

## APPENDIX B. SURVEY METHODOLOGY

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. A further consideration in designing the approach was that the plan had to be applicable to both maintaining current data on the registered nurse population in the country as a whole and to providing certain State by State nurse 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 this survey.<sup>1</sup> The following provides a brief synopsis of the approach taken.

### Sample

The selection of registered nurses to be included in the sample is based on name. Using a sample of names obtained from the 1972 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 segments" represented clusters of names that were alphabetically adjacent to one another. Those "alpha segments" to be included in the survey are then subdivided into smaller portions so that parts of an "alpha segment" can be chosen to be used in sampling from a particular State. 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 have higher sampling rates (though

---

<sup>1</sup>/ Westat, Inc. 1984 National Sample Survey of Registered Nurses, Summary of Results, National Technical Information Services, Springfield, VA, NTIS Accession No. HRP0906851.

smaller sample sizes) than larger States. The different sampling rates are accomplished by varying the size of the portions of the "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 a broader portion of the "alpha segment" and those with lower rates a smaller proportion of that broader portion.

The final sample selected in each State is merged with those from the other States, into a single computer file from which an initial unduplication of nurses from State-to-State is accomplished by removing any duplications from this file. In the 1984 study, the total number of names selected was 44,268. After the initial unduplication, 41,932 were included in the sample.

#### Data Collection Procedures

The data are collected through the use of mail questionnaires. Efforts are made through multiple mailings and the use of telephone follow-ups to ensure an adequate response. The 1984 study had three waves of questionnaire mailings to the individuals in the sample. The final questionnaire mailing was preceded by a mailgram. 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 names selected for Sample</u>	<u>44,268</u>
<u>Respondents</u>	<u>32,100</u>
Completed questionnaires (mail and telephone)	31,822
Deceased nurses	91
Persons ineligible for survey	187
<u>Nonrespondents</u>	<u>8,109</u>
<u>Duplicates</u>	<u>4,059</u>

Calculation of response rate:

$$\frac{32,100}{44,268 - 4,059} = .798 = 79.8 \text{ percent}$$

## Estimation Procedure

To provide estimates of the number of those in the registered nurse population with a certain given characteristic each respondent to the survey is assigned a weight so that the respondent represents a number of nurses who have active licenses to practice in the country as of the survey date. The computation of the weights is based on the number of respondents in relation to the State licensure counts. Individual respondents are provided with the weight for the State to which assigned according to the "priority" order of the States. The State with the largest sampling factor has first priority and the one with the smallest is listed last. Thus, the State of assignment is the one for which the nurse would have a probability of being selected at the highest rate if the nurse has more than one license and could be selected in more than one State. If the nurse has only one license that is the State of assignment for weighting purposes.<sup>2</sup> The estimates of the numbers of persons with a certain type of characteristic, therefore, are an aggregate of the weights of the individual respondent with that characteristic.

## Reliability of the Estimates

Since the estimates provided in this study are based on a sample rather than a total enumeration they may differ somewhat from the numbers that would be obtained if a complete census was taken using the same questionnaire. In all studies there may be errors related to the completeness of respondent coverage, interpretation of questions, and appropriateness of coding and editing of the data. In sample studies there may also be errors due to sampling. These sampling errors can be approximated through the use of standard errors. The standard errors may also measure the effect of some nonsampling errors of response but can not deal with any systematic bias in the data that might result from nonsampling errors such as those indicated previously.

Westat, Inc. the contractor for the 1984 study, developed estimates of standard errors for a number of important variables or characteristics in the study by use of a "jackknife" variance estimation procedure. In this case, the variance estimates were based upon overlapping replicate samples, each replicate sample derived from pairings of the 40 "alpha segments" used for sampling purposes. Each replicate sample consisted of all observations from 19 of the 20 "alpha segment" pairs and one segment of the remaining pair. Each of the 20 replicate samples was put through the same weighting procedure described earlier for the total sample.

From these direct estimates they also developed design effect factors for use in a more generalized approach to the estimation of the standard errors. The following standard error approximation approaches measure the

---

<sup>2</sup>/ It should be pointed out that the State of license is only used for weighting purposes. Nurses are assigned to the State in which they are actually located when findings are presented in terms of geographic distribution.

sampling variability of the national data included in this report. The chances are about 68 out of 100 that an estimate from the survey differs from that which would be obtained in a complete census by less than the standard error. The chances are about 95 out of 100 that the estimate would differ from the census by less than 1.96 standard errors.

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

$$\sigma_{\hat{Y}/\hat{X}} \approx 100 \cdot \sqrt{\left[F/n\right] \left[\hat{Y}/\hat{X}\right] \left[1 - \hat{Y}/\hat{X}\right]}$$

where:

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

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

$n$  = 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:

$$(1) \quad 100 \cdot \frac{1485725}{1887697} = 78.7\%$$

$$(2) \quad 100 \cdot \sqrt{\frac{(2.0)(0.787)(1-0.787)}{31626}} = 0.33\%$$

In about 95 out of 100 chances the true percent of registered nurses in the United States who are employed in nursing would be 78.0% - 79.3%.

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

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

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

$$1485725 \cdot \sqrt{\frac{2.0(1-0.787)}{31626(0.787)} + (0.0024)^2} = 7106$$

Therefore, in about 95 out of 100 chances, the actual number of registered nurses employed in nursing in the United States in November 1984 was 1,471,797 - 1,499,653.

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:

$$\sigma_{\hat{Y}_R/\hat{X}_R} \approx \sqrt{\sum_{s=1}^h \left[ \hat{X}_s^2 \sigma_{\hat{Y}_s/\hat{X}_s}^2 \right] \left[ \sum_{s=1}^h \hat{X}_s \right]^2}$$

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:

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

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 contractor's report of the study.<sup>3</sup>

---

<sup>3</sup>/ Westat, Inc., 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.....	2.0 <sup>1</sup>	.0024
Alabama.....	1.1	.0161
Alaska.....	1.3	.0323
Arizona.....	1.0	.0249
Arkansas.....	1.0	.0356
California.....	1.7	.0085
Colorado.....	1.3	.0173
Connecticut.....	1.2	.0225
Delaware.....	1.3	.0479
District of Columbia.....	1.6	.0578
Florida.....	1.0	.0183
Georgia.....	1.3	.0226
Hawaii.....	1.1 <sup>1</sup>	.0493
Idaho.....	1.1	.0282
Illinois.....	1.2	.0114
Indiana.....	1.0	.0189
Iowa.....	1.0	.0161
Kansas.....	1.0	.0239
Kentucky.....	1.1	.0213
Louisiana.....	1.2	.0229
Maine.....	1.1	.0224
Maryland.....	1.1	.0232
Massachusetts.....	1.3	.0187
Michigan.....	1.3	.0144
Minnesota.....	1.0	.0097
Mississippi.....	1.0	.0290
Missouri.....	1.0	.0145
Montana.....	1.2	.0324
Nebraska.....	1.0	.0261
Nevada.....	1.8	.0476
New Hampshire.....	1.1	.0393
New Jersey.....	1.0	.0236
New Mexico.....	1.3	.0524
New York.....	1.4	.0081
North Carolina.....	1.3	.0189
North Dakota.....	1.2	.0169
Ohio.....	1.3	.0128
Oklahoma.....	1.1	.0227
Oregon.....	1.0	.0195
Pennsylvania.....	1.2	.0149
Rhode Island.....	1.1	.0273
South Carolina.....	1.2	.0313
South Dakota.....	1.0	.0335
Tennessee.....	1.1	.0232
Texas.....	1.5	.0082
Utah.....	1.1	.0239
Vermont.....	1.2	.0646
Virginia.....	1.0	.0196
Washington.....	1.2	.0196
West Virginia.....	1.1	.0325
Wisconsin.....	1.2	.0169
Wyoming.....	2.3	.0590

<sup>1</sup>These design effects apply for all variables except those involving data on race or ethnicity. For those variables use triple the square root of the average design effect as the F factor.