Appendix B
Survey Methodology

This appendix provides a brief summary of the methodology of the study including the sample design and the statistical techniques used in summarizing the data. It also includes a discussion of sampling errors, provides the standard errors for key variables in the study and presents a simplified methodology for estimating standard errors. Much of the material included here has been abstracted from the technical report provided by the Research Triangle Institute (RTI), the contractor who carried out the sampling for and conducted the sixth National Sample Survey of Registered Nurses discussed in this report.

Sample Design

The six surveys carried out to date all followed the same design developed by Westat, Inc. under a contract with the Division of Nursing, BHPr, HRSA in 1975–76. The design approach took into account two key characteristics of the sampling frame. First, no single list of all individuals with licenses to practice as registered nurses in the United States exists although lists of those who have licenses in any one State are available. Second, a nurse may be licensed in more than one State.

A sampling frame was required to select a probability sample of nurses from which valid inferences could be made to the target population of all those with current licenses to practice in the United States. State Boards of Nursing in the 50 States and in the District of Columbia (hereafter also referred to as a State) provided files containing the name, address, and license number of every RN currently licensed in that State. These 51 files constituted a multiple sampling frame containing all the RNs licensed in the 51 States. Because many nurses are licensed in more than one State, their names could appear in the combined list more than once. A nested alpha-segment design was used to properly determine selection probabilities for nurses listed in more than one State.

The target population of this study was the current RN population of the United States as of March 1996. RNs were selected with equal probabilities within States. Whether RNs fell into the sample depended on whether their names fell within one of the alpha-segments or portions of alpha-segments that were selected for the sample. 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.

Both national and State-level estimates were required. While uniform sampling rates would have produced the best national estimates, the resulting sample sizes for the smallest States would have been inadequate to support State-level estimates. Sampling rates were increased in the smaller States to obtain larger State-level sample sizes. Planned sampling rates ranged from less than 1 percent in several of the States with a large RN population to 14 percent in Wyoming. Planned State sizes ranged from a sample of over 2,000 RNs in California to approximately 700 in Wyoming. While this disproportionate sampling improved the precision of estimates in the smaller States, it also reduced precision of national estimates due to unequal weighting effects.

Registered nurses were in the sample on the basis of name, with an RN being included in the sample if the name of licensure fell within a specific portion of the alpha-segments included in the sample from the RN's State of licensure. As stated earlier, an alpha-segment consisted of all alphabetically adjacent names falling between set boundaries. The segments were constructed so that each segment contained approximately the same number of RNs. Specifically, the lower boundary of an alpha-segment was the last name in alphabetical order of all the names included in that segment. The membership of the segment consisted of all names, beginning with the lower boundary, *up to but not including* a name that defined the upper boundary. The latter name fell into the next alpha-segment.

A planned variation in the size of the portions of segments was used to accommodate the differing State sampling rates. The largest portions used were full alpha-segments while other sizes were ½-, ¼-, ⅛-, 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.

From the frame of 250 alpha-segments, 40 alpha-segments were randomly selected. Although each State had 40 sample segments (i.e., portions of alpha-segments), the segments differed in size depending on the State's sampling rate. To identify and account for nurses having multiple licenses, the alpha-segment portions from larger States were "nested," or included, within those from 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 State with a larger RN population) were nested within those of the higher sampling rate (the State with the smaller RN population). The nested alpha-segment design permitted the use of each sample RN's data for State estimates of each of her/his States of licensure and also provided appropriate (multiplicity-adjusted) weights for both State and national estimates.

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 ½-portion selection and ¼-portion, ⅛-portion, 1⁄16-portion, and 1⁄32-portion selections.

The sampling rate for a particular State was obtained from some combination of the alpha-segments and portions. For example, the 40 complete alpha-segments would have 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 40×0.4 percent, or 16 percent, of those names.) The sample for a State with an 8 percent sampling rate consisted of the 40 ½-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 by using the ¼-portions from the first group and ½-portions from the second group $(0.4 \times [(30 \times ¼) + (10 \times ½)] = 5)$.

The survey design specified precisely which alpha-segments and portions would correspond to each of the different sampling rates used. This design 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 boundaries 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 from 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 several of the sample alpha-segments in the periodic surveys to be systematically replaced. Under the original survey design, the 40 sample alpha-segments were randomly assigned to five panels of eight alpha-segments each. For each successive survey, a new panel (consisting of eight new alpha-segments) was entered into the sample, replacing one of the five panels that was in the previous survey. Under this scheme, a nurse who maintained an active license in the same State(s) and whose name did not change could be retained in the sample for up to five surveys. With the reconstruction of the alpha-segments in the fourth RN survey (1988), changes were made so that exact correspondence of the current segments to those established initially may no longer exist; therefore, some nurses may not have been carried through all five surveys.

Each of the 51 State Boards of Nursing provided one or more files that contained the names of currently licensed RNs. These files formed the basis of the sampling frame from which the RNs for each 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. For this current study, States were also asked to identify those for whom the State provided advanced practice nurse (APN) status. In some cases, these APNs were identified on separate lists and their APN status was added to the information on the RN sampling frame list. In other cases, the State identified these nurses on the basic list provided. Once a licensure file provided by a State contained all appropriate names of individuals with active RN licenses and met all specifications, the required sample names in that file were selected. Regardless of the way a State alphabetized and standardized the names in 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"). Whether or not the RN was an APN was not taken into account in the sample selection.

Table B-1 shows the sampling rates and sample sizes that were planned and actually obtained for the 51 States in the survey. Differences between planned and actual sampling rates result from State-specific variation in nurses' names. States are priority ordered by frame size (smaller to larger) so that sampling rates are decreasing down the table.

The original State frame sizes were adjusted to account for duplicate licenses within States and ineligible licenses (i.e., frame errors) found in the sample. Duplicates within States arose primarily from combining RN and APN lists. Most duplicates were identified before selecting the sample and determining the frame size, but a few were identified after sample selection, requiring a frame size adjustment. The ineligible licenses were identified in the process of reconciling the State and nurse reported licenses. Cases that could not be reconciled by RTI were sent to the State Boards of Nursing for resolution. No changes in the sampling rates occurred as a result of the frame adjustments, so the nesting of the alphabetic clusters remained the same even though the ordering by adjusted frame would have changed. It was, therefore, not necessary to change the priority ordering as a result of any changes in relative size.

Weighting Procedures

The probability sample design of the survey 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 in her/his priority State, adjusted to account for nonresponse.

A nurse is uniquely linked on the national sampling frame with her/his "priority State," i.e., the State with the lowest number of licensed RNs in which she or he was licensed and selected into the sample. All nurses with the same priority State had an equal probability of being selected within that State, so 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-priority 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

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

State	Frame[1] Size	Percent Sample Rate[2] Planned	Percent Sample Rate[2] Actual	Actual Sample Size
Total	2,878,444			45,339
Wyoming	4,937	14.00	15.31	763
Alaska	7,320	12.00	9.54	698
North Dakota	7,404	9.00	9.82	728
Vermont	7,521	9.00	8.97	675
Montana	9,656	7.00	7.03	679
South Dakota	9,746	7.00	6.89	674
Idaho	10,060	7.00	6.91	701
Hawaii	10,887	7.00	6.48	706
Nevada	11,590	7.00	6.74	785
Delaware	11,770	6.00	5.64	711
New Mexico	13,546	5.00	5.37	727
Utah	14,872	5.00	5.11	760
Rhode Island	15,939	4.50	3.85	616
New Hampshire	16,110	4.50	4.21	679
Maine	17,510	4.00	3.73	655
Nebraska	19,429	3.50	3.33	650
District of Columbia	20,438	3.50	3.57	733
West Virginia	20,815	3.50	3.30	687
Mississippi	23,513	3.00	3.48	822
Arkansas	25,820	3.00	3.21	830
South Carolina	27,910	3.00	3.03	853
Oklahoma	28,471	3.00	3.17	902
Kansas	28,266	3.00	3.13	892
Oregon	32,553	2.50	2.39	779
Iowa	36,023	2.50	2.40	867
Louisiana	36,541	2.00	2.06	753
Kentucky	38,041	2.00	1.90	728
Alabama	40,223	2.00	2.02	811
Arizona	41,881	2.00	1.85	773
Colorado	41,599	2.00	2.09	884
Connecticut	49,113	1.75	1.53	756
Minnesota	53,887	1.50	1.64	882
Maryland	56,089	1.50	1.53	856
Washington	56,880	1.50	1.49	849
Tennessee	57,898	1.25	1.29	748
Wisconsin	61,875	1.25	1.20	748
Missouri	65,336	1.25	1.23	802
Indiana	67,425	1.25	1.15	780
Georgia	71,389	1.25	1.39	991
North Carolina	73,374	1.25	1.23	899
Virginia	75,469	1.25	1.15	871
Massachusetts	102,628	1.00	0.90	919
New Jersey	111,767	1.00	0.90	1,010
Michigan	116,133	0.90	0.85	984
Ohio	128,230	0.90	0.91	1,161
Illinois	135,553	0.90	0.89	1,225
Texas	147,756	0.90	0.87	1,284
Florida	152,295	0.90	0.88	1,338
Pennsylvania	192,299	0.90	0.91	1,761
New York	219,124	0.90	0.84	1,855
California	253,533	0.90	0.83	2,099

[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 rates.

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 a 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 the State B weight, W(B), the numerator was the total frame count of RNs licensed in State B, N(B), after removing 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 removing 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 weighting scheme incorporated a nonresponse adjustment that inflated the respondents' data to represent the entire universe. The adjusted frame total shown in Table B-1 was used in computing the State I weight.

Estimation Procedure

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:

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

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.

Sampling and Nonsampling 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 are also reflected in the sampling error estimates.

Nonsampling Errors

Some sources of error—such as unusable responses to vague or sensitive questions; no responses 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 this survey were careful planning, keeping nonresponses to the lowest feasible level, and coding and processing the sample data carefully.

If nonsampling errors are random, in the sense that they are independent and tend to be compensating from one respondent to another, then they do not cause bias in estimates of totals, percents, or averages. Furthermore, the contribution from such nonsampling errors will automatically be included in the sampling errors that are estimated from the sample data.

Although nonsampling errors that are randomly compensating do not tend to bias estimates of simple statistics, correlations or relationships in cross-tabulations are often decreased by such errors, and sometimes substantially. Thus, errors that tend to be compensated in estimates of simple aggregates or averages may (but not necessarily will) introduce systematic errors or biases in measures of relationships or cross-tabulations.

Nonsampling errors that are systematic rather than random and compensating are a source of bias for sample estimates. Such errors are not reduced by increasing the size of the sample, and the sample data do not provide an assessment of the magnitude of these errors. Systematic errors are reduced in this study by such things as careful wording of questionnaire items, respondent motivation, and well-designed data-collection and data-management procedures. However, such errors sometimes occur in subtle ways and are less subject to design control than is the case for sampling errors.

Nonresponse to the survey is one source of nonsampling error because a characteristic being estimated may differ, on average, between respondents and nonrespondents. For this reason, considerable effort has been expended in this survey to obtain a high response rate through such actions as respondent motivation and follow-up procedures. A high response rate reduces both random and systematic errors. After taking into account duplicates and frame errors, the overall response rate to this survey was 72.3 percent. State-level response rates ranged from a little over 60 percent in the District of Columbia and Nevada to 85.4 percent in North Dakota.

Sampling 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 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. Standard errors were estimated based upon the assumption that the systematic sample of 40 alpha-segments is equivalent to 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 20 degrees of freedom) for variance estimation, the probability is approximately .95 that the statistic of interest differs from the value of the population characteristic that it estimates by not more than 2.086 standard deviations.

Specifically, a 95 percent confidence interval for an estimated statistic $\hat{x}$ takes the form

$$\hat{x} \pm 2.086\hat{\sigma}_{\hat{x}}$$

where $\hat{\sigma}_{\hat{x}}$ is the estimated standard error of $\hat{x}$.

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 1 alpha-segment from the 20th pair. The actual respondent count in the included segments for a particular replicate was approximately 39/40ths of the full respondent sample and was 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 number of responding nurses from the 39 segments associated with each replicate. Having 20 sets of weights permits construction of 20 replicate estimates to compare with the estimate produced from all of the data; each replicate estimate is based on about 39/40ths of the data.

This procedure was performed 20 times, once for each pair of alpha-segments.

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

$\hat{X}_i$ = an estimated total for replicate I associated with alpha-segment pair I, and

$\hat{X}$ = 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 to the total number of RNs resident in Florida), 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 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. 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 variables in Table B-2. Table B-3 shows the estimated State population of nurses and the standard error of these population totals.

Design Effects and Generalized Variances

The generalized variance is a model-based approximation of 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 is determined by taking the ratio of the estimated sampling variance, $\sigma^2_{\hat{p}}$, obtained by the jackknife method, to the sampling variance of the $\hat{p}$ simple random sample of the same size. For the proportion $\hat{p}$, this is given by

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

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

Table B-2. Estimates and Standard Errors (S.E.) For Selected Variables for U.S. Registered Nurse Population

Description	Estimated Number	S.E. of Estimated Number	Estimated Percent	S.E. of Estimate Percent
Number of Nurses	2,558,874	4,802		
Basic Nursing Education				
Diploma	910,618	7,547	35.59	0.2988
Associate Degree	965,059	12,589	37.71	0.4716
Baccalaureate Degree	675,685	11,140	26.41	0.4391
Master Degree	5,229	1,097	0.20	0.0428
Doctorate (N.D.)	309	170	0.01	0.0066
Unknown/Refused	1,974	453	0.08	0.0177
Employed in Nursing				
Yes	2,115,815	6,647	82.69	0.2721
No	443,059	7,304	17.31	0.2721
Racial/Ethnic Background				
Hispanic	40,559	7,375	1.59	0.2881
American Indian/Alaskan	11,843	1,517	0.46	0.0597
Asian/Pacific Islander	86,434	19,171	3.38	0.7509
Black/Not Hispanic	107,527	14,204	4.20	0.5528
White/Not Hispanic	2,294,092	25,544	89.65	0.9765
Unknown/Refused	18,417	1,629	0.72	0.0637
Employment Status in 1996				
Employed in Nursing FT	1,510,318	10,629	59.02	0.4090
Employed in Nursing PT	605,497	7,780	23.66	0.3105
Not Employed in Nursing	443,059	7,304	17.31	0.2721
Graduation Year				
Before 1961	351,033	8,939	13.72	0.3405
1961 to 1965	173,855	3,254	6.79	0.1271
1966 to 1970	211,971	5,650	8.28	0.2291
1971 to 1975	299,868	6,855	11.72	0.2645
1976 to 1980	374,879	6,289	14.65	0.2339
1981 to 1985	385,167	4,928	15.05	0.1945
1986 to 1990	338,468	6,757	13.23	0.2708
After 1990	417,580	7,235	16.32	0.2805
Unknown/Refused	6,054	885	0.24	0.0345
Highest Nursing Education				
Diploma	696,804	8,352	27.23	0.3319
Associate Degree	812,438	12,457	31.75	0.4755
Baccalaureate	799,507	10,900	31.24	0.4296
Master's	231,978	4,918	9.07	0.1866
Doctorate	16,465	1,314	0.64	0.0514
Unknown/Refused	1,682	432	0.07	0.0168
Age of Nurse				
<25	58,012	3,060	2.27	0.1214
25 to 29	170,277	5,054	6.65	0.1983
30 to 34	297,119	6,844	11.61	0.2763
35 to 39	413,931	8,645	16.18	0.3366
40 to 44	465,188	7,095	18.18	0.2663
45 to 49	378,569	6,458	14.79	0.2518
50 to 54	263,635	6,136	10.30	0.2434

Table B-2. (continues)

Description	Estimated Number	S.E. of Estimated Number	Estimated Percent	S.E. of Estimate Percent
Age of Nurse (continues)				
55 to 59	201,114	6,369	7.86	0.2426
60 to 64	147,951	4,940	5.78	0.1946
>= 65	145,849	5,631	5.70	0.2153
Unknown/Refused	17,230	1,412	0.67	0.0546
Marital Status and Children				
Married Child < 6	217,039	4,484	8.48	0.1796
Married Child ≥ 6	753,218	7,748	29.44	0.2978
Married Child < 6 and ≥ 6	208,027	3,870	8.13	0.1502
Married No Children	663,959	8,082	25.95	0.3021
Married Child Unknown	7,298	888	0.29	0.0345
Wid/Sep/Div Child < 6	12,598	1,557	0.49	0.0608
Wid/Sep/Div Child ≥ 6	170,756	4,675	6.67	0.1813
Wid/Sep/Div Child All	18,513	1,317	0.72	0.0516
Wid/Sep/Div No Children	245,709	9,110	9.60	0.3568
Wid/Sep/Div Child Unknown	1,834	536	0.07	0.0210
Never Married	251,484	5,537	9.83	0.2154
Unknown/Refused	8,438	828	0.33	0.0325
Employment Setting (For nurses employed in nursing)				
Hospital	1,270,870	9,602	49.67	0.3831
Nursing Home Extended Care	170,856	4,810	6.68	0.1902
Nursing Education	48,918	2,699	1.91	0.1053
Public Health Community Health	278,141	5,055	10.87	0.2009
Student Health	62,932	3,505	2.46	0.1364
Occupational Health	21,575	1,525	0.84	0.0604
Ambulatory Care/Not Owned	170,589	6,303	6.67	0.2425
Nurse Owned/Operated Ambulatory Care	8,341	1,111	0.33	0.0432
Other	82,635	2,465	3.23	0.0970
Unknown/Refused	957	231	0.04	0.0090
Type of Position (For nurses employed in nursing)				
Administrator/Assistant Administrator	112,134	3,604	4.38	0.1402
Consultant	27,020	2,112	1.06	0.0825
Supervisor	95,451	3,826	3.73	0.1514
Instructor	73,084	3,536	2.86	0.1376
Head Nurse or Assistant	123,231	3,574	4.82	0.1375
Staff or General Duty	1,309,596	11,085	51.18	0.4556
Nurse Practitioner/Midwife	44,904	2,514	1.75	0.0980
Clinical Specialist	35,620	2,421	1.39	0.0946
Nurse Clinician	30,396	1,754	1.19	0.0680
Certified Nurse Anesthetist	21,827	1,995	0.85	0.0780
Research	12,665	1,581	0.49	0.0620
Private Duty	15,947	1,448	0.62	0.0562
Not Applicable	0	0	0.00	0.0000
Unknown/Refused	2,422	568	0.09	0.0222
Mean Gross Annual Salary for Full-Time RNs	42,071	161		
Mean Scheduled Hours Per Year	1,742	4		
Mean Hours Worked in Week Beginning on March 18,1996	36	0.1		

Table B-3. Direct Estimates of State Nurse Population, Standard Error, and Coefficient of Variation by State, 1996

State	1996 Estimated State Nurse Population	Standard Error	Coefficient of Variation (in Percent)
United States	2,558,874	4,802	0.19
Alabama	37,188	750	2.02
Alaska	6,651	289	4.35
Arizona	40,313	962	2.39
Arkansas	20,890	483	2.31
California	233,404	2,427	1.04
Colorado	37,289	737	1.98
Connecticut	41,296	770	1.86
Delaware	9,538	408	4.27
District of Columbia	9,948	749	7.53
Florida	148,046	2,218	1.50
Georgia	62,526	926	1.48
Hawaii	10,236	474	4.63
Idaho	8,627	250	2.90
Illinois	124,332	1,552	1.25
Indiana	56,420	1,076	1.91
Iowa	32,303	595	1.84
Kansas	24,452	561	2.29
Kentucky	32,427	745	2.30
Louisiana	33,969	492	1.45
Maine	15,507	311	2.00
Maryland	48,789	1,018	2.09
Massachusetts	87,995	1,890	2.15
Michigan	99,676	1,694	1.70
Minnesota	50,909	606	1.19
Mississippi	20,979	461	2.20
Missouri	58,096	798	1.37
Montana	8,417	169	2.01
Nebraska	16,909	319	1.89
Nevada	11,336	466	4.12
New Hampshire	12,938	406	3.14
New Jersey	88,404	1,722	1.95
New Mexico	13,185	363	2.75
New York	195,293	2,526	1.29
North Carolina	69,231	1,154	1.67
North Dakota	7,248	210	2.89
Ohio	118,612	1,400	1.18
Oklahoma	23,583	456	1.93
Oregon	29,239	716	2.45
Pennsylvania	160,149	2,111	1.32
Rhode Island	12,915	338	2.62
South Carolina	29,135	590	2.02
South Dakota	9,035	271	3.00
Tennessee	51,111	877	1.72
Texas	136,757	1,981	1.45
Utah	14,059	398	2.83
Vermont	6,300	275	4.36
Virginia	66,436	1,326	2.00
Washington	52,411	556	1.06
West Virginia	16,810	585	3.48
Wisconsin	53,040	793	1.49
Wyoming	4,512	276	6.11

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.

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

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

where n is the number of survey respondents used to determine the estimate $\hat{X}$. The multiplier F, the median* design effect, depends upon the State for which the estimated proportion was generated. The median design effects are on Table B-4.

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 (such as those employed in hospitals). 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 for each State. The following explanation is made simpler by defining the relative variance of an estimate as the square of its coefficient of variation.

Then the relative variance of the ratio of to (called) can be calculated as:

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

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 relative variance of $\hat{Y}$, denoted $V^2_{\hat{Y}}$, using

*The median design effect was based on all design effects for estimates of proportions computed on selected variables. Using a median instead of mean value avoids the effects of extreme estimates of standard errors which can occur for some relatively rare attributes. In prior years, an average (mean) design effect was computed for selected variables. Given that the distribution of design effects is skewed to the right, it is expected that the true median be less than the true mean.

Table B-4. Median Design Effects for Percentages Estimated from the Sixth National Sample Survey of Registered Nurses, 1996

State	Median Design Effect
United States	1.72
Alabama	1.02
Alaska	1.11
Arizona	0.94
Arkansas	1.01
California	1.17
Colorado	0.96
Connecticut	1.02
Delaware	1.11
District of Columbia	0.94
Florida	1.10
Georgia	1.00
Hawaii	1.27
Idaho	0.99
Illinois	1.04
Indiana	0.93
Iowa	1.01
Kansas	1.01
Kentucky	0.99
Louisiana	1.02
Maine	1.04
Maryland	1.02
Massachusetts	1.06
Michigan	1.01
Minnesota	0.98
Mississippi	0.92
Missouri	1.01
Montana	1.01
Nebraska	1.04
Nevada	0.99
New Hampshire	1.03
New Jersey	1.05
New Mexico	1.11
New York	1.05
North Carolina	1.00
North Dakota	0.94
Ohio	0.95
Oklahoma	1.01
Oregon	0.96
Pennsylvania	1.07
Rhode Island	0.96
South Carolina	1.14
South Dakota	1.00
Tennessee	1.06
Texas	1.26
Utah	1.05
Vermont	1.07
Virginia	1.01
Washington	1.13
West Virginia	0.98
Wisconsin	1.01
Wyoming	1.01

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

This approximation is based on the first-order Taylor series approximation to the variance of a product and the assumption of zero correlation between the estimate of ratio and the denominator of the ratio.

Finally, the variance of $\hat{Y}$ can be estimated by multiplying by the relative variance above by the square of the estimate, $\hat{Y}$. The standard error of $\hat{Y}$, $\sigma_{\hat{Y}}$, is thus estimated as

$$\sigma_{\hat{Y}} = \hat{Y}\sqrt{\hat{V}^2_{\hat{Y}}} \qquad (2)$$

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 making up that particular region. The estimated proportion for the region is

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

here 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

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

where $\sigma_{\hat{Y}_s}/\hat{X}_s$ represents the standard error of the estimated proportion Y_s/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 that make up the region. The formula used is

$$\sigma_{\hat{Y}_R} = \sqrt{\sum_{s=1}^{h} \sigma^2_{\hat{Y}_s}} \qquad (4)$$

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

Illustrative examples of the computation of the generalized variance appear on the following page.

Illustrative Examples of Generalized Variance Estimates

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

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

      $p = 60.7$ $\quad F = 1.05$ $\quad n = 1,151$

      $\sigma = [1.05\,(.607)(.393)/1,151\,]^{1/2} = 0.015$ or 1.5%

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

      $p = 49.7$ $\quad F = 1.72$ $\quad n = 29,837$

      $\sigma = [1.72\,(.497)(.503)/29,837\,]^{1/2} = 0.004$ or 0.4%

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:

      $\hat{Y} = 29,626$ $\quad X = 195,293$ $\quad \hat{Y}/X = 0.1517$ $\quad n = 1,151$ $\quad C.V._x = 1.29\%$

      $F = 1.05$

      $V_{\hat{Y}}^2 = [(1.05)(.8483)/(1,151(.1517))] + (.0129)^2 = 0.0053$

      $\sigma_{\hat{Y}} = 29,626\,(.0053)^{1/2} = 2,157$

   b. Estimated number of nurses located in United States who were not employed in nursing:

      $\hat{Y} = 443,059$ $\quad X = 2,558,874$ $\quad n = 29,837$

      $C.V._x = 0.19\%$ $\quad \hat{Y}/X = 0.1731$ $\quad F = 1.72$

      $V_{\hat{Y}}^2 = [(1.72)(.8296)/(29,837(.1731))] + (.0019)^2 = 0.0003$

      $\sigma_{\hat{Y}} = 443,059\,(.0003)^{1/2} = 7,674$

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

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

   $Y/X = .802$ or 80.2%

   $\sigma_{\hat{Y}/XNJ} = [1.05\,(.7625)(.2375)/594]^{-} = 0.0179$ or 1.79%

   $\sigma_{\hat{Y}/XNY} = [1.05\,(.8483)(.1517)/1,151]^{-} = 0.0108$ or 1.08%

   $\sigma_{\hat{Y}/XPA} = [1.07\,(.7671)(.2329)/1,234]^{-} = 0.0124$ or 1.24%

   $\sigma_{\hat{Y}/XR} = \{[(88,404)^2\,(.0195)^2 + (195,293)^2\,(.0116)^2 + (160,149)^2\,(.0135)^2]/$

   $(88,404 + 195,293 + 160,149)^2\}^{-} = .0081$ or .81%