

## APPENDIX B. Survey Methodology

The data for this study were collected from a sample of all those holding current licenses to practice as registered nurses. The design for the study was initially developed in 1975-76 under a contract the Division of Nursing, Bureau of Health Professions, Health Resources Administration had with Westat, Inc.

The survey design had to take into account a number of aspects about the registered nurse population. For one, there is no overall, up-to-date list of registered nurses in the country. There are only those lists maintained by each of the State Boards of Nursing. Thus, there are 51 separate, nonuniform, sampling frames from which the sample is to be drawn. Moreover, to add to the complexities of the development of the sample, registered nurses may hold licenses in more than one State and, therefore, there are duplications in individuals among the 51 State lists. In addition, a number of registered nurses are not licensed in their State of residence and/or State of employment although they are licensed somewhere in the country. Finally, two other attributes of this particular group have to be considered. One, registered nurses are predominantly female and are subject to name changes. Second, as a predominantly employee rather than self-employed group, they are fairly mobile.

Other considerations that were necessary to take into account in designing the approach were that the plan was to be applicable to maintaining current data on the registered nurse population and to providing certain State by State characteristics.

All of these points were incorporated in the survey design established by Westat, Inc. A fuller explanation of the design can be found in the complete report of the first national sample survey conducted by the American Nurses' Association under contract to the Division of Nursing <sup>1/</sup> and in the report of this current study, prepared by the contractor, Research Triangle Institute. Following is a brief synopsis of the approach taken.

### Sample

While the initial sampling design was predicated on a replication of the survey annually, the design was still appropriate to the repeat of the survey process three years after the initial study was made. The selection of registered nurses to be included in the sample is based on name. Using a sample of names obtained from the Inventory of Registered Nurses, the entire universe was alphabetized by last name and first name initial and proportionately equally sized alphabetic segments were derived. The "alpha

---

<sup>1/</sup> Roth, Aleda V., et al. 1977 National Sample Survey of Registered Nurses. Stock No. HRP 0900603, National Technical Information Service, Springfield, Va. 1978.

segments" represented clusters of names that were alphabetically adjacent to one another. To increase the reliability of the estimates for both large and small States but maintain the study within a limited budget, nurses in different States are sampled at different rates. Thus, States that are small in nurse population had higher sampling rates (though smaller sample sizes) than larger States. The differential sampling rates were accomplished by varying the number of "alpha segments" selected in each State.

Because of the multiple licenses that a significant proportion of the nurses hold, steps had to be taken to ensure that a single probability of sample selection, and therefore a single weight, could be assigned to each nurse in the sample. The system was devised so that the probability that any given nurse would fall in the sample was equal to the highest sampling rate among the States in which the nurse was licensed. This result was achieved by "nesting" each State's sample into the others through the use of an overlapping "alpha segment" procedure. The States having higher sampling rates have the broader alphabetic segment and those with lower rates a smaller proportion of that broader segment.

The final sample selected in each State was placed on a computer tape, merging each of these samples into a single file. An initial unduplication of nurses from State-to-State was then accomplished by removing any duplicates from this file. In total, the aggregated number of names selected for the sample in the 1980 study was 43,986. After the initial unduplication, 39,573 were included in the sample. The data were collected through the use of mail questionnaires.

#### Data Collection Procedures

To ensure adequate response to the study there were three waves of questionnaire mailings to the individuals in the sample, an initial and two follow-ups to those who did not respond to the prior mailing. Finally, there were telephone calls to a sample of those who were still nonrespondents after the mailings were completed. The resulting response rate was as follows:

<u>Total Sample</u>	<u>39,573</u>
<u>Respondents</u>	<u>30,642</u>
Nurses who completed questionnaire	30,535
Deceased nurses	66
Persons ineligible for survey	41
<u>Nonrespondents</u>	<u>7,667</u>
<u>Duplicates</u>	<u>1,264</u>

Calculation of response rate:

$$\frac{30,642}{39,573 - 1,264} = .800 \text{ or } 80.0 \text{ percent}$$

The formula to use to approximate the standard error of an estimated percent of registered nurses nationally having a given characteristic is:

$$\sigma_{\hat{Y}/\hat{X}} = \sqrt{F \cdot \hat{Y}/\hat{X} \cdot (1-\hat{Y}/\hat{X})/n} \cdot 100$$

where:

$\hat{Y}$  = the estimated number of registered nurses with the given characteristics

$\hat{X}$  = the estimated total number of registered nurses from which  $\hat{Y}$  is drawn

$n$  = the actual number of respondents from which  $\hat{X}$  is derived

Thus, the estimated standard error of the percentage of registered nurses who are employed in nursing in the United States would be calculated from the data in Table 1 as follows:

$$(1) 100 \cdot \frac{1272851}{1662382} = 76.6\%$$

$$(2) 100 \sqrt{(1.4)(.766)(1-.766)/30375} = .29\%$$

In about 95 out of 100 chances the true percent of registered nurses in the United States who are employed in nursing would be 76.0% - 77.1%.

To determine the approximate standard error of the estimated number of registered nurses in the United States as a whole with a given characteristic, the following formula would be used:

$$\sigma_{\hat{Y}} = \hat{Y} \sqrt{\frac{F(1-\hat{Y}/\hat{X}) + (C.V.)^2}{n \cdot \hat{Y}/\hat{X}}}$$

The approximate standard error of the estimated number of employed registered nurses is:

$$1,272,851 \sqrt{\frac{1.4(1-.766) + (.0005)^2}{30375 (.766)}} = 4818$$

Therefore, in about 95 out of 100 chances, the actual number of registered nurses employed in nursing in the United States in November 1980 was 1,263,407 - 1,282,295.

The approximation of the standard error of a percentage or a number pertaining to the characteristics of the nurses located within a State would be derived in the same manner as indicated for the national estimates. In this case, the F factors and C.V.s derived for the particular States would be used. These appear in Table B-1.

The standard error of an estimated percentage for a grouping of states would be given by a linear combination of the constituent States:

$$\sigma_{\hat{Y}_R/\hat{X}_R} = 100 \sqrt{\frac{h}{\sum_{s=1}^h (\hat{X}_s^2 \sigma_{\hat{Y}_s/\hat{X}_s}^2)} \cdot \frac{1}{(\sum \hat{X}_s)^2}},$$

Similarly, the formula used to estimate the standard error of an estimated number for a grouping of States is based on the standard errors computed for the constituent States:

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

With regard to the sampling variability, care should be exercised in interpreting data based on small samples such as where the totals are based on less than 25 in the sample or where a percentage is based on less than 10 in the sample out of a sample count of more than 25.

The formulae included here indicate an approach to be used to approximate the standard errors for much of the data in this report. For a much fuller discussion of the reliability of the estimates drawn from the study and the approaches to deriving standard errors of the estimates, however, the reader is urged to consult the full report of the 1977 study 3/ and the contractor's report for the 1980 study.

---

3/ Roth, Op. Cit.

Table B-1. Average design effects (F) and coefficients of variation (C.V.) of the registered nurse population in the United States and each State

State	Average Design Effect F	Coefficient of Variation C.V.
United States.....	1.4	.0005
Alabama.....	1.0	.0227
Alaska.....	1.4	.0683
Arizona.....	1.1	.0242
Arkansas.....	1.0	.0324
California.....	1.1	.0098
Colorado.....	1.1	.0241
Connecticut.....	1.1	.0218
Delaware.....	1.3	.0577
District of Columbia.....	1.1	.0567
Florida.....	1.1	.0188
Georgia.....	1.0	.0228
Hawaii.....	1.2	.0414
Idaho.....	1.0	.0380
Illinois.....	1.0	.0131
Indiana.....	1.0	.0203
Iowa.....	1.0	.0168
Kansas.....	1.0	.0223
Kentucky.....	1.0	.0243
Louisiana.....	1.0	.0224
Maine.....	1.1	.0359
Maryland.....	1.0	.0254
Massachusetts.....	1.1	.0159
Michigan.....	1.0	.0145
Minnesota.....	1.0	.0125
Mississippi.....	1.1	.0323
Missouri.....	1.0	.0189
Montana.....	1.1	.0306
Nebraska.....	1.2	.0248
Nevada.....	1.0	.0304
New Hampshire.....	1.1	.0402
New Jersey.....	1.0	.0181
New Mexico.....	1.0	.0453
New York.....	1.1	.0109
North Carolina.....	1.0	.0195
North Dakota.....	1.0	.0327
Ohio.....	1.0	.0122
Oklahoma.....	1.0	.0204
Oregon.....	1.0	.0184
Pennsylvania.....	1.1	.0139
Rhode Island.....	1.0	.0334
South Carolina.....	1.2	.0294
South Dakota.....	1.1	.0357
Tennessee.....	1.0	.0235
Texas.....	1.1	.0154
Utah.....	1.0	.0302
Vermont.....	1.2	.0521
Virginia.....	1.0	.0225
Washington.....	1.1	.0170
West Virginia.....	1.0	.0339
Wisconsin.....	1.0	.0149
Wyoming.....	1.2	.0575