenses assigned to higher priority States.

The weights were computed sequentially for States A, B, etc., where A was the highest-priority State, and B the next highest State. The weight for State A was the ratio of the count of licenses in the sampling frame for State A to the number of respondents licensed in State A. For State B, and the remaining States, the numerator and denominator of this ratio were adjusted to account for State A and other higher-priority States. To describe the basic method, the following terms are defined:

$N(i)$ = total number of licenses for State i

$m(i)$ = number of respondents for State i that did not have a license in higher-priority State

$n(i,j)$ = number of respondents with a license in both State i and State j (note $n(i,i)$ denotes the number of eligible respondents with a license only in State i)

$W(i)$ = the adjusted weight for eligible respondents who were assigned to the priority State i.

The weight for State A was computed as follows:

$$W(A) = N(A) / m(A) .$$

For State B weight, $W(B)$, the numerator was the total frame count of RNs licensed in State B, $N(B)$, after removal of the estimated total count of State B nurses who were also licensed in State A (i.e., $W(A)n(A,B)$). Similarly, the numerator of $W(C)$ excluded State C nurses who were also licensed in either State A or State B (i.e., $W(A)n(A,C) + W(B)n(B,C)$). That is, for the State B weight and the State C weight, the computations were:

$$W(B) = [N(B) - W(A)\ n(A,B)] / m(B)$$

$$W(C) = [N(C) - W(A)\ n(A,C) - W(B)\ n(B,C)] / m(C) .$$

In either case, the denominator was the number (m(B) or m(C)) of respondents in the State not licensed in a higher-priority State.

In general, the numerator of a State I weight, W(I), was the total frame count licensed in State I after removal of the estimated total count of State I nurses also licensed in higher-priority States. The denominator, m(I), was the number of State I respondents not licensed in a higher-priority State. This weighing scheme incorporated a nonresponse adjustment that inflated the respondents' data to represent the entire universe.

B. Estimation Procedures

State-level estimates can be computed using the final set of sampling weights, W_k (for sample nurse-k). For example, an estimate of the total number of RNs working in Iowa may be based on the following indicator variable X_k:

$$X_k \begin{cases} = 1 \text{ if nurse-k worked in Iowa,} \\ = 0 \text{ otherwise.} \end{cases}$$

The desired estimated total may then be written as

$$\hat{X} = \sum_k W_k X_k ,$$

the sum being over all sample nurses.

Estimates of ratios and averages are obtained as the ratio of estimated totals.

C. Sampling Errors

To the extent that samples are sufficiently large, relatively precise estimates of characteristics of the licensed RN population of the United States can be made because of the underlying probability structure of the sample data. Such estimates are, sometimes, an imperfect approximation of the truth. Several sources of error could cause sample estimates to differ from the corresponding true population value. These sources of error are commonly classified into two major categories: sampling errors and nonsampling errors.

A probability sample such as the one used in this study is designed so that estimates of the magnitude of the sampling error can be computed from the sample data. Nonsystematic components of nonsampling error will also be reflected in the sampling error estimates.

Some sources of error--such as unusable response to vague or sensitive questions, no response from some nurses, and errors in coding, scoring, and processing the data--are, to a considerable extent, beyond the control of the sampling statistician. They are called "nonsampling errors" and also occur in cases where there is a complete enumeration of a target population, such as the U.S. Census. Among the activities that were directed at reducing nonsampling errors to the lowest level feasible for the current survey were careful planning, keeping nonresponses to the lowest feasible level, and coding and processing the sample data carefully.

Nonresponse to the survey is one source of nonsampling error because a characteristic being estimated may differ, on average, between respondents and nonrespondents. It is for this reason that considerable effort is expended to obtain a high response rate through such actions as respondent motivation and followup procedures. A high response rate will reduce both random and systematic errors.

Sample survey results are subject to sampling error. The magnitude of the sampling error for an estimate, as indicated by measures of variability such as its variance or its standard error (the square root of its variance), provides a basis for judging the precision of the sample estimates.

Systematic sampling, which was the selection procedure used in choosing the alpha-segments for this study, is very convenient from certain practical points of view, including providing for panel rotation. However, it does not permit unbiased estimation of the variability of survey estimates unless some assumptions are made. Estimates of standard errors are made based upon the assumption that RTI's systematic sample of 40 alpha-segments is equivalent to selecting a stratified random sample of two alpha-segments from each of 20 strata of adjacent alpha-segments. Ordinarily, this assumption should lead to overestimates of the sampling error for systematic sampling, but in this case (with alpha-segments as the sampling units) RTI believes the magnitude of the overestimate is trivial.

Regarding the sample as consisting of 20 pairs of alpha segments (thus obtaining 19 degrees of freedom) for the purpose of variance estimation, the probability will be approximately

0.95 that the statistic of interest differs from the value of the population characteristic that it estimates by not more than 2.093 standard deviations. Specifically, a 95 percent confidence interval for an estimated statistic $\hat{x}$ takes the form

$$\hat{x} \pm 2.093\hat{\sigma}_{\hat{x}} ,$$

where $\hat{\sigma}_{\hat{x}}$ is the estimated standard error of $\hat{x}$. Inferences involving a small number of respondents (where, for example, 25 is considered a small number) have estimated standard errors of $\hat{\sigma}_{\hat{x}}$ that themselves are subject to high variability. Thus, inferences based on such estimates should be guarded.

1. Direct Variance Estimation

The direct computation of the sampling variance used the jackknife variance estimation procedure with 20 replicates of the sample. Each replicate was based on 19 pairs of alpha-segments and one alpha-segment from the 20th pair. The actual respondent count in the included segments for a particular replicate constituted a sample of approximately 39/40ths of the full respondent sample, and were weighted to represent the full population.

Variance estimates using the jackknife approach require the computation of a set of weights for the full sample and a set for each replicate using the established weight computation procedure (i.e., 20 additional sets of weights). For the replicates, the weights were based on the responding nurses from the 39 segments associated with each replicate. The 20 sets of weights permits the construction of 20 replicate estimates to compare with the estimate produced from all of the data; each replicate estimate will be based on about 39/40ths of the data.

The variance estimate is computed using the following procedure. Define the following,

$\hat{X}_{\cdot i}$ = an estimated total for replicate i associated with alpha-segment pair i;

$\hat{X}_{i}$ = an estimated total obtained over the full sample.

The variance of $\hat{x}$, Var$(\hat{x})$, is estimated by computing

$$\mathrm{Var}(\hat{x}) = \sum_{i=1}^{20} (\hat{x}_i - \hat{x})^2 .$$

If the estimate of interest is a ratio of two estimated totals (e.g., the proportion of RNs resident in Florida between 25 and 29 years old), the variance estimate for the estimated ratio would be of the following form:

$$\mathrm{Var}\left(\frac{\hat{x}}{\hat{y}}\right) = \sum_{i=1}^{20} \left(\frac{\hat{x}_i}{\hat{y}_i} - \frac{\hat{x}}{\hat{y}}\right)^2$$

Following the example, the $\hat{x}$ and $\hat{x}_i$ measurements would be full sample and replicate estimates, respectively, of the total number of RNs resident in Florida who were 25 to 29 years old while $\hat{y}$ and $\hat{y}_i$ would be the corresponding estimates of the total number of RNs resident in Florida. The variance of any other statistic, simple or complex, can be similarly estimated by computing the statistic for each replicate.

The jackknife variance estimator can use either the full sample estimate, $\hat{x}$, or the average of the replicate estimates, x_i. While usually little difference exists between the two estimates, RTI used the estimator $\hat{x}$, which tends to provide more conservative estimates of variance.

Direct estimates of the variance were computed for a variety of variables. These variables were chosen not only due to their importance, but also to represent the range of expected design effects. The average of these design effects (on a State-by-State basis) provides the basis for the variance estimate for variables not included in the set for which direct variance estimates were computed. Direct estimates of the standard error (the square root of the variance) are presented for a selected set of national estimates in Table B-2. Table B-3 shows the estimated State population of nurses, and the standard error of these population totals.

2. Design Effects and Generalized Variances

The generalized variance is a model-based approximation to the sampling variance estimate, which is less computationally complex than the direct variance estimator but is also less accurate. The generalized variance equations use the national-level or State-level estimates of the design effect and, for some estimates, the coefficient of variation (CV) to estimate the sampling variance. The design effect, F, for an estimated proportion $\hat{p}$ is determined by taking the ratio of the estimated sampling variance, $\hat{\sigma}_p^2$, obtained by the jackknife method, to the sampling variance of the $\hat{p}$ simple random sample

Table B-2. Standard errors(S.E.) of selected statistics(number and percentage) for U.S. registered nurse population

Description	Estimated Number	S.E. of estimated Number	Estimated percent	S.E. of estimated percent
Total RN population	2,239,816	5,554	--	--
Basic Nursing Education				
Diploma	951,101	7,679	42.46	0.35
Associate Degree	729,050	11,847	32.55	0.51
Baccalaureate Degree	554,902	7,599	24.77	0.34
Master's degree	2,581	426	0.12	0.02
Doctorate	171	135	0.01	0.01
Not known	2,010	432	0.09	0.02
Employment Status				
Employed in nursing	1,853,024	10,862	82.73	0.38
Not employed in nursing	386,791	8,392	17.27	0.38
Racial/Ethnic Background				
Hispanic	30,441	5,335	1.36	0.24
American Indian/Alaskan native	9,998	1,356	0.45	0.06
Asian/Pacific Islander	75,785	17,554	3.38	0.78
Black	90,611	8,699	4.05	0.39
White	2,018,456	19,439	90.12	0.91
Not known	14,526	1,355	0.65	0.06
Employment setting				
Hospital	1,232,717	12,037	66.53	0.48
Nursing Home Ext. Care	128,983	3,173	6.96	0.14
Nursing Education	36,514	2,023	1.97	0.09
Public/Community Health	180,132	3,948	9.72	0.18
Student Health	50,606	2,753	2.73	0.12
Occupational Health	19,266	1,263	1.04	0.06
Ambulatory care setting				
(Non-Nurse)	138,290	4,289	7.46	0.19
(Nurse)	5,820	955	0.31	0.04
Other	56,263	2,291	3.04	0.10
Not known	4,433	607	0.24	0.03
Type of position				
Administrator or Assistant Administrator	114,619	3,101	6.19	0.14
Consultant	16,771	1,299	0.91	0.06
Supervisor	92,080	2,566	4.97	0.11
Instructor	64,349	2,923	3.47	0.13
Head Nurse of Assistant	84,726	3,635	4.57	0.16
Staff or Gen. Duty	1,233,537	13,066	66.57	0.55
Practitioner/Midwife	26,704	1,331	1.44	0.06
Clinical Specialist	35,524	1,682	1.92	0.08
Nurse Clinician	24,824	1,497	1.34	0.07
Certified Registered Nurse Anesthetist	18,617	1,305	1.01	0.06
Research	7,800	858	0.42	0.04
Private Duty	11,651	1,263	0.63	0.06
Other	120,024	4,418	6.48	0.19
Not known	1,798	363	0.10	0.02

Table B-2. (cont.) Standard errors (S.E.) of selected statistics (number and percentage) for U.S. registered nurse population

Description	Estimated Number	S.E. of estimated Number	Estimated percent	S.E. of estimated percent
Highest Nursing Education				
Diploma	754,848	7,084	33.70	0.32
Associate Degree	632,483	10,760	28.24	0.47
Baccalaureate	671,400	8,353	29.98	0.36
Master's	168,009	3,353	7.50	0.15
Doctorate	11,285	1,004	0.50	0.04
Not known	1,792	432	0.08	0.02
Age of Nurses				
Less than 25 yrs of Age	47,625	2,452	2.13	0.11
25 to 29 Years of Age	198,405	3,595	8.86	0.17
30 to 34 Years of Age	328,190	6,555	14.65	0.29
35 to 39 Years of Age	421,553	5,537	18.82	0.24
40 to 44 Years of Age	355,799	5,762	15.89	0.25
45 to 49 Years of Age	259,093	5,019	11.57	0.22
50 to 54 Years of Age	202,111	4,016	9.02	0.18
55 to 59 Years of Age	164,273	5,908	7.33	0.26
60 to 64 Years of Age	120,687	3,811	5.39	0.17
65 to More Years of Age	126,476	4,234	5.65	0.19
Not known	15,603	1,216	0.70	0.05

Table B-3 Direct Estimates of State Nurse Population, Standard Error, and Coefficient of Variation by State, 1992

State	1992 Estimated State Nurse Population	Standard Error	Coefficient of Variation (in Percent)
United States	2,239,816	5,554	0.25
Alabama	31,908	494	1.55
Alaska	4,153	121	2.92
Arizona	32,988	674	2.04
Arkansas	17,059	426	2.50
California	207,563	1,650	0.79
Colorado	31,948	459	1.44
Connecticut	41,239	912	2.21
Delaware	7,463	339	4.55
District of Columbia	11,812	513	4.34
Florida	119,405	1,692	1.42
Georgia	51,412	741	1.44
Hawaii	9,222	267	2.90
Idaho	7,287	226	3.10
Illinois	110,762	1,692	1.53
Indiana	49,153	833	1.69
Iowa	29,743	399	1.34
Kansas	22,301	417	1.87
Kentucky	28,154	656	2.33
Louisiana	28,169	461	1.64
Maine	12,731	287	2.25
Maryland	43,823	1,229	2.80
Massachusetts	78,322	1,281	1.64
Michigan	85,544	1,401	1.64
Minnesota	45,681	479	1.05
Mississippi	15,334	360	2.35
Missouri	48,167	940	1.95
Montana	7,370	141	1.91
Nebraska	14,829	374	2.52
Nevada	8,943	475	5.32
New Hampshire	13,036	495	3.80
New Jersey	82,555	1,569	1.90
New Mexico	11,121	320	2.88
New York	189,342	2,280	1.20
North Carolina	56,943	733	1.29
North Dakota	6,718	134	2.00
Ohio	107,950	1,376	1.28
Oklahoma	20,032	528	2.64
Oregon	26,561	506	1.90
Pennsylvania	143,449	2,528	1.76
Rhode Island	11,564	282	2.44
South Carolina	24,265	650	2.68
South Dakota	7,704	265	3.44
Tennessee	40,188	615	1.53
Texas	111,778	1,073	0.96
Utah	11,195	193	1.73
Vermont	6,307	391	6.20
Virginia	54,958	924	1.68
Washington	46,626	693	1.49
West Virginia	14,600	612	4.19
Wisconsin	46,905	855	1.82
Wyoming	3,536	105	2.98

of the same size. For the percentage p, this is given by

$$F = n\,\hat{\sigma}_p^2 / [\hat{p}\,(1-\hat{p})] ,$$

where n is the unweighted number of respondents used to determine the denominator of $\hat{p}$.

Direct estimates of the design effect were computed for a set of variables for each State. The averages of the design effects were then computed for each State and the nation. These average design effects can be used in formulas for estimating generalized variances or standard errors. This procedure uses average design effects for a class of estimates instead of calculating direct estimates (with a resulting economy in time and costs), at the sacrifice generally of some accuracy in the variance estimates.

The standard error for an estimated proportion, $\hat{p} = \hat{Y}/\hat{X}$, for a State or for the United States, is provided by the equation:

$$\sigma_{\hat{Y}/\hat{X}} = \sqrt{F \cdot (\hat{Y}/\hat{X}) \cdot (1-\hat{Y}/\hat{X})/n} \qquad (1)$$

where n is the number of survey respondents used to determine the estimate $\hat{X}$. The multiplier F, the average design effect, depends upon the State for which the estimated proportion was generated. Table B-4 contains the list of average design effects for the United States and each State.

Generalized estimates of standard errors can also be computed for estimated numbers (or totals) of RNs in a State, $\hat{Y}$, with a particular characteristic. The estimate $\hat{Y}$ is a subtotal of the estimate $\hat{X}$, the estimated total of RNs working and/or living in the State. The standard error and coefficient of variation of $\hat{X}$ (represented by $C.V._{\hat{X}}$) were determined for the nation and each State (see Table B-3). The following exposition is made simpler by defining the rel-variance of an estimate as the square of its coefficient of variation.

Then the rel-variance of the ratio of $\hat{Y}$ to $\hat{X}$ (called $V^2_{\hat{Y}/\hat{X}}$) can be calculated as:

$$V^2_{\hat{Y}/\hat{X}} = \frac{F(1-\hat{Y}/\hat{X})}{n(\hat{Y}/\hat{X})} ,$$

Table B-4 Average Design Effects for Percentages Estimated from the Fifth National Sample Survey of Registered Nurses, 1992

State	Average Design Effect
United States	1.91
Alabama	1.06
Alaska	1.03
Arizona	1.12
Arkansas	1.13
California	1.13
Colorado	1.10
Connecticut	1.24
Delaware	1.40
District of Columbia	1.15
Florida	1.15
Georgia	1.02
Hawaii	1.35
Idaho	1.21
Illinois	0.98
Indiana	1.01
Iowa	1.06
Kansas	1.06
Kentucky	1.02
Louisiana	1.07
Maine	1.06
Maryland	1.07
Massachusetts	1.15
Michigan	1.02
Minnesota	0.99
Mississippi	1.17
Missouri	1.06
Montana	0.96
Nebraska	1.03
Nevada	1.71
New Hampshire	1.18
New Jersey	1.06
New Mexico	1.28
New York	1.14
North Carolina	1.05
North Dakota	1.20
Ohio	1.01
Oklahoma	1.07
Oregon	0.97
Pennsylvania	1.20
Rhode Island	0.97
South Carolina	1.05
South Dakota	1.52
Tennessee	1.08
Texas	1.11
Utah	1.08
Vermont	1.59
Virginia	1.09
Washington	1.25
West Virginia	1.23
Wisconsin	1.16
Wyoming	1.07

where F is the design effect for the State of interest and n is the number of respondents to the survey (i.e., the number in the sample that were weighted to obtain the estimate $\hat{X}$).

Then we can approximate the rel-variance of $\hat{Y}$, denoted $V^2_{\hat{Y}}$, using

$$V^2_{\hat{Y}} = V^2_{\hat{Y}/\hat{X}} + (C.V._{\hat{X}})^2 \ .$$

Finally, the variance of $\hat{Y}$ can be estimated by multiplying by the rel-variance above, $V^2_{\hat{Y}}$. The standard error of $\hat{Y}$, $\sigma_{\hat{Y}}$, is thus estimated as

$$\sigma_{\hat{Y}} = \hat{Y}\sqrt{\hat{V}^2_{\hat{Y}}} \quad . \tag{2}$$

The standard error of an estimated percentage for a region of the United States depends upon a linear combination of the variance of the same estimated percentages for the States comprising that particular region. The estimated proportion (or percentage) for the region is

$$\hat{Y}_R/\hat{X}_R = \frac{\sum_{s=1}^{h} \hat{Y}_s}{\sum_{s=1}^{h} \hat{X}_s} \quad .$$

where h is the number of States in region (R), and $\hat{Y}_s$ and $\hat{X}_s$ are estimates for a particular State. The formula used to approximate the standard error of an estimated proportion for a region is

$$\sigma_{\hat{Y}_R/\hat{X}_R} = \sqrt{\sum_{s=1}^{h} (\hat{X}_s^2 \sigma^2_{\hat{Y}_s/\hat{X}_s}) \ / \ (\sum_{s=1}^{h} \hat{X}_s)^2} \tag{3}$$

where $\sigma_{\hat{Y}_s/\hat{X}_s}$ represents the standard error of the estimated proportion $\hat{Y}_s/\hat{X}_s$ for the States and the standard errors are estimated from equation (1) or from direct estimation.

The direct standard error for an estimated number for a region of the United States also depends upon a linear combination of the variance of the same estimated numbers for the States which comprise the region. The formula used is

$$\sigma_{\hat{Y}_R} = \sqrt{\sum_{s=1}^{h} \sigma_{\hat{Y}_s}^2} \tag{4}$$

where the standard error $(\sigma_{\hat{Y}_s})$ of the estimated number $\hat{Y}_s$ is available either from the direct procedures or from formula (2).

Illustrative examples of the computation of the generalized variance are given in Chart B-1.

Chart B-1 Illustrative Examples of Generalized Variance Estimates

1. Estimated Percentages (or proportions) for a State or the United States

   a) Percent of nurses located in New York who were employed in nursing on a full-time basis:

   $p = 60.2 \qquad F = 1.14 \qquad n = 1{,}409$

   $\hat{\sigma} = [1.14\ (.602)(.398)/1409]^{1/2} = .014$ or 1.4%

   b) Percent of employed nurses in the United States who were working in hospitals:

   $p = 66.5 \quad F = 1.91 \qquad n = 27{,}336$

   $\hat{\sigma} = [1.91\ (.665)(.335)/27{,}336]^{1/2} = .0039$ or .39%

2. Estimated number for a State or the United States

   a) Estimated number of nurses located in New York State who were not employed in nursing:

   $\hat{Y} = 30{,}045 \quad \hat{X} = 189{,}342 \quad \hat{Y}/\hat{X} = .159 \quad n = 1{,}409 \quad C.V._{\hat{x}} = 1.20\%$

   $F = 1.14$

   $\hat{V}^2_{\hat{Y}} = [(1.14)(.841)/1409(.159)] + (.0120)^2 = .0044$

   $\sigma_{\hat{Y}} = 30{,}045\ (.0044)^{1/2} = 1{,}998$

   b) Estimated number of nurses located in United States who were employed in nursing:

   $\hat{Y} = 1{,}853{,}024 \qquad \hat{X} = 2{,}340{,}816 \qquad n = 32{,}304$

   $C.V._{\hat{x}} = .0025 \quad \hat{Y}/\hat{X} = .827 \qquad F = 1.91$

   $\hat{V}^2_{\hat{Y}} = [(1.91)(.173)/32{,}304(.827)] + (.0025)^2 = .0000186$

   $\sigma_{\hat{Y}} = 1{,}853{,}024\ (.0000186)^{1/2} = 7{,}996$

Chart B-1 (cont.) Illustrative Examples of Generalized Variance Estimates

3. Standard error of a regional estimate (or a grouping of States)

Estimated percent of nurses employed in nursing in the Middle Atlantic region:

$Y/X = 80.0\%$

New Jersey:

$$\sigma_{Y/X} = [1.06(.7815)(.2185)/673]^{1/2} = .016 \text{ or } 1.6\%$$

New York:

$$\sigma_{Y/X} = [1.14(.8413)(.1587)/1409]^{1/2} = .010 \text{ or } 1.0\%$$

Pennsylvania:

$$\sigma_{Y/X} = [1.20(.7575)(.2425)/1164]^{1/2} = .014 \text{ or } 1.4\%$$

Middle Atlantic Region:

$$\sigma_{Y/X} = \{[(82{,}555)^2(.016)^2+(189{,}342)^2(.010)^2+(143{,}449)^2 (.014)^2]/(82{,}555 + 189{,}342 + 143{,}449)^2\}^{1/2} = .0074 \text{ or } .74\%$$