

APPENDIX B

SURVEY METHODOLOGY

APPENDIX B

SURVEY METHODOLOGY

The eighth cycle of the National Sample Survey of Registered Nurses (NSSRN) followed the same basic sample design as its predecessors. The sample design was originally developed by Westat, Inc. under a contract with the Division of Nursing, BHP, HRSA in 1975-76 and can be best described as a systematic sample of alphabetic clusters of names in each State using a 'nested alpha segment design'. Prior to sampling, each State was ranked by the sampling rate such that the highest priority States were those with the highest sampling rate (for the most part, small States). As a result, the alphabetic clusters of names for lower priority States are 'nested', or included, within those of higher priority States. This means that a sample name selected in one State (such as California) will also have been selected in every State with a higher priority (in the case of California, this is all other States).

This design approach takes 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. The advantage of the nested alpha-segment design is that one can determine the probabilities of selection and appropriate multiplicity adjusted weights for those nurses that are listed in more than one State. In addition, the design also permits the use of each

sample registered nurses' data for State estimates of each of her/his States of licensure.

This appendix provides a brief summary of the methodology of the NSSRN including the sampling frame, 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.

Sampling Frame

The target population for the eighth NSSRN included all registered nurses with an active license in the United States as of March 2004. A sampling frame was required to select a probability sample of nurses from which valid inferences could be made to the target population. The sampling frame for the eighth NSSRN consisted of all registered nurses who are currently eligible to practice as an RN in the U.S. This sampling frame included RNs who have received a specialty license or have been certified by a State agency as an advanced practiced nurse (APNs) such as nurse practitioner, certified nurse midwives, certified registered nurse anesthetist, or clinical nursing specialist and excluded licensed practical nurses (LPNs)/licensed vocational nurses (LVNs).

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 holding an active license in that State. 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 diskette or compact disk (twenty States), or electronically as an attachment to an e-mail message (twenty-seven States). Three States sent the data via FTP and another provided the data on their website. For this study, States were also asked to identify nurses for whom the State provided advanced practice nurse (APN) status. In some cases, the State identified these nurses on the basic list provided. However, some APNs were identified on separate lists and their APN status was appended to the information on the RN sampling frame.

Each of the 51 State files was checked for consistency, names were standardized, and duplicates and ineligible records were removed from the State list to prepare the list for sampling.

Sample Design

The NSSRN 2004, the eighth in the series, continued to oversample nurses in small States in order to better support HRSA's National Center for Health Workforce Analysis' State level supply and demand projections for registered nurses. The basic design was enhanced by using sample design optimization methodology developed by Chromy¹ to determine the sample allocation to the States that would simultaneously satisfy variance constraints defined by the 51 States and the total U.S.

In the original sample design, and in the 1988 redesign, the universe of RNs was sorted alphabetically by last name and approximately equal-sized clusters of RNs were constructed by partitioning the alphabetically ordered list into 250 alpha-segment clusters with equal (or nearly equal)

numbers of RNs. An alpha-segment was defined as all alphabetically adjacent names falling within pre-specified boundaries. For example, all names beginning with the lower boundary, up to but not including the name that defined the upper boundary.

From the frame of 250 equally divided alpha-segments, a total of 40 alpha-segments were randomly selected, representing a 16 percent sampling rate overall. Registered nurses are selected in the sample based on their name, with an RN being included in the sample if the name of licensure falls into one of the alphabetic segments that are in sample for that State.

Although each State had 40 sample segments, the sample size of each State differed in size depending on the State's sampling rate. 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. Since both national and State-level estimates are required for the 2004 NSSRN, as was done in prior surveys, sampling rates were increased in the smaller States to obtain larger State-level sample sizes. 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.

To accommodate the differing State sampling rates, a planned variation in the size of the segments, i.e., "portions of alpha segments" was used. Each of the 40 alpha-segments selected for sample was divided into $\frac{1}{2}$ -, $\frac{1}{4}$ -, $\frac{1}{8}$ -, $\frac{1}{16}$ -, and $\frac{1}{32}$ - portions. These fractions indicate the size of the alpha segment portion relative to the size of the basic alpha-segment.

The sampling rate for a particular State was achieved using a combination of the alpha-segment portions. As a result, each State contains some sample (i.e., a portion) from each of the 40 alpha-segments, depending on the sampling rate for the State. For example, selecting the entire 40 complete alpha segments on a State list is expected to constitute a 16 percent sampling rate ($40 \div 250 = 0.16$) in the State. This is because each alpha segment contained an expected 0.4% of the State's

¹ Chromy, James R. "Design Optimization with Multiple Objectives". American Statistical Association of the Section on Survey Research Methods, Arlington, VA., pp A4-199

RN names (40×0.4 percent = 16 percent). Likewise, the sample for a State with an 8 percent sampling rate consisted of the 40 $\frac{1}{2}$ portion selections. Several sampling rates use a combination of portions for each alpha-segment in sample (rather than one fractional portion for all alpha-segments). For example, 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 $\frac{1}{4}$ portions from the first group and the $\frac{1}{2}$ portions from the second group (0.4 percent $\times [(30 \times \frac{1}{4}) + (10 \times \frac{1}{2})] = 5$ percent).

To identify and account for nurses appearing in more than one of the 51 State lists, the portions were constructed such that each portion was “nested” (or included) in the boundaries of the larger portion. As a result, the alpha segment clusters from the States with lower sampling rates (typically larger States) were automatically included in the alpha segment clusters selected from the States with higher sampling rates (typically smaller States).

As a result, a RN who was licensed under the same name in two States *with identical sampling rates* was selected (or not selected) for both States, since the alphabetic name boundaries defining the portions are the same for both States. However, if the RN was licensed under the same name in two States that are sampled *with different sampling rates*, then, if the RN was sampled in the State with a lower sampling rate, they were also included in the sample for the State with the higher sampling rate (as the alphabetic name boundaries defining the portions for the State with the lower sampling rate are nested within those of the State with the higher sampling rate). This nesting property of the sample design maximizes the chances that the RN will be selected in all States that they have an active license in. A nurse that is licensed in two or more States under the same name will have a probability of selection corresponding to the State with the highest sampling rate.

Sample design optimization techniques developed by Chromy (1996) were used to determine how to allocate the sample of 54,000 RNs to the 51 State lists. This sample size was then converted to a

sampling rate, and the rate was rounded to one of the admissible rates for the nesting design. For example, the original rate for the State of Washington was 1.59%, the closest admissible rate was 1.5%. Rates were rounded down only such that the change in sampling rate still left their effective sample size at or above the 1996 NSSRN level.

After determination of frame sizes and expected sampling rates, the States were assigned a priority order to properly determine selection probabilities for nurses appearing on more than one of the 51 State lists. Traditionally, States were ordered by size, with larger States having lower sampling rates and smaller States having higher sampling rates. However, as in the 2000 NSSRN, States were priority ordered based on their sampling rate. As such, it is mostly, but not necessarily, the case that States with larger RN populations had lower sampling rates.

Essentially the same procedure was followed for sample selection for all States. Once a State provided a licensure file containing all appropriate names of individuals with active RN licenses and meeting 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”).

Registered nurses were selected 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 alpha-segment portion as defined by the State sampling rate. In other words, the sample for a given State consisted of all RN names falling into any one of the State’s pre-designated 40 alphabetic portions that corresponded to the State sampling rate (one portion from each of the complete 40 alpha-segments in sample).

The pairs of names that defined the alpha-segment portion constituted the lower and upper boundaries corresponding to the sampling rate. Thus, the membership of the alpha-segment portion was

defined by all names, beginning with the lower boundary (i.e., the last name in alphabetical order of all the names included in that segment), *up to but not including* a name that defined the upper boundary. This latter name fell into the next alpha-segment. As was done in the NSSRN 2000, any deviations of more than 8 percent were candidates for either an increased or decreased rate.

Because the survey is longitudinal in nature, a panel structure was constructed to allow for several of the sample alpha-segments to be systematically replaced each survey. Under the original survey design, the 40 sample alpha-segments were arranged in alphabetical order and then partitioned into eight groups of five successive alpha-segments each. One segment from each group was randomly assigned to each panel, so that each panel consisted of segments that spanned the entire alphabet. For each successive survey, a new panel (consisting of eight new alpha-segments or 20 percent of the sample) was entered into the sample, replacing one of the five panels from the previous survey. Under this scheme, a nurse who maintained an active license in the same State(s) could be retained in the sample for up to five surveys.

The planned NSSRN 2004 sample size was 54,000 cases, similar to that of the NSSRN 2000, and up from the 45,000 used in previous studies. Planned sampling rates ranged from 1.125 percent in several of the largest States to 15 percent in Wyoming. This translated into planned sample sizes ranging from 3,225 RNs in California to approximately 796 in Wyoming. The initial round of sampling, however, yielded a much smaller sample than expected due to the variable size of the alpha-segments in each State. Thus, a second round of sampling was done by increasing the sampling rates from 1 to 1.125 in the eleven largest States and “adding to” the sample selected in the first round, yielding a total of 56,917 sample cases. After eliminating cross-State duplications, the expected sample size to be fielded was still approximately 54,000 cases.

Table B-1 in Appendix B 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 the

distribution of nurses’ names. States are priority ordered by sampling rate and size.

Because many nurses are licensed in more than one State, their names could be selected in the sample more than once. In accordance with the sample design, we ensured that each sampled RN was retained in the outgoing sample file exactly once to avoid multiple questionnaires being sent to nurses. If we identified an exact duplicate, the nurse in the lower priority State was coded as a duplicate of the sample member in the higher priority State. For example, an Alaska record was coded as a duplicate to the sample record in Wyoming. Following data collection, these expected duplicates were reviewed to ensure that the nurse reported a license in both of the States.

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

State	Priority Order	Frame Size	Sampling Rate Percentage		Actual Sample Size
			Planned	Actual ²	
TOTAL		3,252,548			56,917
Wyoming	1	5,309	15.00%	15.60%	828
Alaska	2	7,389	13.00%	11.88%	878
Vermont	3	8,728	10.00%	9.53%	832
District of Columbia	4	17,104	10.00%	9.71%	1,661
North Dakota	5	8,139	9.00%	9.74%	793
Delaware	6	10,407	9.00%	8.87%	923
Montana	7	10,885	8.00%	8.15%	887
South Dakota	8	10,773	7.00%	6.88%	741
Idaho	9	12,769	7.00%	6.75%	862
Hawaii	10	13,548	7.00%	7.44%	1,008
Nevada	11	19,201	7.00%	6.25%	1,200
Rhode Island	12	17,203	5.50%	5.37%	923
New Mexico	13	17,544	5.00%	4.98%	874
New Hampshire	14	19,108	5.00%	4.71%	900
Utah	15	19,210	4.50%	4.97%	954
Maine	16	19,869	4.50%	4.50%	894
Nebraska	17	20,100	3.50%	3.56%	716
Arkansas	18	27,878	3.50%	3.52%	982
West Virginia	19	21,295	3.50%	3.13%	667
Mississippi	20	31,734	3.00%	3.13%	994
Oklahoma	21	32,185	3.00%	2.93%	944
Kansas	22	34,047	3.00%	3.10%	1,057
Iowa	23	40,312	2.50%	2.31%	933
South Carolina	24	38,265	2.50%	2.47%	944
Oregon	25	38,453	2.00%	1.95%	750
Louisiana	26	43,299	2.00%	1.75%	757
Colorado	27	48,586	2.00%	2.14%	1,042
Connecticut	28	52,364	2.00%	1.96%	1,025
Alabama	29	46,974	1.75%	1.81%	852
Kentucky	30	47,123	1.75%	1.77%	832
Arizona	31	51,482	1.75%	1.72%	887
Maryland	32	56,922	1.50%	1.47%	835
Washington	33	66,397	1.50%	1.44%	954
Minnesota	34	66,434	1.50%	1.59%	1,056
Wisconsin	35	63,865	1.25%	1.24%	793
Tennessee	36	65,827	1.25%	1.29%	849
Indiana	37	70,488	1.25%	1.23%	867
Missouri	38	74,508	1.25%	1.28%	953
Georgia	39	86,369	1.25%	1.26%	1,086
Virginia	40	85,705	1.25%	1.21%	1,036
North Carolina	41	96,877	1.125%	1.146%	1,110
Massachusetts	42	105,206	1.125%	1.350%	1,420
New Jersey	43	109,726	1.125%	1.067%	1,171

² Since the actual distribution of names differs for each State from the frame distribution used to develop the 250 alpha-segments, some variation occurs between the planned and actual sampling rates.

THE REGISTERED NURSE POPULATION

Sampling Rate Percentage					
State	Priority Order	Frame Size	Planned	Actual ²	Actual Sample Size
Michigan	44	117,360	1.125%	1.161%	1,363
Ohio	45	140,689	1.125%	1.124%	1,581
Illinois	46	154,572	1.125%	1.124%	1,738
Texas	47	176,652	1.125%	1.066%	1,883
Pennsylvania	48	191,628	1.125%	1.037%	1,988
Florida	49	201,113	1.125%	1.086%	2,184
New York	50	244,288	1.125%	1.061%	2,592
California	51	286,639	1.125%	1.018%	2,918

Weighting Procedures

The probability sample design of the survey permits the computation of unbiased estimates of characteristics of the RN population at the National and State level. 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 sampling weight for an RN is the reciprocal of the nurse's probability of selection in her/his priority State, adjusted to account for nonresponse and multiple licenses.

Before computing the weights, the original State frame sizes (shown above) were adjusted to account for duplicate licenses within States and ineligible licenses (i.e., frame errors) found in the sample. Most within-State duplicates were identified at the time of initial list processing, but a few were identified after sample selection. The ineligible licenses were identified in the process of reconciling the State and nurse reported licenses. Some of the inconsistencies between the State reported data and the nurse reported data are due to the time period that elapsed between frame construction and data collection (a period during which changes and license expirations naturally occur). Other differences are due to errors in either the State list or the nurse's questionnaire. Cases that could not be reconciled by Gallup were sent to the State Boards of Nursing for resolution.

In both cases, the frame total is computed by subtracting the estimated number of ineligible and duplicate licenses from the State's original frame count. The adjusted frame total used to compute the resulting weights for State i can be computed as:

$$N'_i = N_i - \hat{D}_i - \hat{E}_i$$

where:

N_i = the total number of licenses on State i list,

\hat{D}_i = the estimated number of within-State duplicates in State i , and

\hat{E}_i = the estimated number of frame errors in State i (e.g., licenses listed by State that were not reported by a responding nurse).

Each responding nurse was assigned a weight corresponding to their unique 'priority State'; that is, *the State with the highest sampling rate from which he or she was licensed and selected into the sample*. In other words, the weight is reflective of the probability of selecting the sampled nurse in their "priority" State. All nurses with the same priority State have an equal probability of being selected and, consequently, have equal initial sampling weights. The sum of the weights for all nurse respondents assigned to a specific priority State will equal, approximately, the total number of active licenses on the list (at the time the sample was drawn) less the number of those licenses assigned to higher priority lists.

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

$N(i)$ = total number of licenses for State i (adjusted for within-State duplicates and frame errors)

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

$n(i,j)$ = number of eligible 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 higher 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 adjusted frame count of licenses for 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:

$$\begin{aligned} W(B) &= [N(B) - W(A) n(A,B)] / m(B) \\ W(C) &= [N(C) - W(A) n(A,C) - W(B) n(B,C)] / m(C) . \end{aligned}$$

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 adjusted frame count of RN licenses 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 both a nonresponse adjustment that inflated the respondents' data to account for those that did not respond to the survey and a duplication adjustment to account for duplication in the sampling frame across States. These final analysis weights will serve to differentially weight responding nurses to reflect the level of disproportionality in the final respondent sample relative to the population.

Estimation Procedure

Final NSSRN 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 a particular State is based on the following indicator variable, X_k :

$$\begin{aligned} X_k &= 1 \text{ if nurse } k \text{ worked in a particular State,} \\ &= 0 \text{ otherwise.} \end{aligned}$$

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. In addition, 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 included careful planning, keeping nonresponses to the lowest feasible level, and coding and processing of the sample data.

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. However, correlations or relationships in cross-tabulations are often decreased by such errors, and sometimes substantially. Thus, random errors that tend to be compensated for 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 efforts 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 of the largest sources 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 by respondent motivation and follow-up procedures. A high response rate reduces both random and systematic nonsampling errors. After taking into account duplicates and frame errors, the overall response rate to this survey was 70.47 percent. State-level response rates ranged from 61.98 percent to 81.57 percent except for the District of Columbia where the response rate (46.12 percent) was significantly lower.

Sampling Errors

All sample survey estimates 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. Thus, 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) the magnitude of the overestimate is believed to be 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 0.95 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 for \hat{x} .

Direct Variance Estimation

Similar to prior cycles of the NSSRN, direct estimates of sampling variance were obtained for a set of important variables for each State and for the United States using the jackknife variance estimation procedure with 20 replicates of the sample. 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).

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.

Each replicate was formed from 19 pairs of alpha-segments (38 alpha-segments total) and 1 alpha-segment from the 20th pair. Alpha-segments were randomly removed from each pair to form the replicate estimates. This procedure was performed 20 times, once for each pair of alpha-segments. Thus, actual respondent count in the included segments for a particular replicate was approximately 39/40^{ths} of the full respondent sample and was weighted to represent the full population.

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

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

where:

\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.

If the estimate of interest is a ratio of two estimated totals (e.g., the total number 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:

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

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, the estimator \hat{x} was used 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. Table B-2 in Appendix B presents direct estimates of the standard error (the square root of the variance) for a selected set of variables. Table B-3 in Appendix B shows the estimated population of nurses in each State 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 \hat{p} is determined by taking the ratio of the estimated sampling variance, $\hat{\sigma}_{\hat{p}}^2$, obtained by the jackknife method, to the sampling variance of the \hat{p} in a simple random sample of the same size. This design effect, F , can be computed as follows:

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

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

Direct estimates of the design effect were computed for a set of variables for each State. The median of the design effects was then computed for each State and the nation. These median design effects can be used in formulas for estimating generalized variances or standard errors. This procedure uses median 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:

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

where n is the number of survey respondents used to determine the estimate \hat{X} . The multiplier F , the median² design effect, depends upon the State for which the estimated proportion was generated. The median design effects are listed on Table B-4 in Appendix B.

Generalized estimates of standard errors can also be computed for estimated numbers (or totals) of RNs in a State with a particular characteristic \hat{Y} , (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. Note that 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 (see Table B-3).

To calculate the standard error of a total, one must first compute the relative variance (or square of the coefficient of variation) of the ratio of \hat{Y} to \hat{X} (called $V_{\hat{Y}/\hat{X}}^2$). The relative variance can be calculated as:

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

where F is the design effect for the State of interest and n is the number of respondents to the survey that were weighted to obtain the estimate \hat{X} .

Then, from the relative variance of the ratio, one can approximate the relative variance of the total \hat{Y} , denoted $V_{\hat{Y}}^2$, by using:

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

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

$$\sigma_{\hat{Y}} = \hat{Y} \sqrt{V_{\hat{Y}}^2} \quad (2)$$

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:

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

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

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

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).

²The 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-2. Estimates and Standard Errors (S.E.) For Selected Variables of U.S. Registered Nurse Population

Description	Estimated Number	S.E. of Estimated Number	Estimated Percent	S.E. of Estimated Percent
UNITED STATES, Total Number Of Nurses	2,909,357	7,000		
Basic Nursing Education				
Diploma Program	733,377	9,749	25.21	0.32
Associate Degree	1,227,256	16,571	42.18	0.54
Baccalaureate Degree	887,114	13,366	30.49	0.47
Master's Degree	14,979	1,412	0.51	0.05
Doctorate	532	271	0.02	0.01
Not Reported	46,098	2,568	1.58	0.09
Employed in Nursing				
Yes	2,421,351	10,124	83.23	0.27
No	488,006	7,792	16.77	0.27
Racial/Ethnic Background				
White (non-hispanic)	2,380,529	28,004	81.82	0.89
Black/African American (non-hispanic)	122,495	16,737	4.21	0.57
Asian (non-hispanic)	84,383	15,540	2.90	0.54
American Indian/Alaskan Native (non-hispanic)	9,453	972	0.32	0.03
Native Hawaiian/Pacific Islander (non-hispanic)	5,594	1,091	0.19	0.04
Two or more races (non-hispanic)	41,244	2,641	1.42	0.09
Hispanic/Latino (White)	38,530	7,745	1.32	0.27
Hispanic/Latino (Black/African American)	2,924	633	0.10	0.02
Hispanic/Latino (Two or more races)	3,096	741	0.11	0.03
Hispanic, Other	3,460	921	0.12	0.03
Not Reported	217,651	5,689	7.48	0.19
Employment Status in 2004				
Employed In Nursing Full Time	1,696,807	12,210	58.32	0.44
Employed In Nursing Part Time	720,283	11,059	24.76	0.35
Employed In Nursing, Full/Part Time Unknown	4,261	523	0.15	0.02
Not Employed In Nursing	488,006	7,793	16.77	0.27
Graduation Year				
Before 1961	150,147	4,332	5.16	0.15
1961 To 1965	146,805	4,047	5.05	0.14
1966 To 1970	203,313	4,150	6.99	0.14
1971 To 1975	300,072	7,685	10.31	0.26
1976 To 1980	378,607	7,543	13.01	0.25
1981 To 1985	385,145	7,064	13.24	0.24
1986 To 1990	321,070	6,472	11.04	0.22
1991 To 1995	406,125	5,902	13.96	0.22
1996 To 2000	367,557	6,094	12.63	0.20

THE REGISTERED NURSE POPULATION

Description	Estimated Number	S.E. of Estimated Number	Estimated Percent	S.E. of Estimated Percent
After 2000	196,086	5,069	6.74	0.17
Not Reported	54,430	2,524	1.87	0.09
Employment Setting				
Hospital	1,360,847	13,063	46.77	0.43
Nursing Home Extended Care	153,172	3,369	5.26	0.12
Nursing Education	63,444	2,879	2.18	0.10
Public Health/Community Health	259,911	4,347	8.93	0.15
School Health Service	78,022	3,095	2.68	0.10
Occupational Health	22,447	1,820	0.77	0.06
Ambulatory Care (Except Nurse Owned/Operated)	265,273	5,346	9.12	0.18
Nurse Owned/Operated Ambulatory Care Setting	12,500	1,112	0.43	0.04
Insurance Claims/Benefits	43,641	1,976	1.50	0.07
Planning/ Regul /Licensing Agency	8,733	933	0.30	0.03
Other	103,310	3,974	3.55	0.13
Not Reported	538,058	8,227	18.49	0.29
Type of Position				
Administrator Or Assistant Administrator	125,011	2,522	4.30	0.08
Consultant	35,617	1,707	1.22	0.06
Supervisor	74,201	2,976	2.55	0.10
Instructor/Faculty	62,255	2,403	2.14	0.08
Head Nurse Or Assistant Nurse	148,210	3,880	5.09	0.13
Staff Nurse	1,431,053	11,735	49.19	0.39
Nurse Practitioner	84,042	3,424	2.89	0.12
Nurse Midwife	7,274	990	0.25	0.03
Clinical Specialist	28,623	1,900	0.98	0.07
Nurse Clinician	32,954	1,908	1.13	0.07
Certified Nurse Anesthetist	27,287	1,452	0.94	0.05
Research	19,263	1,250	0.66	0.04
Private Duty	11,762	1,280	0.40	0.04
Informatic Nurse	8,570	929	0.29	0.03
Home Health	45,621	1,834	1.57	0.06
Survey Or Auditors/Regulator	12,097	1,031	0.42	0.04
Patient Coordinator	138,404	3,205	4.76	0.11
Other	82,352	3,226	2.83	0.11
Not Reported	534,760	7,774	18.38	0.27
Highest Nursing Education				
Diploma In Nursing	510,209	8,062	17.54	0.27
Associate Degree In Nursing Or Related Field	981,238	14,852	33.73	0.49
Baccalaureate In Nursing	922,696	12,963	31.71	0.45
Baccalaureate In Related Field	71,580	1,946	2.46	0.07
Masters In Nursing	256,415	5,251	8.81	0.18

Description	Estimated Number	S.E. of Estimated Number	Estimated Percent	S.E. of Estimated Percent
Masters In Related Field	94,386	3,057	3.24	0.10
Doctorate In Nursing	11,548	645	0.40	0.02
Doctorate In Related Field	14,552	1,192	0.50	0.04
Not Reported	46,733	2,300	1.61	0.08
Age of Nurse				
<25	61,778	1,486	2.12	0.05
25 To 29	171,659	3,751	5.90	0.13
30 To 34	243,182	5,572	8.36	0.19
35 To 39	289,525	6,598	9.95	0.23
40 To 44	408,248	6,721	14.03	0.23
45 To 49	508,708	7,695	17.49	0.26
50 To 54	463,565	9,646	15.93	0.32
55 To 59	338,078	6,534	11.62	0.22
60 To 64	210,196	5,764	7.22	0.20
65+	185,254	5,092	6.37	0.17
Not Reported	29,165	1,525	1.00	0.05
Marital Status and Children				
Married, Children < 6	225,572	5,474	7.75	0.19
Married, Children > = 6	650,793	8,062	22.37	0.28
Married, Children All Ages	162,791	3,393	5.60	0.11
Married, No Children	994,588	10,942	34.19	0.34
Married, Children Unknown	16,916	1,275	0.58	0.04
Widowed/ Separated/ Divorced, Children < 6	13,300	1,023	0.46	0.04
Widowed/ Separated/ Divorced, Children > = 6	137,283	4,514	4.72	0.15
Widowed/ Separated/ Divorced, Children All Ages	14,683	898	0.50	0.03
Widowed/ Separated/ Divorced, No Children	355,309	8,582	12.21	0.29
Widowed/ Separated/ Divorced, Children Unknown	5,795	817	0.20	0.03
Never Married, Children < 6	9,131	1,063	0.31	0.04
Never Married, Children > = 6	18,657	1,606	0.64	0.06
Never Married, Children All Ages	2,854	609	0.10	0.02
Never Married, No Children	234,208	5,167	8.05	0.18
Never Married, Children Unknown	3,897	680	0.13	0.02
Not Reported	63,581	2,497	2.19	0.09
Mean Gross Annual Salary for Full-Time RNs	57,784.86	180.85		
Mean Hours Worked per year	2,160.00	5.63		
Mean Hours Worked in Last Full Workweek	38.55	0.13		

Table B-3. Direct Estimates of State Nurse Population, Standard Error, and Coefficient of Variation by State, 2000

State	2004 Estimated State Nurse Population	Standard Error	Coefficient of Variation (in Percent)
United States	2,909,357	7,001	0.24
Alabama	42,894	472	1.10
Alaska	7,567	420	5.54
Arizona	48,284	910	1.89
Arkansas	23,818	569	2.39
California	255,858	1,734	0.68
Colorado	43,719	695	1.59
Connecticut	42,894	1,199	2.80
DC	9,352	324	3.47
Delaware	12,118	675	5.57
Florida	169,460	2,168	1.28
Georgia	78,898	1,070	1.36
Hawaii	11,146	387	3.47
Idaho	11,068	256	2.32
Illinois	138,092	1,236	0.90
Indiana	64,396	858	1.33
Iowa	37,777	614	1.63
Kansas	29,892	790	2.64
Kentucky	42,971	812	1.89
Louisiana	39,449	731	1.85
Maine	17,785	465	2.61
Maryland	53,061	759	1.43
Massachusetts	89,358	972	1.09
Michigan	103,697	1,406	1.36
Minnesota	60,214	621	1.03
Mississippi	27,303	517	1.89
Missouri	66,551	973	1.46
Montana	9,416	149	1.58
Nebraska	20,026	604	3.01
Nevada	16,206	427	2.63
New Hampshire	18,473	493	2.67
New Jersey	92,425	1,476	1.60
New Mexico	15,027	435	2.89
New York	215,309	2,377	1.10
North Carolina	92,391	1,238	1.34
North Dakota	7,966	206	2.58
Ohio	133,064	1,224	0.92
Oklahoma	29,268	574	1.96
Oregon	34,946	713	2.04
Pennsylvania	164,433	1,834	1.12
Rhode Island	13,847	337	2.44
South Carolina	35,204	741	2.11
South Dakota	10,223	213	2.09
Tennessee	62,266	989	1.59

State	2004 Estimated State Nurse Population	Standard Error	Coefficient of Variation (in Percent)
Texas	168,368	1,363	0.81
Utah	18,169	413	2.27
Vermont	7,137	254	3.56
Virginia	73,526	1,361	1.85
Washington	59,761	913	1.53
West Virginia	17,742	452	2.55
Wisconsin	62,044	640	1.03
Wyoming	4,498	122	2.72

**Table B-4. Median Design Effects for Percentages
Estimated from the Eighth National Sample Survey
of Registered Nurses, 2004**

State	Median Design Effect
United States	1.63
Alabama	1.06
Alaska	1.24
Arizona	1.01
Arkansas	0.98
California	1.11
Colorado	1.04
Connecticut	1.05
Delaware	0.97
DC	1.33
Florida	1.08
Georgia	1.03
Hawaii	0.99
Idaho	0.98
Illinois	1.01
Indiana	1.02
Iowa	1.10
Kansas	0.98
Kentucky	1.08
Louisiana	1.04
Maine	1.04
Maryland	1.16
Massachusetts	1.02
Michigan	0.95
Minnesota	1.01
Mississippi	1.01
Missouri	1.05
Montana	0.99
Nebraska	0.99
Nevada	1.07
New Hampshire	1.09
New Jersey	1.00
New Mexico	1.04
New York	1.04
North Carolina	1.01
North Dakota	0.97
Ohio	1.05
Oklahoma	1.02
Oregon	1.03
Pennsylvania	0.98
Rhode Island	1.00
South Carolina	1.03
South Dakota	1.06
Tennessee	0.98
Texas	1.04

State	Median Design Effect
Utah	1.02
Vermont	0.98
Virginia	1.13
Washington	1.07
West Virginia	0.93
Wisconsin	1.07
Wyoming	0.95