APPENDIX B

SURVEY METHODOLOGY

The design for the national sample surveys of registered nurses was 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 a 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. All four sample surveys of registered nurses carried out to date have used this same design. The description of the survey methodology included here has been abstracted from a technical report of the adaptation of the survey design to this current study prepared by Research Triangle Institute (RTI), the contractor for the 1988 study.

Sample Design

The sample was selected from a universe of possible RN names developed from the respective lists of those with current licenses to practice in each State and the District of Columbia. The sampling frame was an alphabetically ordered list of names. This list was partitioned into 250 alphabetic segments, or alpha-segments, of nearly equal size (i.e., equal number of names).

Each of the 51 State Boards of Nursing provided one or more files that contain the names of currently licensed RNs. The files provided by a State formed the basis of the sampling frame from which the sample of RNs for that State was selected.

Thus, RNs are selected for inclusion in the sample, with equal probabilities of selection within States, on the basis of name. Whether an RN falls into the sample depends on whether her/his name falls within one of the alphabetic segments (called alpha-segments) or portions of alpha-segments that are selected for the sample. An alpha-segment consists of all alphabetically adjacent names falling between boundaries determined in such a manner that each pair of boundaries contains approximately the same proportion of RN names in the country. Approximately equal-sized alpha-segments are constructed by partitioning an alphabetically ordered list of RN names into the 250 equal (or as nearly equal as possible) segments.

In response to the concern that the RN name distribution, and hence alpha-segment sizes, may have changed since the original alpha-segment construction was carried out in 1977 when the first of this series of studies was conducted, RTI did an analysis of those alpha-segments in relation to lists of licensees received from States for the 1984 study. On the basis of that analysis, the 250 alpha-segments were reconstructed for the 1988 study.

Because State-level estimates are desired, different sampling rates were set for the States based on considerations of the statistical precision of the estimates and the costs involved. States in which smaller numbers of RNs are currently licensed were assigned higher sampling rates than were larger States to yield a sample large enough to provide State estimates of reasonable precision. The use of differential State sampling rates substantially reduced variations in State sample sizes and thus permitted more precise State-level estimation.

From the universe of RN names, divided into 250 alpha-seqments, each State's sample consists of 40 primary sampling units (portions of alpha-segments). Therefore, a total of 40 alpha-segments are used for sample selection for the study. However, although each State has the same number (40) of sample segments (i.e., portions of alpha-segments), the segments differ in size depending on the State's sampling rate.

To accommodate the differing State sampling rates there is a planned variation in the size of the portions of segments used. The size of an alpha-segment portion is the proportion of all RNs that it contains. The largest portions used are the full alpha-segments, while those of other sizes are called 1/2-portions, 1/4-portions, 1/8-portions, 1/16-portions and 1/32-portions. The fractions indicate 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 is achieved using a combination of these portions of alpha-segments.

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

The "nesting" results from the manner in which the 40 basic alpha-segment selections are used in defining the sample for each State. Each of these alpha-segments, or one of the fractional portions of it, constitutes one of the 40 sample clusters for each State. Accordingly, each of the basic alpha-segments has 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 is selected at random from one of its two halves, and that same half-portion is in the sample for all States for which half-portions are sampled from this alpha-segment. Similarly, each of the smaller sample portions is randomly selected from the two halves of the next largest sample portion; e.g., a sample 1/8-portion is a randomly selected half of the corresponding sample 1/4-portion.

The sampling rate for a particular State is obtained from some combination of the alpha-segments and portions. For example, the 40 alpha-segments, in their entirety, constitute the sample for States with a 16 percent sampling rate (because each segment contains an expected 0.4 percent of the State's RN names, taken together they contain an expected 40 x 0.4 percent or 16 percent of those names). The sample for a State with an 8 percent sampling rate consists of the 40 1/2-portion selections. A 5-percent sampling rate is 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 x [(30x1/4) + (10x1/2)] = 5).

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. Table B1 shows the sampling rates and sample sizes that were planned and actually obtained for the 51 States in the 1988 survey. Column (A) in the table gives the number of those with current RN licenses in each State as available in the sampling frame for the State. Column (B) presents the sampling rate planned for each State. The table lists the States in "priority" order by frame size (smaller to larger) so that sampling rates are decreasing down the table. The expected sample size given in column (D) is the product of the planned sampling rate (B) and the State total frame number of licensees.

RNs in the sample 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 and assigned, for weighting purposes, to the highest priority State on whose sample frame the RN appeared. 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, 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.

Table B1. State Sampling Rates and Sample Sizes (Priority-Ordered)

State	Frame Size	Sampling Rate*		Sample Size		Percent
	Active Licenses	Planned	Actual	Planned	Actual	Difference
	(A)	(B)	(C)	(D)	(E)	(F)
Wyoming	3,746	16.0	15.2	599	568	-5.2
Alaska	4,568	12.0	11.2	548	512	-6.6
North Dakota	6,636	9.0	7.5	597	498	-16.6
Vermont	6,828	9.0	9.2	615	626	1.8
Nevada	7,219	9.0	8.2	650	598	-8.0
Idaho	7,387	9.0	9.5	665	705	6.0
South Dakota	7,503	9.0	8.7	675	655	-3.0
Montana	8,300	8.0	7.9	664	654	-1.5
Delaware	8,438	8.0	8.5	675	721	6.8
Hawaii	8,732	8.0	8.0	699	694	-0.7
Utah	9,672	7.0	7.1	677	690	1.9
New Mexico	9,682	7.0	7.0	678	676	-0.3
Rhode Island	13,322	6.0	6.4	799	850	5.4
Maine	13,690	6.0	6.4	821	870	6.0
District of Columbia	15,783	5.0	4.7	789	736	-6.7
Mississippi	16,457	5.0	5.1	832	840	2.1
Nebraska	16,525	5.0	5.2	826	855	3.5
West Virginia	16,904	5.0	4.8	845	811	-4.0
New Hampshire	17,117	5.0	4.1	856	704	-17.8
Arkansas	17,428	4.5	4.4	784	773	-1.4
South Carolina	19,841	4.4	4.1	794	817	2.9
Oklahoma	22,453	4.0	4.3	898	956	6.5
Kansas	22,831	4.0	4.3	914	1,015	11.0
Kentucky	25,987	3.0	3.2	780	844	8.2
Oregon	26,850	3.0	3.1	806	826	2.5
Louisiana	26,999	3.0	2.9	810	770	-4.9
Alabama	29,840	3.0	3.2	895	942	5.2
Iowa	30,609	3.0	3.3	918	1,004	9.4
Arizona	32,782	3.0	2.9	983	968	-1.5
Colorado	33,730	3.0	3.2	1,012	1,067	5.4
Tennessee	40,811	2.0	2.0	816	803	-1.6
Minnesota	42,501	2.0	2.0	850	848	0.2
Washington	45,332	2.0	1.9	905	867	-4.2
Georgia	64,672	2.0	2.1	933	999	7.1
Maryland	48,769	2.0	2.0	975	974	-0.1
Connecticut	49,236	1.5	1.5	739	736	-0.4
Indiana	50,034	1.5	1.5	751	772	2.8
Wisconsin	50,647	1.5	1.5	760	744	-2.1
Missouri	50,906	1.5	1.4	764	721	-5.6
North Carolina	51,196	1.5	1.5	768	756	-1.6
Virginia	57,279	1.5	1.5	859	868	1.0
New Jersey	88,966	1.0	1.0	890	842	-5.4
Michigan	94,413	1.0	0.9	944	832	-11.9
Florida	103,736	1.0	1.0	1,037	1,047	1.0
Texas	111,174	1.0	1.0	1,112	1,095	-1.5
Ohio	111,920	1.0	1.0	1,119	1,124	0.4
Massachusetts	113,816	1.0	1.0	1,138	1,085	-4.7
Illinois	122,064	1.0	1.0	1,221	1,188	-2.7
Pennsylvania	168,551	1.0	1.0	1,686	1,689	0.2
California	206,358	1.0	0.9	2,064	1,931	-6.4
New York	219,898	1.0	1.0	2,199	2,124	-3.4
Total	2,362,138			45,627	45,290	-0.8

*In percent

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 and an additional unduplicating process was undertaken when questionnaire responses demonstrated duplication.

Statistical Techniques

A. Weighting

The probability sample design for the national sample survey of registered nurses 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, to the extent feasible, for the potentially biasing effect of nonresponse. The weight 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 was 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 licenses in the same priority State had an equal probability of being selected within that State and, consequently, all sampled nurses from that State had equal weights. The sum of the weights for all respondents assigned to a State was equal, approximately, to the State's total number of active licenses at the time the sample was drawn less the number of those licenses assigned to higher priority States.

Weights are assigned to nurses sequentially. Thus, all nurses licensed in the highest priority State (i.e., Wyoming) are assigned weights first. These weights are derived by dividing Wyoming's total number of active licenses by the number of responses from nurses licensed in Wyoming. For the next highest State (i.e., Alaska), the sum of the weights of the respondents who are licensed in both Alaska and Wyoming having already been assigned a weight in the State of Wyoming are subtracted from Alaska's total number of active licenses and the number of respondents who are licensed in both Alaska and Wyoming are subtracted from the total respondents with licenses in Alaska before the weight is derived for the remaining Alaska licensees. This procedure is continued down the priority order of the States with the accounting for all those with licenses in higher priority States before deriving the weight for the remaining nurses licensed in that State.

B. 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 an 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 careful coding and processing of the sample data.

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 was expended to obtain a high response rate through such things as respondent motivation and followup procedures. A high response rate will reduce both random and systematic errors.

Sample survey results, however, 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. However, it does not permit unbiased estimation of the variability of survey estimates unless some assumptions are made. Estimates of standard errors were 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. 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 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 of respondents (where, for example, 25 is considered a small number) have estimated standard errors that themselves are subject to high variability. Thus, inferences based on such estimates should be guarded.

Direct Variance Estimation

Using the jackknife variance estimation procedure with 20 replicates, each replicate is based on the use of 19 pairs of alpha-segments and one alpha-segment from the 20th pair (there are 40 alpha-segments altogether). The respondents in the included segments for a particular replicate constituted a sample approximately 39/40th's as large as the full sample. A set of weights was computed for the full sample and for each replicate using the established procedure but, for the replicates, employing only the responding nurses from the 39 segments associated with each replicate.

In order to obtain variance estimates via the jackknife approach, 20 additional sets of weights are computed, each based on approximately 39/40th's of the respondents. This permits the construction of 20 replicate estimates, each based on about 39/40th's of the data, to compare with the estimate produced from all of the data. To compute the variance suppose

$\hat{x}_i$ = an estimated total for replicate i associated with alpha-segment pair i;

$\hat{x}$ = an estimated total obtained over the full sample.

The variance of $\hat{x}$, VAR $(\hat{x})$, can be 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:

$$\mathrm{Var}\,(\hat{Y}/\hat{X}) = \sum_{i=1}^{20} (\hat{Y}_i/\hat{X}_i - \hat{Y}/\hat{X})^2 .$$

Following the example mentioned previously,

$\hat{Y}$ and $\hat{Y}_i$ = full sample and replicate estimates, respectively, of the total number of RNs resident in Florida who are 25 to 29 years old, and

$\hat{X}$ and $\hat{X}_i$ = the corresponding estimates of the total number of RNs resident in Florida.

The variance of any other statistic, simple or complex, is similarly estimated by computing the statistic for each replicate.

Another point of consideration was whether to use the full sample estimate in the variance estimate computation or to use the average of the replicate estimates. There is usually little difference between the two. RTI used the estimator full sample estimate, which tends to provide more conservative estimates of variance.

Direct estimates of the variance were computed for a variety of different variables. Two considerations entered into the choice of the variables for which direct variances were developed. For one, variables were chosen in relation to their importance. In addition, others were chosen to represent the gamut of expected design effects. As indicated in the next section, the average of these design effects (on a State-by-State basis) provides the basis for estimating the variance for variables not included in the set where direct variance estimates were computed. Direct estimates of the variance, or the standard error (the square of the variance), for a selected set of national estimates are included in Table B2. Table B3 shows the standard errors of estimates of the nurse population located in each State.

Generalized Variances

The design effect, D, for an estimated proportion is determined by taking the ratio of the estimated sampling variance, obtained

Table B2. Standard errors (S.E.) of selected statistics (numbers and percentages) for U.S. registered nurse population

Description	Estimated Percent	S.E. of estimated Number	Estimated Percent	S.E. of estimated Number
<u>Basic Nursing Education</u>				
Diploma	989,941	7,512	48.69	0.40
Associate Degree	576,169	9,809	28.34	0.43
Baccalaureate Degree	451,985	6,352	22.23	0.32
Master's Degree	1,804	335	0.09	0.02
Doctorate	114	114	0.01	0.01
Unknown	13,021	1,401	0.64	0.07
<u>Employed in Nursing</u>				
Employed in Nursing	1,627,035	10,313	80.03	0.41
Not employed in nursing	405,997	8,171	19.97	0.41
<u>Racial/Ethnic Background</u>				
Hispanic	26,163	4,27	31.29	0.21
American Indian/Alaskan native	8,358	1,194	0.41	0.06
Asian/Pacific islander	46,691	10,993	2.30	0.54
Black	73,647	7,938	3.62	0.39
White	1,864,157	14,102	91.69	0.73
Unknown	14,016	1,092	0.69	0.05
<u>Employment Status in 1988</u>				
Employed in nursing full-time	1,099,576	12,424	54.09	0.53
Employed in nursing part-time	526,489	5,444	25.90	0.30
Employed in nursing full-or part-time	970	280	0.05	0.01
Not employed in nursing	405,997	8,171	19.97	0.41
<u>Field of Employment</u>				
Hospital	1,104,978	11,852	67.91	0.44
Nursing Home Ext. Care	107,805	3,280	6.63	0.19
Nursing Education	30,004	1,711	1.84	0.10
Public Health/Community Health	110,886	3,047	6.82	0.20
Student Health	47,791	2,824	2.94	0.17
Occupational Health	21,856	1,657	1.34	0.10
Ambulatory Care	125,813	2,804	7.73	0.19
Private Duty	19,988	1,290	1.23	0.08
Self Employment	13,203	833	0.81	0.05
Other	3,321	2,288	2.66	0.14
Unknown	1,386	395	0.09	0.02

Table B2. Standard errors (S.E.) of selected statistics (numbers and percentages) for U.S. registered nurse population (continuted)

Description	Estimated Percent	S.E. of estimated Number	Estimated Percent	S.E. of estimated Number
Type of Postion				
Administrator/Assistant Administrator	98,382	2,359	6.05	0.14
Consultant	17,625	1,233	1.08	0.08
Supervisor	91,538	3,706	5.63	0.22
Instructor	62,557	2,738	3.84	0.17
Head Nurse of Assistant	85,911	2,573	5.28	0.15
Staff or Gen. Duty	1,087,878	9,969	66.86	0.40
Practitioner/Midwife	23,535	1,147	1.45	0.09
Clinical Specialist	28,975	1,732	1.78	0.11
Nurse Clinician	17,628	1,198	1.08	0.07
Certified Registered Nurse Anesthetist	16,831	1,279	1.03	0.08
Research	4,783	771	0.29	0.05
Private Duty	19,988	1,290	1.23	0.08
Other	65,031	2,185	3.40	0.13
Unknown	6,369	983	0.39	0.06
Highest Nursing Education				
Diploma	821,201	7,546	40.39	0.38
Associate Degree	511,763	9238	25.17	0.31
Baccalaureate	556,966	7,217	27.40	0.36
Master's	124,981	3,702	6.12	0.19
Doctorate	5,415	683	0.27	0.03
Unknown	12,705	1,421	0.62	0.07
Age of Nurse				
Less than 25 Yrs of Age	78,555	2,694	3.86	0.13
25 to 29 Years of Age	237,578	6,307	11.69	0.31
30 to 34 Years of Age	368,982	6,261	18.15	0.30
35 to 39 Years of Age	332,159	5,375	16.34	0.26
40 to 44 Years of Age	267,578	4,335	13.16	0.21
45 to 49 Years of Age	212,846	3,876	10.47	0.19
50 to 54 Years of Age	173,828	2,754	8.55	0.14
55 to 59 Years of Age	140,190	4,503	6.90	0.22
60 to 64 Years of Age	109,019	3,218	5.39	0.16
65 to More Years of Age	101,064	3,115	4.97	0.15
Unknown	11,232	1,325	0.55	0.07
Mean Salary	28,383	102		

Table B3. Direct estimates of standard error (SE) and coefficient of variation (CV), and registered nurse population by state, 1988

State	1988 Estimated State Nurse Population	SE	CV (in percent)
United State	2,033,032	5,611	0.28%
Alabama	26,763	363	1.36
Alaska	4,243	209	4.93
Arizona	29,860	297	0.99
Arkansas	14,394	282	1.96
California	191,947	1,217	0.63
Colorado	28,917	418	1.45
Connecticut	39,550	789	1.99
Delaware	6,860	408	5.95
D. C.	10,928	506	4.63
Florida	102,470	1,360	1.33
Georgia	41,873	703	1.68
Hawaii	7,024	303	4.31
Idaho	6,329	302	4.77
Illinois	104,697	1,596	1.52
Indiana	43,203	734	1.70
Iowa	27,472	381	1.39
Kansas	20,247	388	1.92
Kentucky	23,279	556	2.39
Louisiana	23,625	397	1.68
Maine	12,317	286	2.32
Maryland	41,182	872	2.12
Massachusetts	87,694	1,352	1.54
Michigan	79,330	864	1.09
Minnesota	40,116	409	1.02
Mississippi	14,252	365	2.56
Missouri	45,102	598	1.33
Montana	6,748	140	2.07
Nebraska	13,536	263	1.94
Nevada	7,677	547	7.13
New Hampshire	13,525	504	3.73
New Jersey	73,321	1,331	1.82
New Mexico	9,180	391	4.26
New York	178,912	2,012	1.12
North Carolina	47,467	899	1.89
North Dakota	6,752	188	2.78
Ohio	97,258	888	0.91
Oklahoma	18,851	460	2.44
Oregon	23,477	428	1.82
Pennsylvania	131,357	1,633	1.24
Rhode Island	11,156	302	2.71
South Carolina	19,249	527	2.74
South Dakota	6,397	180	2.81
Tennessee	33,631	397	1.18
Texas	100,874	1,369	1.36
Utah	9,294	311	3.35
Vermont	5,837	280	4.80
Virginia	45,865	966	2.11
Washington	40,671	693	1.70
West Virginia	13,704	437	3.19
Wisconsin	41,402	702	1.70
Wyoming	3,031	112	3.70

by the jackknife method, to the sampling variance of an estimated proportion for a simple random sample of the same size. For the percentage p this is given by

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

n = the unweighted number used to determine the denominator of $\hat{p}$

Direct estimates of the design effect have been obtained, by the jackknife method, for a judgmentally selected set of the important variances for each State. The average of the design effects, F, obtained for each State and the Nation may 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 formula:

$$\hat{\sigma}_{\hat{Y}/\hat{X}} = [F \cdot (\hat{Y}/\hat{X}) \cdot (1-\hat{Y}/\hat{X})/n]^{\frac{1}{2}} \qquad (1)$$

n = the number of 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 B4 contains the list of average design effects derived for each of the States and for the United States.

It is also possible to make generalized estimates of standard errors for estimated numbers (totals) of RN's in a State (or the nation) with a particular characteristic, where

$\hat{Y}$ = the subtotal of the estimate $\hat{X}$, and

$\hat{X}$ = the estimated total of RN's working and/or living in the State (or nation).

The standard error and coefficient of variation of $\hat{X}$ are shown in Table B3 for each State and the United States. The following exposition is made simpler by defining the rel-variance of an estimate as the square of its coefficient of variation.

Table B4. Average design effects (F) for percentages estimated from the National Survey of Registered Nurses by State, 1988*

State	Average design effect (F)	State	Average design effect (F)
United States	1.98%	Missouri	1.11
Alabama	1.04	Montana	0.96
Alaska	1.31	Nebraska	1.00
Arizona	1.24	Nevada	1.67
Arkansas	1.11	New Hampshire	0.99
California	1.24	New Jersey	1.00
Colorado	1.11	New Mexico	1.37
Connecticut	1.15	New York	1.14
Delaware	1.34	North Carolina	1.02
D. C.	1.21	North Dakota	1.20
Florida	1.22	Ohio	1.06
Georgia	0.96	Oklahoma	1.12
Hawaii	1.01	Oregon	1.05
Idaho	1.43	Pennsylvania	1.11
Illinois	1.03	Rhode Island	1.07
Indiana	1.03	South Carolina	1.22
Iowa	1.00	South Dakota	1.03
Kansas	1.08	Tennessee	0.98
Kentucky	1.12	Texas	1.15
Louisiana	1.05	Utah	1.23
Maine	1.06	Vermont	1.39
Maryland	1.00	Virginia	0.99
Massachusetts	1.18	Washington	1.12
Michigan	1.10	West Virginia	0.97
Minnesota	1.13	Wisconsin	1.07
Mississippi	1.10	Wyoming	1.09

*These design effects represent the approximate average of the ratio of the sampling variance of a percentage estimated from the Registered Nurses survey to the corresponding sampling variance of a simple random sample with the same number of respondents. These design effects apply for all variables except those involving data on race or ethnicity.

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

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

F = the average design effect for the State of interest, and

n = the number of respondents weighted to obtain the estimate $\hat{X}$.

The rel-variance of $\hat{Y}$, denoted $V^2_{\hat{Y}}$, is approximated using

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

$C.\ V._{\hat{X}}$ = the coefficient of variation of the respective State or the United States.

Finally, the standard error of $\hat{Y}$, $\sigma_{\hat{Y}}$, is thus estimated as

$$\sigma_{\hat{Y}} = \hat{Y}\,(\hat{V}^2_{Y})^{1/2} \qquad (2)$$

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 comprising that particular region. The estimated proportion (or percentage) for the region is

$$\hat{Y}_R/\hat{X}_R = \sum_{s=1}^{h} \hat{Y}_s \Big/ \sum_{s=1}^{h} \hat{X}_s \qquad \text{where}$$

h = the number of States in region (R), and

$\hat{Y}_s$ and $\hat{X}_s$ = estimates for a particular State.

The formula used to approximate the standard error of an estimated proportion for a region is

$$\hat{\sigma}_{\hat{Y}_R/\hat{X}_R} = \left[\sum_{s=1}^{h} \left(X_s^2 \hat{\sigma}^2_{\hat{Y}_s/\hat{X}_s} \right) / \left(\sum_{s=1}^{h} \hat{X}_s \right) \right]^{1/2} \qquad (3)$$

where

$\hat{\sigma}_{\hat{Y}_s/\hat{X}_s}$ = the standard error of the estimated proportion $\hat{Y}_s/\hat{X}_s$ computed from direct procedures or Equation (1)

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

$$\hat{\sigma}_{Y_R} = \left(\sum_{s=1}^{h} \hat{\sigma}^2_{Y_s} \right)^{1/2} \qquad (4)$$

where

$\hat{\sigma}_{Y_s}$ = the standard error of the estimated number $\hat{Y}_s$ computed either from direct procedures or from Equation (2).

Chart B1

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:

   $p = 56.5\%$ $F = 1.14$ $n = 1,408$

   $$\hat{\sigma} = [1.14(.565)(.435)/1408]^{\frac{1}{2}} = .014 \text{ or } 1.4\%$$

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

   $p = 67.9\%$ $F = 1.98$ $n = 27,026$

   $$\hat{\sigma} = [1.98(.679)(.321)/27026]^{\frac{1}{2}} = .0040 \text{ or } .40\%$$

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:

   $\hat{Y} = 36,013$ $\hat{X} = 178,912$ $\hat{Y}/\hat{X} = .201$ $n=1,408$ $C.V._{\hat{X}} = 1.12\%$
   $F = 1.14$

   $$\hat{V}^2_{\hat{Y}} = [(1.14)(.799)/1408(.201)] + (.0112)^2 = .0033$$

   $$\hat{\sigma}_{\hat{Y}} = 36013\ (.0033)^{\frac{1}{2}} = 2,069$$

Chart B1(continued)

Illustrative Examples of Generalized Variance Estimates

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

$\hat{Y} = 1,627,035$ $\hat{X} = 2,033,032$ $n = 33,047$

$C.V._{\hat{X}} = .0028$ $\hat{Y}/\hat{X} = .800$ $F = 1.98$

$$\hat{V}^2_Y = [1.98(.200)/33047(.800)] + (.0028)^2 = .000023$$

$$\hat{\sigma}_Y = 1627035(.000023)^{1/2} = 7,803$$

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

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

$$Y / X = 76.6\%$$

$$\hat{\sigma}_{\hat{Y}/\hat{X}_{NJ}} = [1.00(.726)(.274)/613]^{1/2} = .018 \text{ or } 1.8\%$$

$$\hat{\sigma}_{Y/X_{NY}} = [1.14(.799)(.201)/1408]^{1/2} = .011 \text{ or } 1.1\%$$

$$\hat{\sigma}_{\hat{Y}/X_{PA}} = [1.11(.745)(.255)/1238]^{1/2} = .013 \text{ or } 1.3\%$$

$$\hat{\sigma}_{\hat{Y}/X_R} = ([(73321)^2 (.018)^2 +(178912)^2 (.011)^2 +(131357)^2 (.013)^2] / (73321+178912+131357)^2)^{1/2} = .0076 \text{ or } .76\%$$

Final Comments

The design effect results reflect the clustering intrinsic to the selection of nurses in groups of alphabetically neighboring names. Variance increasing effects due to clustering occur to the extent that nurses with similar names tend to have similar characteristics, e.g., because they may belong to the same ethnic group. Naturally, these effects are more pronounced for survey measures related to ethnic and racial characteristics of sample nurses.

The clustering effects on the sampling variance increase both with the homogeneity of the clusters--the alphabetic segments or portions in this case--and with the number of units (i.e., nurses) sampled from each cluster. These effects remained relatively large for race/ethnicity variables. Nevertheless, average design effects were generally reduced, compared to previous RN sample surveys, by virtue of the revised alpha-segments constructed for this survey. Segments of about equal size reduced the unequal weighting effects on the variance of survey estimates.

The sampling error estimates do not take into account nonsampling errors. One type of nonsampling error is due to discrepancies between the target population and the frame population. Although the former population is the one about which inferences are desired, the latter population is the one available in the sampling frame. These discrepancies, which may be referred to as coverage errors, are in this case mostly due to the dynamic nature of the population and to the unavoidable time period that elapsed between frame construction and data collection. Other frame inefficiencies and errors that exist in the lists and files provided by States also give rise to such errors.

Another type of nonsampling error is nonresponse bias. In spite of all efforts to enhance response rates, some nonresponse inevitably remains in any large survey. In the RN-IV survey, there was some evidence that nonresponse rates are higher for non-white ethnic groups. This nonresponse bias is thus compounded with the high sampling variability of estimated numbers and percents for Blacks, Hispanics and Asians, for example, negatively affecting the overall accuracy of these estimates. These comments about the racial/ethnic data need to be kept in mind when reviewing the characteristics of the registered nurse population according to race or ethnicity.