

The U.S. Nursing Workforce: Technical Documentation



Health Resources and Services Administration
Bureau of Health Professions
National Center for Health Workforce Analysis



Table of Contents

Data Sources for Analysis of the Nurse Supply 3
Standard Error Calculation for Census 2000 Sample Data 4
Standard Error Calculation for 2008-2010 American Community Survey PUMS Data 4
Use of the ACS PUMS Data for Rural-Urban Analyses 5

The U.S. Nursing Workforce: Technical Documentation

This technical document is a companion to the report *The U.S. Nursing Workforce: Trends in Supply and Education*.¹ It provides additional information on data and methodology for analysis of the nurse supply, including information on the calculation of standard errors, tests of significance for change over time, and use of the data for rural-urban analysis.

Data Sources for Analysis of the Nurse Supply

Data from Census 2000 and the American Community Survey were used to analyze the nurse supply over a period of approximately one decade. The 2000 Census data file was obtained from the Integrated Public Use Microdata Series (IPUMS)-USA maintained at the Minnesota Population Center at the University of Minnesota.² The 2000 Census data are a 5 percent national random sample of the population that completed the Census long form.³ The 2000 Census sample file has 14,081,466 records.

The American Community Survey (ACS) Public Use Microdata Sample (PUMS) data were downloaded from the U.S. Census Bureau.⁴ The 2008-2010 ACS data file is an approximately 3 percent sample of the U.S. population constructed by combining the 1 percent samples in the 2008, 2009, and 2010 ACS. There are 9,093,077 records in the 3-year 2008-2010 ACS data file. The 3-year ACS file was used, rather than the most recent single-year ACS file, in order to have sufficient sample sizes for state-level analysis and multi-variable cross tabulations.

The 2000 Census represents a single year in which data were collected between about March and July of 2000. The ACS data are collected throughout each calendar year and represent the aggregate characteristics over a 3-year period.

¹ Available on the web at: <http://bhpr.hrsa.gov/healthworkforce/index.html>

² Steven Ruggles, J. Trent Alexander, Katie Genadek, Ronald Goeken, Matthew B. Schroeder, and Matthew Sobek. Integrated Public Use Microdata Series (IPUM): Version 5.0 [Machine-readable database]. Minneapolis: University of Minnesota, 2010. The data file from IPUM is a national data file that combines the individual state data files available from the U.S. Census Bureau (http://www2.census.gov/census_2000/datasets/PUMS/).

³ U.S. Census Bureau. Technical Documentation: Public Use Microdata Sample, 2000 Census of Population and Housing. Washington, DC: U.S. Government Printing Office; October 2008 (<http://www.census.gov/prod/cen2000/doc/pums.pdf>).

⁴ U.S. Census Bureau at http://www2.census.gov/acs2010_3yr/pums/. See U.S. Census Bureau. A Compass for Understanding and Using American Community Survey Data: What PUMS Data Users Need to Know. Washington, DC: U.S. Government Printing Office; February 2009 (http://www.census.gov/acs/www/guidance_for_data_users/handbooks/).

Standard Error Calculation for Census 2000 Sample Data

The standard errors of 2000 estimates were calculated using the design factor method outlined in documentation from the U.S. Census Bureau.⁵ The design factor represents the effects of the sample design and estimation procedure used for the Census 2000 sample data. The Census long form public use data file does not contain variables for the parameters of the complex sample design or replicate weights for direct estimation of standard errors. In general, the design factor method provides a conservative estimate of standard errors.

In the calculation of a standard error in the Census 2000 sample, first the unadjusted standard error is calculated. Then the unadjusted standard error is multiplied by the design factor for the characteristic. For example, the design factor for race is 2.2, and that for sex is 1.2.⁴

The formula for calculating the unadjusted standard error (SE) of a percent in the Census 2000 sample is:

$$SE(p) = \sqrt{\left(\frac{19}{B}\right) p(100 - p)}$$

where p = the percent and B = the estimated (weighted) base or denominator used in calculating the percent.⁴ The result from this formula is multiplied by the design factor to derive the approximate adjusted standard error.

Standard Error Calculation for 2008-2010 American Community Survey PUMS Data

The 2008-2010 3-year ACS PUMS file contains 80 replicate weights for direct calculation of standard errors. As stated in documentation for 2008-2010 3-year ACS:⁶

The standard error of X can be computed after the replicate estimates X_1 through X_{80} are computed [using each replicate weight]. The standard error is estimated using the sum of squared differences between each replicate estimate X_r and the full sample estimate X .

The standard error (SE) formula using replicate weights is:

$$SE(X) = \sqrt{\left(\frac{4}{80}\right) \sum_{r=1}^{80} (X_r - X)^2}$$

⁵ U.S. Census Bureau. Chapter 4: Accuracy of the Microdata Sample Estimates and Table E: Census 2000 PUMS Standard Error Design Factors—United States in Technical Documentation: Public Use Microdata Sample, 2000 Census of Population and Housing. Washington, DC: U.S. Government Printing Office; October 2008 (<http://www.census.gov/prod/cen2000/doc/pums.pdf>).

⁶ U.S. Census Bureau. 2008-2010 PUMS Accuracy of the Data.

http://www.census.gov/acs/www/Downloads/data_documentation/pums/Accuracy/2008_2010AccuracyPUMS.pdf

In the SUDAAN statistical software package, there are two methods of calculating standard errors using replicate weights: (1) the jackknife method and (2) balance repeated replication (BRR). Using BRR for the formula above, the code for a crosstab procedure is:⁷

```
proc crosstab data=[name] design = BRR;  
weight pwgtp;  
repwgt pwgtp1-pwgtp80 / adjfay = 4;
```

where pwgtp is the name of the person weight variable in the ACS data file and pwgtp1-pwtp80 are the names of the respective replicate weight variables. The statement “adjfay=4” adjusts for the “4/80” in the formula. Either the jackknife or BRR method can be used, with the relevant specified adjustment, as each produced the same standard errors.⁸ The replicate weights in the ACS can have a negative value. SUDAAN treats negative replicate weights as zero.

The significance of the differences between 2000 and 2008-2010 were based upon the Z value calculated from the following formula:⁹

$$Z = \frac{A - B}{\sqrt{[SE(A)]^2 + [SE(B)]^2}}$$

A is the percent in 2008-2010 and B is the percent in 2000. Differences are reported as statistically significant if $p < .05$, i.e., $Z < -1.96$ or $Z > 1.96$.

Use of the ACS PUMS Data for Rural-Urban Analyses

The American Community Survey (ACS) public-use data files do not contain a variable to indicate the metropolitan status (e.g., rural or urban) of the household location of an individual. The PUMS files do contain a geographical variable to indicate the Public Use Microdata Area (PUMA) in which a household is located. A PUMA is an area with a minimum population of 100,000. Smaller geographical units are not provided in ACS PUMS files in order to protect the

⁷ See U.S. Census Bureau. Estimating ASEC Variances with Replicate Weights. Available at the link “Estimating ASEC Variances with Replicate Weights” on the web page at <http://usa.ipums.org/usa/repwt.shtml>.

⁸ The SUDAAN setup for the jackknife method is: “proc crosstab data=[name] design = jackknife; weight pwgtp; jackwgt pwgtp1-pwgtp80 / adjjack = .05;”. The “adjjack = .05” is the adjustment for the “4/80” in the standard error formula for the ACS using replicate weights [4/80 = .05]. (See jackknife example at <http://usa.ipums.org/usa/repwt.shtml>.) The differences in the adjustments used for jackknife and BRR relates to the different formulas used in jackknife and BRR for calculating standard errors (see Research Triangle Institute. SUDAAN Language Manual, Release 9.0, 2004, pp. 59-62). We calculated standard errors using BRR and jackknife in SUDAAN to compare resulting standard errors. BRR (using adjfay=4) and jackknife (using adjjack = .05) produced identical standard errors in the analysis using the 2008-2010 ACS presented in this report.

⁹ U.S. Census Bureau. Instructions for Applying Statistical Testing for the 2008-2010 3-Year Data and the 2006-2010 ACS 5-Year Data.

http://www.census.gov/acs/www/Downloads/data_documentation/Statistical_Testing/2010StatisticalTesting3and5year.pdf

confidentiality of respondents in the survey. PUMAs are defined every ten years in conjunction with the decennial census survey.¹⁰

The Economic Research Service at the U.S. Department of Agriculture (USDA) has constructed a classification of PUMAs as metropolitan or nonmetropolitan. This classification was applied in the analysis presented in this report.¹¹ The term “urban” is used for the PUMAs classified as metropolitan and the term “rural” is used for the PUMAs the USDA scheme labels as nonmetropolitan.

PUMAs can be comprised of multiple counties or subparts of a county. The USDA classified each county or county subpart in a PUMA as either metropolitan or nonmetropolitan using the USDA 2003 Rural Urban Continuum Codes (RUCC).¹² The values of the RUCC are 1 through 9 in which metropolitan counties were defined as those with RUCC values of 1, 2, or 3. All other counties were defined as nonmetropolitan.

For each county or subpart county in a PUMA, the population (in 2000) was obtained by USDA from state-level data files from the Census Bureau. The total population in a PUMA was thus divided into the number in metropolitan counties (based upon the RUCC scheme) and the number in nonmetropolitan counties. There were 1596 PUMAs (77%) where all of the population lived in metropolitan counties. In 225 PUMAs (11%), the population all lived in a nonmetropolitan county. The remaining 250 PUMAs (12%) contained both metropolitan and nonmetropolitan counties. If over 50 percent of the total population in a PUMA was attributed to metropolitan counties making up the PUMA, then the PUMA was classified as a metropolitan area. If half or less of the PUMA population was in metropolitan counties, then the PUMA was classified as nonmetropolitan.

¹⁰See A Compass for Understanding and Using ACS Data: What PUMS Data Users Need to Know, U.S. Census Bureau, February 2009, <http://www.census.gov/acs/www/Downloads/handbooks/ACSPUMS.pdf>

¹¹ The data file containing the USDA classification of PUMAs as metropolitan or nonmetropolitan was obtained by the Health Resources and Services Administration from USDA in September 2012 along with the paper describing the USDA method authored by Tom Hertz at USDA entitled, “Mapping County Typology Codes and Metro Status onto the Census Public Use Microdata Areas (PUMAs).”

¹² For information on the RUCC scheme see <http://www.ers.usda.gov/data-products/rural-urban-continuum-codes.aspx>