Toward a Method for Identifying Facilities and Communities with Shortages of Nurses

February 2007

Preface

This report summarizes the findings of a research study conducted to identify and evaluate different methods for assessing the extent to which health care facilities and geographic areas are experiencing shortages of registered nurses (RNs). It documents the strengths and weaknesses of different methods and identifies approaches that appear to be especially effective or promising. A companion report is available that provides additional details about the different statistical models and analyses summarized in this report.

The study was conducted by the Center for Health Workforce Studies (the Center) at the School of Public Health at the University at Albany, State University of New York under a contract with the Division of Shortage Designation at the Health Resources and Services Administration (HRSA) of the USDHHS. The report was prepared by Paul Wing, Sandra McGinnis, and Jean Moore of the Center staff, with the assistance of Zulkarnain Pulungan, Tracey Continelli, and Ajita De, all graduate research assistants at the Center. The authors acknowledge the contributions of Diane Douglas, the HRSA project officer, and her colleagues from HRSA for their help in framing the tasks to be performed and reviewing drafts of documents. The contributions of a formal advisory committee are also gratefully acknowledged. Responsibility for the accuracy of the report rests solely with the authors.

The study team gratefully acknowledges the special contributions of Linda Lacey of the North Carolina Center for Nursing to this research effort. The provision of the responses to their surveys made possible much of the empirical analysis conducted in the early phases of the study. The cooperation of Patricia Moulton of the Center for Rural Health at UND in North Dakota, who also provided data for analysis, is also acknowledged. Other organizations and states are also acknowledged for their assistance early in the study by participating in discussions of possible pilot testing of different methods, including agencies in Iowa, California, Delaware, and Pennsylvania.

The Center was established in 1996 to collect, analyze, and present data about health care workers to inform provider, professional, government, and education organizations; policy makers; and the public. Today, the Center is a national leader in the field of health workforce studies. It supports and improves health workforce planning and access to quality health care through its efforts to compile, collect, track, analyze, evaluate, and disseminate information about the health workforce at the national, state, and local levels. Additional information about the Center and copies of many Center reports can be found on its Web site: http://chws.albany.edu.

Table of Contents

I. Background 1
A. Federal Initiatives to Address Nursing Shortages 1
B. Study Overview 3
1. Project Goals and Objectives 3
2. Expert Advisory Panels 4
3. Characteristics of an Ideal Shortage Designation Method 5
C. Initial Literature Review 6
1. Characteristics of RNs 6
2. Employment Settings 7
3. Trends in Supply 7
4. Geographic Distribution 8
5. Projections of Future Supply 9
6. Nursing Shortages 11
II. Data Sets and Compilations 12
III. Models and Analyses Based on Facility Data 13
A. Preliminary Ordinary Least Squares (OLS) Regressions 14
B. Ordered Probit Models 19
C. Validation of North Carolina Results 21
D. Apply North Carolina Ordered Probit Coefficients in North Dakota 21
E. OLS Regressions for Vacancy Rates Using Combined Data from NC and ND 21
IV. Models and Analyses Based on Geographic Data 25
A. Limitations and Challenges 25
B. Measuring RN Supply at the County Level 26
C. Adjusting for Commuting 26
D. Geography-Based Methods 26
1. RNs to Population Ratio Method 27
2. RNs to Adjusted Population Method 27
3. RN to Physician Ratio 27
4. County Cluster Adjustments 27
5. Cross-County Patient Flow Adjustments 28
6. Factor Analysis of Nursing Shortage Indicators 28
V. Preferred Method 31
A. Estimate Health Care Utilization 31
B. Estimate Current National RN Staffing 31
C. Estimating RN Demand by County 33
D. Use Supply of RNs to Estimate RN Shortages 33
VI. Outstanding Issues Related to the Preferred Method 34
A. The Problem of Patient Acuity 34
B. RN Commuting Patterns 36
1. Models to Predict RN Commuting Patterns 36
2. Using Commuting Patterns to Estimate RN Supply 38
VII. Study Recommendations 39
VIII. References 41

List of Tables

Table 1. Active RNs in the U.S. by Gender and Age Group, 2004 6
Table 2. Coefficients for Full and Abbreviated OLS Regression Models to Predict Number of Adverse Effects of Nursing Shortages in Hospitals in NC 16
Table 3. Coefficients for Full OLS Regression Model to Predict RN Vacancy Rates in Nursing Homes in NC 17
Table 4. Coefficients for Full OLS Regression Model to Predict Number of Adverse Effects of Nursing Shortages in Home Health Agencies in NC 17
Table 5. Coefficients for Full OLS Regression Model to Predict Number of Adverse Effects of Nursing Shortages in Public Health Agencies in NC 18
Table 6. Coefficients for Abbreviated OLS Regression Model to Predict Number of Adverse Effects of Nursing Shortages in Public Health Agencies in NC 19
Table 7. Coefficient Estimates for the Ordered Probit Nursing Shortage Model Based on All Facilities in NC 20
Table 8. OLS Coefficient Estimates for RN Vacancy Rates in Hospital Settings for Combined NC & ND Model 22
Table 9. Coefficient Estimates for RN Vacancy Rates in Home Health Setting for Combined NC & ND Model 23
Table 10. Coefficient Estimates for Long-Term Care Setting for Combined NC & ND Model. 23
Table 11. Coefficient Estimates for Public Health Setting for Combined NC & ND Model 24
Table 12. Standardized Factor Analysis Coefficients Related to Nursing Shortages in Counties in the U.S. 29
Table 13. Numbers and Percentages of Counties in Factor Analysis Categories, by Census Division 30
Table 14. National RN Staffing Ratios by Type of Care 32
Table 15. Ordinary Least Squares Regression Coefficients Predicting RN Incommuting, By Type of County 37
Table 16. Percentage of Cases in Which Estimate Differs From Actual by More Than 10% 38

List of Figures

Figure 1. RN Employment by Setting, 2004 7
Figure 2. Active RNs per 100,000 Population, U.S., 1980 to 2000 8
Figure 3. RNs per 100,000 Population in the U.S., 2004 9
Figure 4. National Supply and Demand Projections for RNs, 2000 to 2015 10
Figure 5. Indexed Projections of RNs per 100K Population, RNs per 100K Age 65+ Population, and Projected Numbers of Active RNs, 2004 to 2024 10
Figure 6. Facilities in North Carolina Reporting Different Levels of Difficulty Recruiting Nurses, 2004 14
Figure 7. Percentage of Facilities in NC Reporting That Recruiting Nurses Was Either 'Difficult' or 'Very Difficult', 2004 14
Figure 8. Estimated RN Shortage Percentages for Counties in the U.S. 35

Toward a Method for Identifying Facilities and Communities with Shortages of Nurses

I. Background

In 2004, the Health Services and Resources Administration (HRSA) issued a Request for Proposals for a two-year research project to gather information and insights in support of the development of a new methodology for identifying health care facilities and communities with critical shortages of registered nurses (RNs). HRSA's decision to support this research was based in large part on their concern that its current method for identifying facilities and communities with shortages of RNs was too narrow in scope and that RN shortages were likely to worsen over the next 20 years. The New York Center for Health Workforce Studies at SUNY Albany was selected to conduct this study.

This report summarizes the findings of the various components of this empirical research study. It describes a number of methods for identifying facilities and communities with shortages of nurses. It documents the strengths and weaknesses of different methods for assessing the extent of shortages of RNs in facilities and communities. The report is presented in seven sections, each summarizing a different aspect of the study:

- Federal Initiatives to Address Nursing Shortages
- Initial Literature Review
- Data Sets and Compilations
- Methods and Analyses Based on Facility Data
- Methods and Analyses Based on Geographic Data
- Preferred Method
- Study Recommendations

In addition to summarizing these research components of the study, this report presents a series of conclusions designed to inform policy makers and other researchers who may be interested in implementing or adapting one or more of these methods in the future. Additional details about the different methods, including estimates of the supply and demand for RNs in different jurisdictions, can be found in the technical report prepared as part of the larger study.

A. Federal Initiatives to Address Nursing Shortages

The Federal government has had a long-standing interest in the nursing workforce. For more than two decades, through its National Center for Health Workforce Analysis, Division of Nursing and the Shortage Designation Branch of HRSA has collected data on nurses in the U.S. and developed quantitative models to estimate the current and future supply of and demand for RNs. Several programs to encourage new RNs to practice in facilities and communities with severe shortages of RNs, including the Nursing Education Loan Repayment Program (NELRP) and the Nursing Scholarship Program, have been operating for many years. These programs help to alleviate persistent shortages of RNs.

In framing the parameters for this research study, HRSA identified a number of issues that needed resolution including:

- Should indicators developed to measure critical shortages of RNs be based on *need* for RNs or *demand* for RNs?
- Can standard indicators that measure critical shortages of RNs be applied to all of the eligible settings[1] included in this study?
- Can variations in the supply of and demand for RNs by region, geography (i.e., rural or urban), setting, or facility be accounted for in indicators that measure RN shortages?
- Are setting-specific data sets available at the national level that include the elements needed to measure critical shortages of RNs?
- Can a process be developed that identifies facilities with the most serious shortages of RNs so that Federal resources can be targeted to the neediest facilities?
- How can true shortages of nurses be distinguished from shortages created by poor management practices?

An effective study should take all of these issues into account while researching and testing the development of a national methodology to measure shortages of RNs. Current methods are inadequate. A better method would support several government incentive programs to attract new nurses. It would also provide a better basis for monitoring RN shortages locally and nationally.

One important Federal response to the national nursing shortage was the Nurse Reinvestment Act, which was enacted in August 2002. The Act reauthorized the NELRP, which provides loan repayment to RNs in return for work at facilities or in communities with a shortage of RNs, and established the Nursing Scholarship Program. Eligible placement sites for these programs were expanded to include:

- Ambulatory surgical centers;
- Federally designated migrant, community public housing, or homeless health centers;
- Federally qualified health centers;
- Home health agencies;
- Hospice programs;
- Hospitals;
- Indian Health Service centers;
- Native Hawaiian health centers;
- Nursing homes;
- Rural health clinics; and
- State or local health department clinics or skilled nursing facilities.

The method used for the identification of qualified placement sites included a combination of geographic and facility designations. In 2002, the New York Center for Health Workforce Studies assisted the Bureau of Health Professions by developing an up-to-date list of nursing

[1] The eligible settings included in this study are: hospitals, home health agencies, hospices, ambulatory care sites (including community health centers and other public clinics), long-term care facilities, and state or local health departments.

shortage hospitals and counties throughout the U.S. and its territories. The Center used two separate methodologies, one to identify private, non-profit hospitals with shortages of nurses and the second to identify counties with shortages of nurses. Appendix C of the full technical report provides a brief description of the methodologies used for the designation of hospitals and areas with nursing shortages.

Because this approach relied on hospital nursing data to identify facilities with nursing shortages, it failed to quantify nursing shortages experienced by any providers except hospitals. Most of the other types of facilities included on the list above were considered categorically eligible, based on the premise that they faced critical shortage of nurses.

B. Study Overview

In the general context described above, this study was conducted over a two-year period, starting in the fall of 2004. After a brief summary of the study goals, objectives, and other characteristics of the study, the ten study components are summarized below.

1. Project Goals and Objectives

The primary goal of this study was to conduct research on the necessary components of a comprehensive, nationwide methodology to identify facilities and communities with critical shortages of RNs across the U.S. and its territories in order to target the placement of Federally-obligated RN scholars and loan repayers. This research, which involved statistical analysis supported by expert opinion, took into account population needs, practice settings, appropriate staffing levels, and nursing education, among other aspects of the supply of and demand for RNs. As a secondary benefit, the project revealed important insights about the differences in the use and distribution of RNs across the various settings and geographic areas of the country.

The study's staff worked to achieve the following objectives in support of the primary goal of the study:

- Identify and define indicators and measures that reflect critical RN shortages for the four types of facilities;
- Assess the availability of data sets that can be used to determine RN staffing needs nationally in each of the settings listed above;
- Develop quantifiable key measures of nursing shortages based on key indicators described above as well as the available data sets that include the necessary data to calculate the key measure.
- Determine whether these key measures of shortage can be incorporated into a comprehensive national methodology to identify facilities and agencies with critical nursing shortages based on the following criteria:
    - the measure accurately quantifies nursing shortages in a specific health care setting; and
    - the measure either can be calculated using an available national data set or the data can be collected and validated at the facility level.
- Establish an analytic framework that can be used for a comprehensive methodology to determine critical nursing shortages across a variety of health care settings.

Ultimately, this research will support the development of a comprehensive method for identifying the health care facilities and agencies with critical shortages of RNs. This will permit more effective targeting of Federal and other resources to encourage service-obligated RNs to work in the facilities and communities with the greatest needs.

2. Expert Advisory Panels

The study was conducted under the guidance of four expert advisory panels, one for each of four types of health care organizations: hospitals, home health agencies, nursing homes, and public health agencies. The names of the panelists can be found in Appendix B of the full, technical report.

These panels met face-to-face twice. The first meetings were held separately early in the study to discuss preliminary findings and agree on strategies for accomplishing study goals and objectives. The second meeting convened all the panels together toward the end of the study to gain the benefit of cross-fertilization of ideas. In between these meetings the panelists were invited to participate in two conference calls in which interim progress reports were provided to solicit feedback and suggestions.

An important outcome of the initial meetings of the advisory panels was agreement on a list of "guiding principles" to inform and direct our efforts. These principles can be roughly classified as relating to theoretical, practical, or fairness concerns. The list also included some specific recommendations about methodology.

The theoretical principles and ideals included:

- **Context: facility within community**. Both facility and community characteristics must be considered, but community characteristics are more important than facility characteristics.
- **Demand over need**. Analyses should primarily focus on employer demand for RNs (e.g., what the local labor market will actually support) rather than the health needs of the population. High-need areas that have no resources or infrastructure to employ additional RNs would find little benefit in the NELRP program.
- **Identify standards for data**. Ultimately, it will be important to upgrade Federal, state, and local data systems to support better planning for the nursing workforce, including the designation of facilities and communities with shortages of RNs.
- **Consider facility culture**. Some facilities may experience high RN vacancies not because of difficulties recruiting RNs, but because of persistent RN turnover due to problems of organizational culture within the facility (e.g., poor management). This is not a "shortage" issue, and the NELRP program is not intended to address such problems.
- **Define shortage based on outcomes**. Theoretically, a facility can be said to have "too few" RNs when there are not enough RNs for the facility to effectively function. This will be observed in certain outcome measures relating to quality of care and facility functioning.

The principles and ideals relating to practical concerns included:

- **Low administrative burden on facilities and HRSA**. Data used in the final methodology should not require a large-scale data collection or manipulation.

- **Applicable to all facility types**. The final shortage methodology should be applicable to and appropriate for all facility types.
- **Readily available data over time**. Ideally, the final methodology should be supported by existing data that are easy to access and available over time for updating.
- **Commonly accepted data elements and indicators**. Using established indicators of supply, demand, and shortage is preferable to developing new ones.
- **Easy to update to reflect changing environment**. Data used for identifying shortages should be easy to update so that designations can be periodically reexamined.

The principles and ideals relating to fairness included:

- **Attention to rural and urban differences**. The shortage designation method should not systematically disadvantage either rural or urban facilities.
- **Special needs of some facilities**. The shortage designation method should recognize extenuating circumstances (e.g., facing critical problems, serving special populations).
- **Case mix of patients**. The method should recognize that some facilities have higher patient acuity than others, which may signify that some facilities require more intensive staffing.
- **Accommodate data manipulation**. The method should minimize opportunities for facilities and communities to "game" the system to achieve a shortage designation.

Specific recommendations for the method included:

- **Look beyond clinical care**. It should be recognized that overall demand for RNs extends beyond just those at the bedside to those in non-clinical positions.
- **Consider overall staff mix**. Some employees may substitute for RNs with other personnel. This may be more or less appropriate depending upon the facility type.
- **Consider RN staff mix** (e.g., specialty, education). Facilities with enough RNs overall may still have a shortage of RNs with certain credentials or in some services (e.g., ICUs).
- **Separate out different units within hospital care**. Different units have different staffing needs (e.g. intensive care units will require more RNs than general medical-surgical units).

Most of these guiding principles were addressed in at least some of the analyses, either directly or indirectly, and many are incorporated into the Preferred Method proposed by the study.

3. Characteristics of an Ideal Shortage Designation Method

Early in the study a number of characteristics were identified as especially desirable for any method to identify facilities and communities with shortages of RNs. These characteristics, some of which may not be attainable, included:

- A common method to be used across the nation;
- Ease of calculation of the RN shortage index for individual facilities and communities;
- Implementation using existing data sets, with no additional data collection required;
- Comparison of shortages of RNs both within and between different types of facilities;
- Comparison of RN shortages across different states and other geographic jurisdictions;

- Consistency of shortage severity estimates with shortage assessments by local experts;
- Identification of shortages in facilities due to poor management; and
- Easy updates to the method to reflect more recent conditions, situations, and relationships.

C. Initial Literature Review

The first component of the research involved a careful review of the literature, focusing on characteristics of RNs relevant to the task of understanding current and future shortages. The discussion that follows summarizes a variety of relevant statistics.

1. Characteristics of RNs

- Table 1 shows that although 6.1% of RNs were men in 2004, which is higher than in previous years, nursing remains a female-dominated profession. This means that, at least in the near future, recruiting more men to the profession is not likely to be an important avenue for increasing the supply.
- By 2014 it will be necessary to recruit more than 400,000 new RNs just to replace those RNs older than age 55 who are expected to retire from active nursing practice.
- The latest estimates developed by the Bureau of Labor Statistics [BLS, 2006] indicate that the U.S. will require 1.2 million new RNs by 2014 to meet the nursing needs of the country, 500,000 to replace those leaving practice and an additional 700,000 to meet growing demands for nursing services.

Table 1. Active RNs in the U.S. by Gender and Age Group, 2004

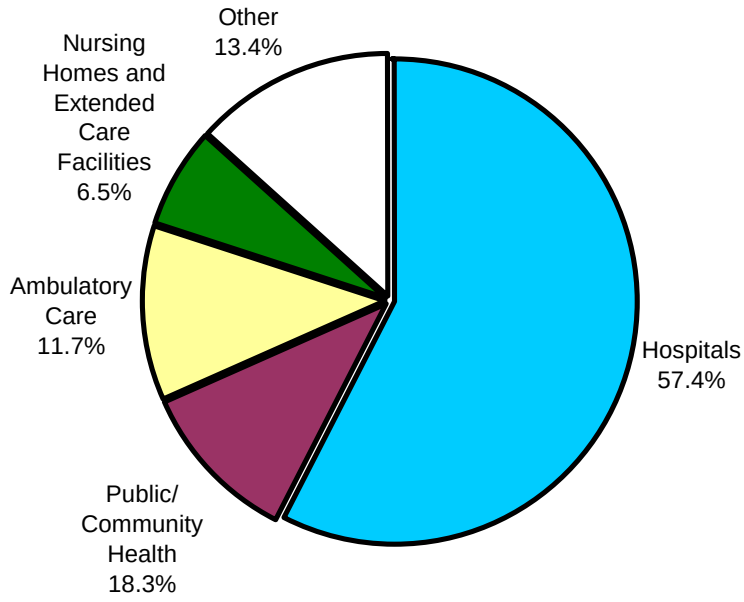
Age Group	Male	Female	% in Age Group
< 25	1,731	57,843	**2.5%**
25 to 29	10,955	148,721	**6.7%**
30 to 34	15,508	205,543	**9.2%**
35 to 39	19,217	237,693	**10.7%**
40 to 44	23,951	336,195	**15.0%**
45 to 49	30,986	418,634	**18.8%**
50 to 54	24,098	382,650	**17.0%**
55 to 59	13,469	257,640	**11.3%**
60 to 64	4,909	131,281	**5.7%**
65 +	1,819	73,486	**3.1%**
% Gender	**6.1%**	**93.9%**	**2,396,329**

Source: 2004 NSSRN

2. Employment Settings

- In 2004, more than 57% of RNs worked in hospitals, 11% worked in public or community health, nearly 12% worked in ambulatory care settings, and almost 7% worked in nursing homes and extended care facilities (Figure 1).

Figure 1. RN Employment by Setting, 2004

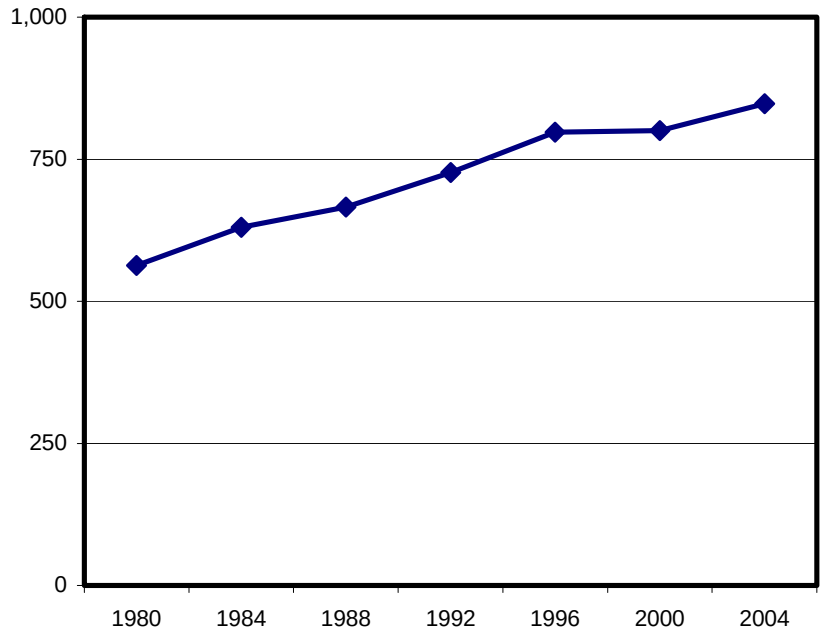


Source: The Registered Nurse Population, March 2004. USDHHS, Bureau of Health Professions, Division of Nursing, November 2005.

3. Trends in Supply

- Between 1980 and 2004, the number of active RNs in the U.S. grew by nearly 90%. In 2000, there were more than 2.4 million active RNs, an increase of more than 1.1 million over 1980.
- Between 1996 and 2000, the total number of RNs grew by only 1.3% each year, compared with average annual growth of 2% to 3% in earlier years (Figure 2). This slowdown in growth between 1996 and 2000 is attributable to two trends: a declining number of candidates passing the RN licensing examination annually and an increasing number of RNs leaving the field [1]. This slowdown was temporary, however, as the growth in the supply of RNs of early years resumed between 2000 and 2004, more than keeping up with the growth in the population over the same period.
- The number of candidates passing the RN licensure examination has decreased steadily since 1995. Between 1995 and 2001, the number of RNs passing the licensing exam declined by nearly 28% [2].

Figure 2. Active RNs per 100,000 Population, U.S., 1980 to 2000



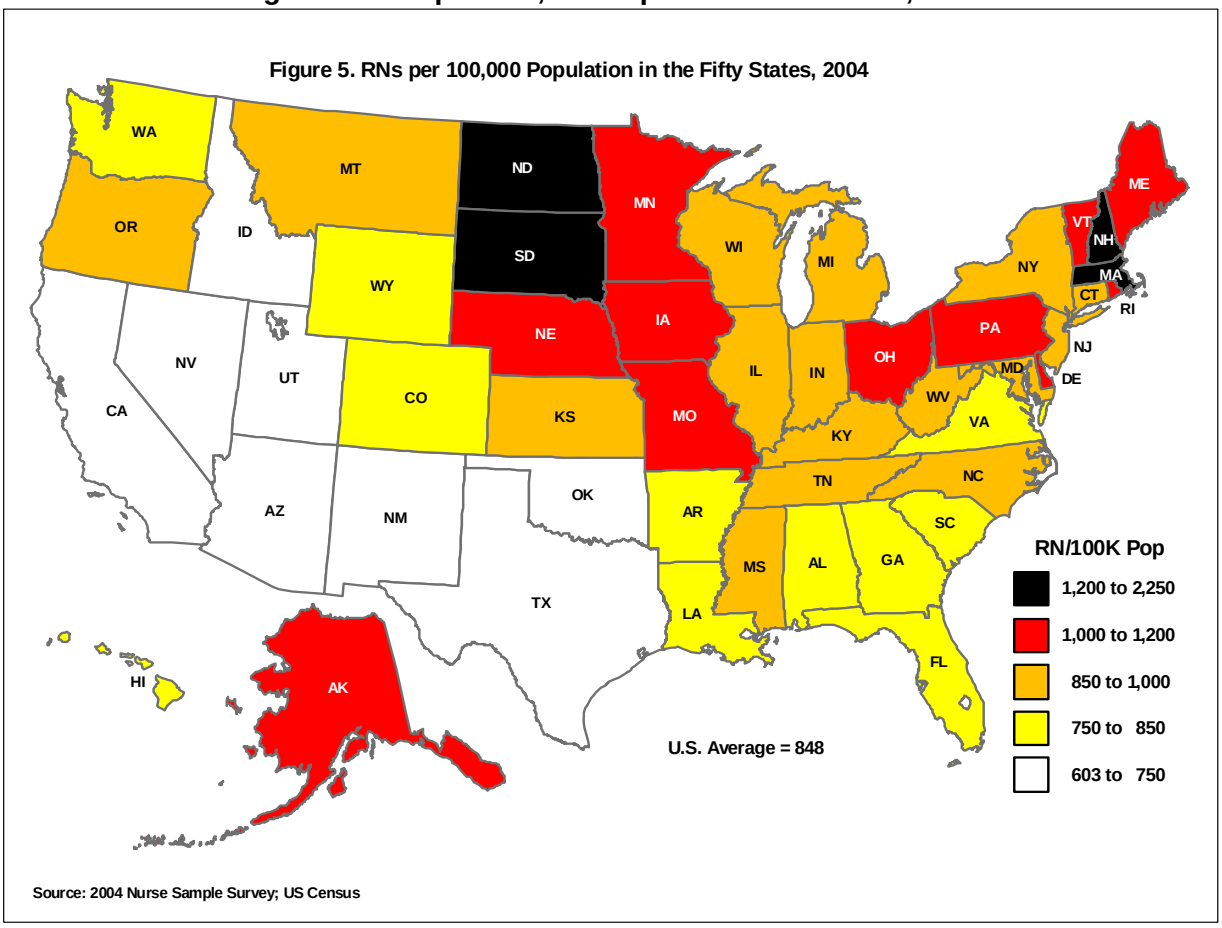
Sources: U.S. Department of Health and Human Services, National Sample Survey of RNs, 2004 and earlier; Population Estimates Program, Population Division, U.S. Census Bureau.

- Between 1995 and 2001, the number of graduates of RN education programs in the U.S. declined. Nearly 29% fewer RNs graduated in 2001 than in 1995 [2]. The numbers of RN graduates have been steadily increasing since 2002 and in some states, including New York, they are now higher than the previous peak number in 1996 [3, 4]. There is evidence capacity limits in many nursing education programs (due to such factors as the inability to recruit enough faculty) are limiting the ability of the system to accept all qualified nursing school applicants.

4. Geographic Distribution

The geographic dispersion of active RNs in 2004 was far from uniform across the U.S. In fact, Figure 3 shows that the ratio of the highest to lowest RN per capita ratios was nearly 4:1, with the highest ratios in the District of Columbia (2,236 RNs per 100,000 population) and New Hampshire (1,321), and the lowest in California (603) and Nevada (612).

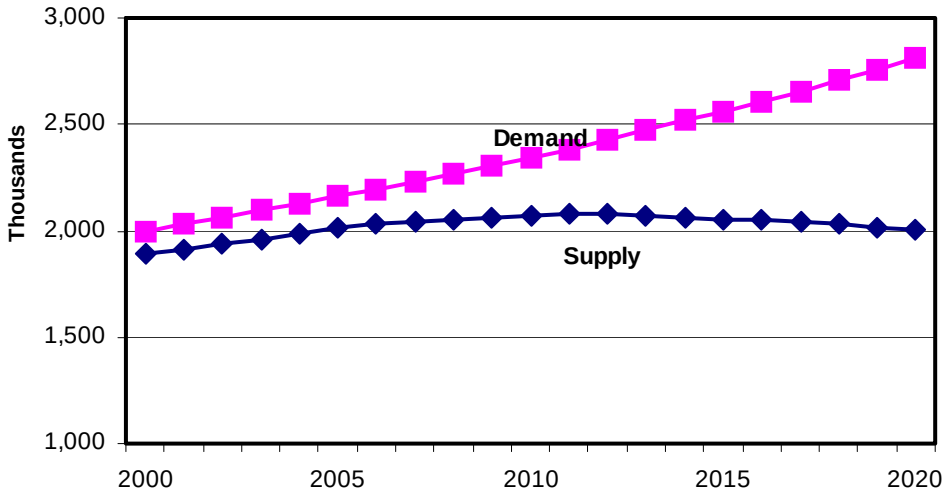
Figure 3. RNs per 100,000 Population in the U.S., 2004



5. Projections of Future Supply

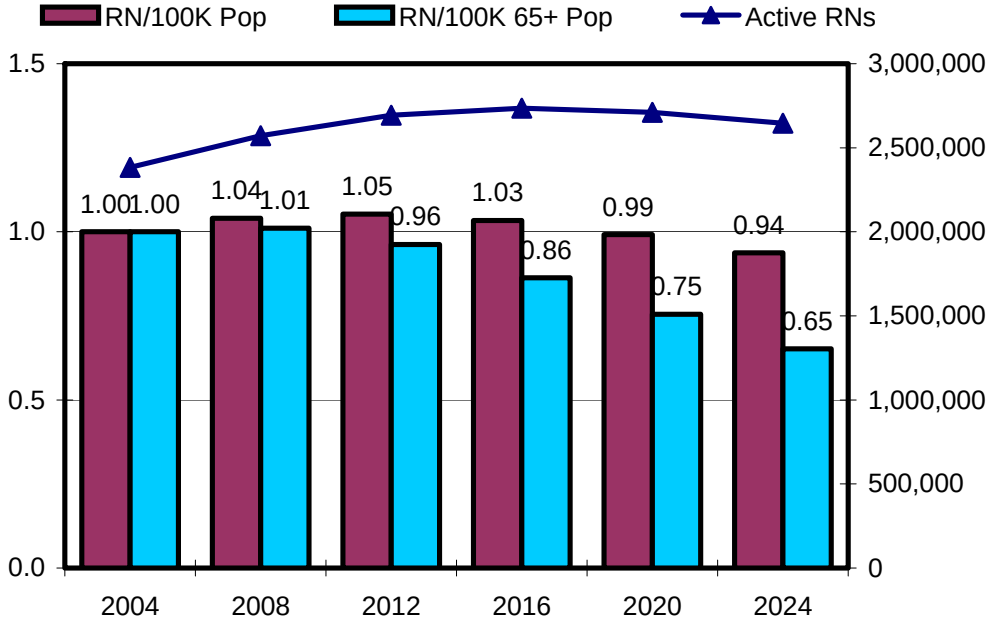
- A growing shortage of RNs has been projected over the next 15 fifteen years, with a 12% shortage by 2010 and a 20% shortage by 2015 (Figure 4). The projected shortage is the result of the expected increase in demand coupled with a relatively stable supply of RNs [5].
- Figure 5 updates these projections based in part on the 2004 National Sample Survey of Registered Nurses (NSSRN). Total numbers of RNs may rise until 2016 if age-specific cohorts follow patterns observed in the NSSRN between 2000 and 2004. This is in large part because the sizes of birth cohorts in nursing tend to increase well into ages 50-55, and so a number of baby boomers (currently ages 43 to 60) may still enter nursing as a second career over the next 10 years.
- This does not mean that problems will not be felt until after 2016, however. Using these projections of numbers of RNs and projections of the total population and the population age 65 and older from the U.S. Census Bureau, Figure 5 shows that the number of RNs per 100,000 population will peak in 2012, while the number of RNs per 100,000 population age 65 and older will peak in 2008 and decline by 5% (falling below current rates) by 2012.

Figure 4. National Supply and Demand Projections for RNs, 2000 to 2015



Source: Bureau of Health Professions, RN Supply and Demand Projections

Figure 5. Indexed Projections of RNs per 100K Population, RNs per 100K Age 65+ Population, and Projected Numbers of Active RNs, 2004 to 2024



Source: CHWS, 2006

6. Nursing Shortages

A review of the literature revealed a number of studies examining future shortages of RNs relevant to this study. Some of the key findings are summarized briefly below.

- Health care providers across a variety of settings reported increasing difficulty recruiting and retaining RNs, particularly in hospital settings [6, 7].
- There were indications that the attrition from clinical settings may be related to dissatisfaction with working conditions. The 2004 NSSRN asked RNs about job satisfaction and found that 76% of RNs employed by hospitals and 75% of RNs employed by nursing homes were satisfied with their jobs, compared to 82% of RNs employed in nursing education and 83% of RNs employed in occupational health.
- There is growing concern about the impact of RN shortages on the quality of health care. A growing body of evidence demonstrates that hospitals with lower ratios of RNs to patients had more adverse events than hospitals with higher RN to patient ratios [8, 9, 10].

II. Data Sets and Compilations

A number of data sets were analyzed over the course of this study. The most important of these are described briefly below.

Survey of Nurse Employers in North Carolina and the North Dakota Nursing Needs Study Facility Survey

Two datasets, one from the North Carolina Center for Nursing and the other from the University of North Dakota School of Medicine & Health Sciences Center for Rural Health were used extensively in this study. These were the best sources of data for facility-level analyses in hospitals, long-term care facilities, home health agencies, and public health agencies. Both datasets included data on RN staffing, turnover, vacancies, and recruiting difficulty. Because the surveys were very similar, they allowed many of the same analyses and some direct cross-state comparisons. These data were used for descriptive analyses, ordinary least squares regression, and ordered probit models.

Area Resource File (ARF)

The ARF, which is maintained by HRSA, contains county-level data from various sources on health care utilization, health care infrastructure, the health workforce, health care spending, and population demographics and economics.

U.S. Census Bureau

The U.S. Census data includes population counts by sex and age, and also occupational data at the county level. Data on sex and age are taken from the entire population. The occupational data is taken from a 1-in-6 sample of the population, and may not be completely accurate for small counties, but it is probably the best national source for the number of RNs at the county level. The Census also includes data on nursing home residents by county, which is also taken from the entire population.

National Sample Survey of Registered Nurses (NSSRN)

The NSSRN is the most detailed source of data on RNs in the U.S. Unfortunately, the sampling design makes it unsuitable for sub-state analysis, but it is a comprehensive source of the number of RNs working in various types of settings nationwide. Most of the NSSRN data used in this study were from the 2004 survey, but data were used from the 2000 and 1996 surveys as well.

Health, United States, 2005

Health, United States is an annual compilation of national health statistics. It is a good source of national utilization data for various types of health care, which can be combined with nurse staffing data from the NSSRN to produce national benchmarks for RN staffing (RNs per inpatient day, for example).

III. Models and Analyses Based on Facility Data

All of the analyses using facility data are based on North Carolina (NC) and North Dakota (ND). These datasets included a number of possible measures of shortage that could be used as dependent variables:

Effects of Nursing Shortage on Facility Operations. The surveys asked respondents an open-ended question about how nursing shortages have affected the operations of their facility. Responses were then coded into five broad categories: labor cost increase, reduced services, strain on staff, patient care problems, and organizational disturbance. More detailed codes within categories were also given (e.g., labor cost increase included breakouts for increases in agency use, recruitment costs, overtime, wages, retention expenses, development of float pools, and orientation expenses). This was an interesting variable because of in-depth discussions in the first advisory panel meeting about how the true measure of a nursing shortage should be related to patient care and facility operations. Although subjective, this variable touches on those issues. Caution was warranted, however, because the question asked about nursing shortage generally, and respondents may have answered the question thinking about LPNs as well as RNs, particularly if they were from a setting that relies heavily on LPNs (e.g., long-term care). Nonetheless, this variable was used as the dependent variable in a series of preliminary ordinary least squares (OLS) regressions.

RN Vacancy Rates. Both the NC and ND datasets included RN vacancy rates. Many facilities, however, had vacancy rates of 0, which limited the variation in the variable. Interestingly, there was very little correlation between RN vacancy rates and the number of reported effects of the nursing shortage, which was cause to question the utility of the consequences variable given its subjectivity. Vacancy rates were also used as the dependent variable in OLS regressions.

RN Turnover Rates. Turnover rates were not used in any of the in-depth analyses. In the first set of advisory panel meetings, the panelists pointed out that facilities that had a genuinely limited supply of RNs to draw from should be separated from facilities in which poor management led to large numbers of departures. Turnover can certainly reflect limited supply, but also seems likely to reflect problems of organizational culture, particularly in facilities that had low vacancy rates but high turnover (meaning that they had no trouble recruiting RNs, but had trouble retaining them.)

Time to Recruit RNs. Both datasets contained information on the average number of weeks reported to fill RN vacancies. Although theoretically a good indicator of shortage, the large amount of missing data for this variable ruled it out for practical reasons.

Difficulty Recruiting RNs. This ordinal variable was used in a series of ordered probit models conducted as part of the study. The variable used a five-point Likert scale with categories: Very Difficult, Difficult, Neutral, Easy, and Very Easy. Figure 6, which summarizes the responses for North Carolina, shows that somewhat more facilities reported difficulty than ease in recruiting RNs in 2004.

Figure 7 shows that only 4.6% of hospitals in North Carolina reported that recruiting RNs was either difficult or very difficult. The percentages were higher for home health agencies (15.2%), long-term care facilities (21.1%), and public health agencies (26.4%).

Figure 6. Facilities in North Carolina Reporting Different Levels of Difficulty Recruiting Nurses, 2004

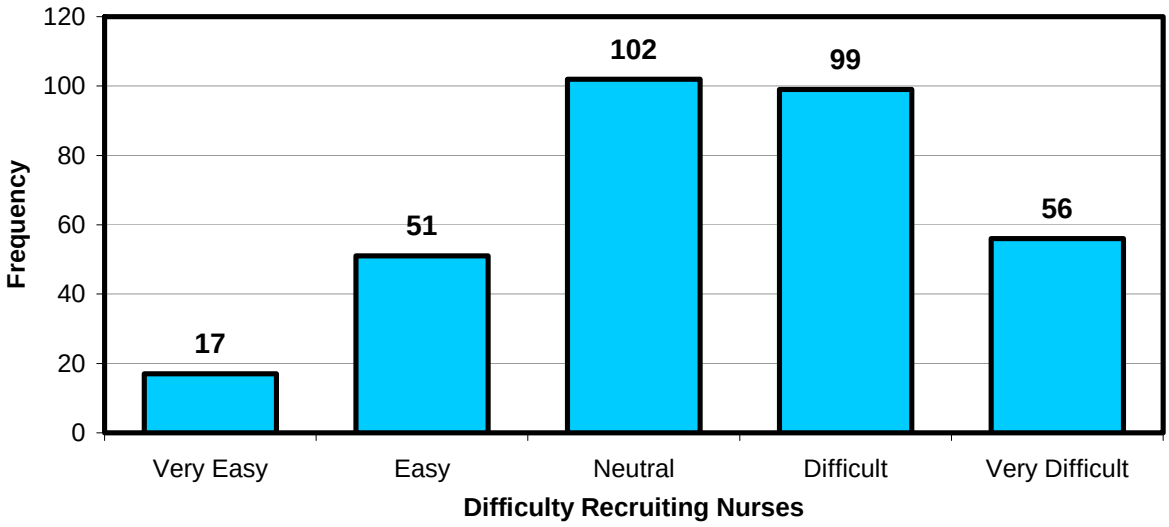
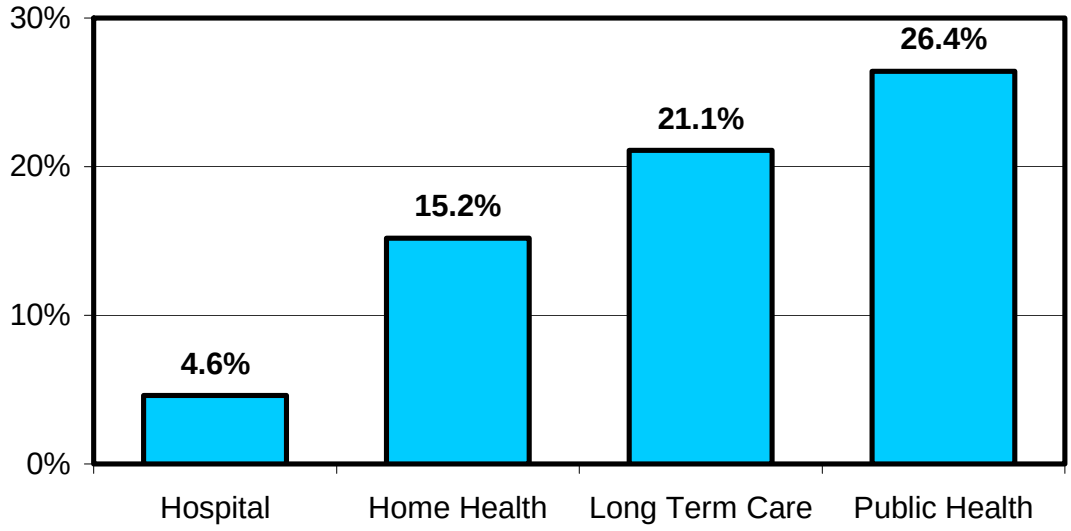


Figure 7. Percentage of Facilities in NC Reporting That Recruiting Nurses Was Either 'Difficult' or 'Very Difficult', 2004



A. Preliminary Ordinary Least Squares (OLS) Regressions

OLS regression equations were estimated to predict and explain the number of adverse consequences and vacancy rates in all four types of facilities in North Carolina. First the models were estimated with both facility- and county-level explanatory variables, which was the ideal model. In recognition of the fact that facility-level variables were not available in most states, an abbreviated model using only county-level data was estimated for each facility type as well. The results for the models in which adverse consequences were the dependent variables are shown in Tables 2 through 6.

The results of these models were not particularly satisfying. Relatively few variables were strongly correlated to adverse consequences, and the explanatory power of the models (as measured by the R^2 statistic) was generally low. Although there were some statistically significant explanatory (independent) variables in the models for both predicted consequences and vacancy rates, the models explained only a relatively small percentage of the variation in the dependent variables. The explanatory power was even smaller when the facility-level variables (which would not be available outside of NC and ND without new data collection) were removed from the models, and only community variables were used.

The conclusion based on these models is that the variables collected by North Carolina were not adequate to accurately predict and explain either adverse consequences or vacancy rates.

Table 2. Coefficients for Full and Abbreviated OLS Regression Models to Predict Number of Adverse Effects of Nursing Shortages in Hospitals in NC

Explanatory (Independent) Variable	Full Model					Abbreviated Model				
	Unstandardized Coefficient		Standard-ized Coefficient	t	p Value	Unstandardized Coefficient		Standard-ized Coefficient	t	p Value
	B	Std Err				B	Std Err			
Constant	-0.683	3.02	-	-0.226	0.822	1.295	2.374	-	0.546	0.587
RNs per 100,000 Adjusted Need	**-0.004**	**0.002**	**-0.353**	**-2.002**	**0.052**	-0.001	0.001	-0.132	-0.880	0.382
RN Salary to Average Salary	0.518	0.707	0.132	0.732	0.468	0.281	0.582	0.081	0.482	0.631
# Nursing/Personal Care Facilities	**0.032**	**0.015**	**0.663**	**2.176**	**0.035**	0.023	0.012	0.494	1.905	0.061
% Population Below Poverty, 2000	0.078	0.065	0.308	1.202	0.236	0.033	0.053	0.136	0.622	0.536
RNs per Hospital Bed	0.265	0.445	0.082	0.596	0.555	0.044	0.402	0.013	0.108	0.914
Hours of Agency RNs	0.002	0.043	0.008	0.058	0.954	-	-	-	-	-
Hours of RN Overtime	-0.001	0.016	-0.007	-0.052	0.959	-	-	-	-	-
RN Vacancy Rate	0.032	0.032	0.142	0.985	0.330	-	-	-	-	-
RN Turnover Rate	0.011	0.021	0.077	0.505	0.616	-	-	-	-	-
Persons per Square Mile (natural log)	0.156	0.358	0.146	0.436	0.665	-0.158	0.271	-0.159	-0.582	0.563
# Short-term Community Hospitals, '01	**-0.359**	**0.134**	**-0.610**	**-2.69**	**0.010**	**-0.227**	**0.109**	**-0.414**	**-2.076**	**0.041**
RN Students per 100K Adjusted Need	**-0.010**	**0.004**	**-0.392**	**-2.828**	**0.007**	**-0.005**	**0.003**	**-0.226**	**-1.967**	**0.053**
% Population White Non-Hispanic, 2004	-0.011	0.012	-0.167	-0.902	0.372	-0.005	0.010	-0.086	-0.511	0.611

Full model R^2 = 0.429 Abbreviated model R^2 = 0.177

Table 3. Coefficients for Full OLS Regression Model to Predict RN Vacancy Rates in Nursing Homes in NC

Independent Variables	Unstandardized Coefficients		Standardized Coefficients	t	p Value
	B	Std. Error	Beta		
(Constant)	-15.65	18.185	-	-0.861	0.392
RNs per 100,000 Adjusted Need	0.032	0.022	0.234	1.444	0.152
RN Salary to Average Salary	**13.83**	**6.945**	**0.316**	**1.992**	**0.049**
# Nursing/Personal Care Facilities	-0.215	0.127	-0.320	-1.687	0.095
% Population Below Poverty, 2000	**-0.939**	**0.460**	**-0.276**	**-2.039**	**0.044**
RNs per Hospital Bed	-9.236	5.976	-0.161	-1.545	0.126
Hours of Agency RNs	-0.281	0.165	-0.182	-1.704	0.092
Hours of RN Overtime	0.138	0.114	0.116	1.214	0.228
RN Turnover Rate	0.027	0.026	0.117	1.063	0.291
Persons per Square Mile (natural log)	1.824	2.768	0.120	0.659	0.512
# Short-Term Community Hospitals, '01	0.840	1.257	0.104	0.669	0.506
LPN Vacancy Rate	**0.356**	**0.083**	**0.401**	**4.287**	**0.000**
LPNs per 100,000 Adjusted Need	-0.080	0.108	-0.090	-0.740	0.461
LPNs per RN	**1.126**	**0.402**	**0.257**	**2.801**	**0.006**
LPN Turnover Rate	0.050	0.040	0.128	1.274	0.206

$R^2 = 0.35$

Table 4. Coefficients for Full OLS Regression Model to Predict Number of Adverse Effects of Nursing Shortages in Home Health Agencies in NC

Independent Variable	Unstandardized Coefficients		Standardized Coefficients	t	p Value
	B	Std Err	Beta		
(Constant)	2.270	2.216	-	1.024	0.310
RNs per 100,000 Adjusted Need	0.0022	0.002	0.214	1.412	0.163
RN salary to Average Salary	**1.570**	**0.607**	**0.480**	**2.587**	**0.012**
# Nursing/Personal Care Facilities	0.014	0.013	0.255	1.137	0.260
% Population Below Poverty, 2000	**-0.118**	**0.052**	**-0.519**	**-2.266**	**0.027**
RNs per Hospital Bed	-0.200	0.337	-0.062	-0.594	0.555
Hours of Agency RNs	**0.046**	**0.022**	**0.232**	**2.069**	**0.043**
Hours of RN overtime	-0.011	0.030	-0.041	-0.369	0.713
RN Vacancy Rate	**0.024**	**0.008**	**0.374**	**3.078**	**0.003**
RN Turnover Rate	**0.0069**	**0.003**	**0.265**	**2.339**	**0.023**
Persons per Square Mile (natural log)	-0.436	0.290	-0.392	-1.502	0.139
# Short-Term Community Hospitals, '01	-0.020	0.116	-0.027	-0.170	0.865
RN Students per 100K Adjusted Need	-0.00088	0.001	-0.202	-1.605	0.114
% Population White Non-Hispanic, 2004	-0.0136	0.010	-0.230	-1.340	0.185

$R^2 = 0.44$

Table 5. Coefficients for Full OLS Regression Model to Predict Number of Adverse Effects of Nursing Shortages in Public Health Agencies in NC

Independent Variable	Unstandardized Coefficients		Standardized Coefficients	t	p Value
	B	Std. Error	Beta		
(Constant)	2.183	2.839	-	0.769	0.447
RNs per 100,000 Adjusted Need	-0.0013	0.002	-0.123	-0.639	0.527
RN Salary to Average Salary	0.408	0.864	0.088	0.473	0.639
# Nursing/Personal Care Facilities	0.017	0.034	0.118	0.517	0.608
% Population Below Poverty, 2000	-0.066	0.056	-0.276	-1.176	0.247
RNs per Hospital Bed	0.578	0.619	0.159	0.934	0.356
Hours of Agency RNs	0.0386	0.075	0.080	0.516	0.609
Hours of RN Overtime	0.0905	0.057	0.227	1.585	0.121
RN Vacancy Rate	**0.0282**	**0.014**	**0.353**	**1.979**	**0.055**
RN Turnover Rate	0.0041	0.007	0.088	0.555	0.582
Persons per Square mile (natural log)	0.190	0.353	0.162	0.537	0.594
# Short-Term Community Hospitals 2001	-0.352	0.287	-0.250	-1.228	0.227
RN Students per 100K Adjusted Need	**-0.0015**	**0.001**	**-0.409**	**-2.321**	**0.026**
% Population White Non-Hispanic, '04	**-0.024**	**0.011**	**-0.404**	**-2.179**	**0.036**

$R^2 = 0.34$

Table 6. Coefficients for Abbreviated OLS Regression Model to Predict Number of Adverse Effects of Nursing Shortages in Public Health Agencies in NC

Independent Variable	Unstandardized Coefficients		Standardized Coefficients	t	p Value
	B	Std. Error	Beta		
(Constant)	3.607	2.172	-	1.661	0.102
RNs per 100,000 Adjusted Need	-0.00085	0.002	-0.074	-0.405	0.687
RN Salary to Average Salary	0.571	0.612	0.146	0.932	0.355
# Nursing/Personal Care Facilities 2000	0.037	0.030	0.400	1.236	0.221
Percent of Population Below Poverty, 2000	-0.086	0.051	-0.338	-1.684	0.098
Ratio of RNs to Beds	0.365	0.444	0.116	0.822	0.415
Ln Population Density	-0.084	0.262	-0.072	-0.321	0.750
# Short-Term Community Hospitals '01	**-0.430**	**0.174**	**-0.441**	**-2.468**	**0.017**
RN Students per 100,000 Adjusted Need	-0.00087	0.001	-0.203	-1.675	0.099
Number of Hospital Beds	0.00033	0.001	0.124	0.360	0.720
Percent White Non-Hispanic, 2004	**-0.0246**	**0.010**	**-0.412**	**-2.525**	**0.014**

$R^2 = 0.30$

B. Ordered Probit Models

The next set of models estimated for North Carolina used the dependent variable of difficulty recruiting RNs. Although this variable was not available for RNs overall, facilities in North Carolina did rate RN recruiting difficulty on a scale of one to five for several types of RNs in several types of units (e.g., staff RNs in ICUs, nurse managers in ob/gyn floors, etc.). To translate this set of ratings into a single summary variable, a median value was calculated for all the positions that each facility had provided. Although few facilities had valid values for all of the different categories of hires because they had not recruited for particular positions in the past year, the median did provide an estimate of the overall difficulty.

A series of ordered probit models were estimated to predict and explain variations in this new median self-reported difficulty in recruiting RNs. Coefficients for the different explanatory and independent variables were estimated for the four facility types both separately and together (to predict recruiting difficulty relative to facilities of their own type and relative to all facilities). The combined model is shown in Table 7 below; the facility-specific models are available in the technical report *Methods for Identifying Facilities and Communities with Shortages of Nurses*.

These models showed promise in explaining difficulty recruiting RNs. Nonetheless, the models were dependent upon a number of facility-level variables, and it was not clear whether a subjective assessment of the difficult recruiting was an adequate basis for rating nursing shortages in facilities.

Table 7. Coefficient Estimates for the Ordered Probit Nursing Shortage Model Based on All Facilities in NC

Variable	Hospital		Home Health		Long-Term Care		Public Health	
	Coeff	p	Coeff	p	Coeff	p	Coeff	p
Demographic Variables								
Dummy for metropolitan area	-0.343	0.323			**-0.750**	**0.016**	-0.474	0.289
Proportion of population < 5 years					**-7.032**	**0.009**		
Proportion of population age 20 - 65 years			**25.836**	**0.001**				
Proportion of population >65 years			8.543	0.145	**-20.231**	**0.001**	**27.654**	**0.001**
Proportion of White population							**-59.011**	**0.005**
Proportion of Black population			2.270	0.121			**-50.752**	**0.014**
Proportion of Hispanic population			**1.207**	**0.039**	**-1.844**	**0.000**	**-4.511**	**0.033**
Proportion of AIAN population	1.202	0.150			**0.586**	**0.020**		
Income per capita ($10,000)	**0.692**	**0.099**			-0.593	0.296	**-2.144**	**0.066**
Percentage of population in poverty			**-0.232**	**0.004**	**-0.110**	**0.099**	**-0.262**	**0.014**
Proportion of population using Medicare					**1.5818**	**0.040**		
Proportion of population using Medicaid							**2.177**	**0.052**
Nursing Variables								
# of RNs per 100 individuals					**-1.103**	**0.009**		
# of Med Records & Health Info Techs per 1,000 individuals					**1.942**	**0.008**		
# of hospitals per 10,000 individuals			**2.242**	**0.039**			**-4.656**	**0.000**
# of Hospices per 10,000 individuals	-1.035	0.454	0.696	0.450			**2.457**	**0.048**
Dummy for county having hospital with nursing school	**-1.210**	**0.061**			0.399	0.427	**2.457**	**0.048**
# of hospital full time personals per 10 individuals	1.176	0.469			-2.89	0.101		
# of nursing home full time personals per 1,000 individuals			**-0.550**	**0.038**				
Ratio of average RN salary to median income			**2.530**	**0.010**	**-1.877**	**0.018**	**-4.023**	**0.004**
Facility Variables								
Facility type	**-5.384**	**0.078**	**-22.06**	**<0.0005**	**9.801**	**0.022**	**63.513**	**0.001**
Total number of budgeted RN positions	**-0.130**	**0.092**	-1.946	0.121	1.834	0.438	**-2.491**	**0.012**
RN vacancy rate	**1.936**	**0.046**	**50.736**	**<0.0005**	**35.816**	**0.010**		
Total number of budgeted LPN positions	-0.854	0.115						
LPN vacation rate					14.321	0.114		
RN turnover rate	1.729	0.322			0.1987	0.291	**6.396**	**0.005**

McKelvey-Zavoina R^2 = 0.71

Recruiting Difficulty Thresholds

Very easy (1) to recruit if score	≤ -5.494
Easy (2) to recruit if score	≤ -4.429
Not difficult (3) to recruit if score	≤ -3.348
Difficult (4) to recruit if score	≤ -2.159
Very difficult (5) to recruit if score	> -2.159

C. Validation of North Carolina Results

To address some of the questions regarding the adequacy of the "difficulty recruiting" variable, project staff attempted to validate the reported difficulty with a series of follow-up calls to those facilities that reported the most and least difficulty recruiting RNs. A list of the ID codes for the top- and bottom-ranked facilities was given to the North Carolina Center for Nursing, which provided contact information for those facilities without linking them to the identifiers in order to preserve the confidentiality of the data provided on the original survey. Consequently, the validation process was partially "blind," with no one involved in the validation knowing whether the facility had original reported a very high or a very low level of recruiting difficulty. The interviewer asked for a retrospective evaluation of difficulty recruiting RNs in 2004 (the data year used in the analysis), and to control for the possibility that people would provide retrospective data based on the current situation, an assessment of the current difficulty recruiting RNs was also obtained. Results were then sent back to the North Carolina Center for Nursing, where names of facilities were stripped and original survey identifiers were reattached in order to compare original with retrospective responses.

The rank order correlation between the original data reported in 2004 and retrospective data obtained through the validation process was only 0.347, an indication that difficulty recruiting RNs was a less than ideal measure of shortage, even though the correlation was greater than would be expected by random chance ($p = 0.016$). Not only was the difficulty recruiting in 2004 from the interviews not highly correlated with current difficulty, but it also was not highly correlated with the original assessments made in 2004. Because this process provided only marginal support for the validity of this dependent variable, it was decided that subjective indicators of shortage were likely to be too highly influenced by personal judgments and biases of the person completing the survey (e.g., overall disposition, momentary mood) to justify using them as the basis for a shortage assessment and rating process.

D. Apply North Carolina Ordered Probit Coefficients in North Dakota

Another attempt to validate the models to use the North Carolina data on recruiting difficulty involved an attempt to apply these models to another state. The coefficients from the ordered probit models were applied to comparable data from North Dakota to compare predicted to actual reported recruiting difficulty. The coefficients from the North Carolina models proved to be poor predictors of reported difficulty in North Dakota.

This raised serious questions about the possibility of using coefficients from one state to predict or estimate the extent of shortages in another state. Although further investigation might reveal that coefficients from one state might be used in another state with similar demographic characteristics, interstate variations in health care and labor market environments seem to preclude nationwide use of a model constructed in only one state.

E. OLS Regressions for Vacancy Rates Using Combined Data from NC and ND

It was hypothesized that the relatively small sample size for models based solely on data from North Carolina might have contributed to the limited number of statistically significant coefficients, and that increasing the number of cases might yield better results. This hypothesis led to a final set of models in the study incorporating facility-level data by the study team and models based on a combined data set from both North Carolina and North Dakota. OLS

regression models were estimated to predict vacancy rates at facilities in those two states combined.

The hypothesis, in fact, proved to be true. Models based on the combined dataset (shown below in Tables 8 to 11) highlighted a greater number of statistically significant explanatory variables for RN vacancy rates than models for either state alone. The overall explanatory power of these models remained only moderate, however, with much unexplained variation in vacancy rates. The long-term care model, in particular, had very limited explanatory power (R^2 = 0.238; Table 10). Furthermore, these models continued to rely heavily on facility-specific data that would be difficult to obtain at the national level.

Table 8. OLS Coefficient Estimates for RN Vacancy Rates in Hospital Settings for Combined NC & ND Model

Independent Variable	Estimate	Std Err	t-stat	p-value
Intercept	-0.7335	0.3863	-1.899	0.061
Dummy for North Dakota	0.0155	0.0203	0.7619	0.448
Dummy for metropolitan area	0.0239	0.0194	1.2323	0.222
Income per capita ($10,000)	0.0327	0.0270	1.2096	0.230
Proportion of Hispanic population *10	-0.0226	0.0358	-0.629	0.531
Total Medicare inpatient days per Pop	**-0.0728**	**0.0173**	**-4.219**	**0.0001**
Proportion of population < 5 years *10	0.2115	0.1239	1.7076	0.092
Proportion of population >65 years	**0.8423**	**0.3865**	**2.1792**	**0.032**
Proportion of population age 20 - 65 years	0.5597	0.4809	1.1639	0.248
# Full time RNs per 100 individuals	**0.0828**	**0.0392**	**2.1113**	**0.038**
Ratio of average RN salary to median income	0.0739	0.0440	1.6775	0.098
Number of budgeted RN positions	-0.0021	0.0021	-1.008	0.317
RN turnover rate	**0.2252**	**0.0655**	**3.4395**	**0.001**
LPN vacancy rate	**0.1661**	**0.0523**	**3.1785**	**0.002**
LPN turnover rate	0.0048	0.0159	0.3003	0.765

R^2 = 0.400

Table 9. Coefficient Estimates for RN Vacancy Rates in Home Health Setting for Combined NC & ND Model

Independent Variable	Estimate	Std Err	t-stat	p-value
Intercept	**-0.5407**	**0.2488**	**-2.174**	**0.032**
Dummy for North Dakota	-0.0811	0.0450	-1.801	0.075
Dummy for county w/ hosp w/ prof nursing schl	0.0945	0.0699	1.3522	0.180
Income per capita ($10,000)	**0.0789**	**0.0329**	**2.3969**	**0.018**
Proportion of Hispanic population x 10	-0.0966	0.0606	-1.593	0.114
# Hospitals per 10,000 individuals	-0.0240	0.0217	-1.105	0.272
# Med records & health info techs per 1,000 Pop	0.0607	0.0579	1.0491	0.297
Proportion of population < 5 years x 10	0.4565	0.2532	1.8031	0.074
Proportion of population >65 years	**1.1646**	**0.5227**	**2.2279**	**0.028**
Number of budgeted RN positions	-0.2623	0.1473	-1.781	0.078
RN turnover rate	**0.1234**	**0.0360**	**3.4259**	**0.001**
LPN vacancy rate	**0.1937**	**0.0687**	**2.8196**	**0.006**
Number of budgeted LPN positions	0.6455	0.5568	1.1593	0.249

$R^2 = 0.346$

Table 10. Coefficient Estimates for Long-Term Care Setting for Combined NC & ND Model
(Dependent Variable is RN Vacancy Rate)

Independent Variable	Estimate	Std Err	t-stat	p-value
Intercept	0.2447	0.2397	1.0207	0.309
Dummy for North Dakota	0.0392	0.0337	1.1655	0.246
Income per capita ($10,000)	-0.0324	0.0380	-0.8517	0.396
Proportion of Hispanic population *10	0.0567	0.0443	1.2798	0.203
Proportion of population < 5 years *10	-0.2825	0.1789	-1.5792	0.116
Proportion of population >65 years	-0.5939	0.3957	-1.5007	0.136
# Full time RNs per 100 individuals	0.0343	0.0459	0.7470	0.456
Ratio of average RN salary to median income	0.0389	0.0569	0.6834	0.495
Number of budgeted RN positions	-0.1725	0.1351	-1.2765	0.204
RN turnover rate	0.0223	0.0199	1.1191	0.265
LPN vacancy rate	**0.3508**	**0.0754**	**4.6521**	**0.000**
Number of budgeted LPN positions	**0.4843**	**0.1735**	**2.7920**	**0.006**

$R^2 = 0.238$

Table 11. Coefficient Estimates for Public Health Setting for Combined NC & ND Model
(Dependent variable is RN Vacancy Rate)

Independent Variable	Estimate	Std Err	t-stat	p-value
Intercept	0.0755	0.1079	0.6999	0.486
Dummy for North Dakota	**-0.0848**	**0.0288**	**-2.9455**	**0.004**
Dummy for Cnty w/ Hosp w/ Prof Nursing Schl	0.0622	0.0708	0.8792	0.382
Proportion of AIAN population x 10	**0.0652**	**0.0226**	**2.8856**	**0.005**
Proportion of black population	0.1231	0.0937	1.3133	0.193
# Hospitals per 10,000 individuals	**0.0466**	**0.0239**	**1.9535**	**0.054**
# Hospices per 10,000 individuals	-0.0541	0.0293	-1.8447	0.069
Total Medicaid inpatient days per Pop	**-0.0910**	**0.0378**	**-2.4066**	**0.018**
Proportion of population < 5 years x 10	-0.2130	0.1284	-1.6591	0.101
Percentage of population in poverty	**-0.0081**	**0.0041**	**-1.9653**	**0.053**
Ratio of mean RN salary to median income	**0.1357**	**0.0428**	**3.1690**	**0.002**
RN turnover rate	0.0710	0.0421	1.6854	0.096

$R^2 = 0.389$

IV. Models and Analyses Based on Geographic Data

Given the practical and methodological shortcomings evident in the analyses using facility-level data, the project team shifted its attention to models based only on county-level data that were nationally available and frequently updated. This shift seemed justified theoretically as well, because the inability of a facility to recruit and retain RNs in a county with sufficient overall supply of RNs may be a result of organizational culture rather than a genuine shortage. Limiting analyses to easily obtainable county level data seemed to serve these ends better than further pursuit of a model incorporating facility-level data.

A. Limitations and Challenges

There are limitations and challenges to a method based solely on geographic factors. For one, patterns of RN employment and health service utilization often transcend county (and state) lines. Knowing where RNs and patients live does not necessary tell researchers where services were provided or received, and thus where shortages actually existed.

Furthermore, the use of county-level data can mask large differences in facilities within counties. This is particularly true in the largest metropolitan counties. For example, New York County (Manhattan) may not meet the criteria for worst county-level RN shortage, but this ignores the fact that some facilities within Manhattan have a much harder time recruiting RNs than others (e.g., public facilities, those located in neighborhoods perceived as unsafe). Geography-based methodologies also may not adequately account for special circumstances specific to facilities.

Regardless of whether a facility is in a large county or not, it may have extenuating circumstances. There may be adequate numbers of RNs in the county, for example, but it may still be difficult to recruit RNs to work with the homeless.

Supplementing geography-based models with other procedures can minimize some of these limitations. Primary care Health Professional Shortage Areas (HPSAs) are currently designated based on geography-level characteristics, on facility-level characteristics, or on service to special populations. A similar tiered process could be developed for nursing shortage designations. Geographic designations could also be supplemented with an application process that allows facilities to submit facility-specific data. Special rules could be established to address sub-county variations in large urban areas (e.g., certain facilities in counties with population greater than one million—public, in a HPSA, or in a high-poverty Census tract—might automatically qualify).

One thing that emerged clearly in the analyses of facility-level data is that certain types of facilities were disadvantaged in the competition for RNs relative to others. The current methodology for awarding nursing loan repayment funds is based on categories of facilities, and this could be preserved so that certain types of facilities continue to receive preference, but in combination with geographic designations. Geographic designations could also be combined with facility type, in recognition of the fact that certain types of facilities (e.g., long-term care) may face greater disadvantages than others (e.g., hospitals). Facilities located in shortage counties could be given priority based on facility type, or conversely, facilities within priority categories (e.g., disproportionate share hospitals, community health centers) could be given priority designations based on county-level shortages.

An application procedure would allow facilities that feel they have been unfairly disadvantaged by a county-level designation to submit facility-level data to document their situation. This

would ease the burden on HRSA because most designations would be based on geography, but would provide facilities with special circumstances with an opportunity to qualify.

B. Measuring RN Supply at the County Level

The counts of RNs by county were taken from the 2000 U.S. Census long-form data, which is a 1-in-6 sample of the U.S. population. These data gave RNs by county of residence, not employment, and were less accurate when the actual number of RNs in the county was low (due to sampling error), but this was probably the best source available for county-level counts of RNs nationally.

In larger counties, the sample size should be sufficiently accurate. But in smaller counties, sampling error could have the effect of either undercounting or overcounting RNs. One person in the sample represents, on average, six people. If a small county has 102 RNs, theoretically one would expect 17 to be selected by the Census sample. If only 13 were in fact selected, the county would appear to have only 78 RNs, and might inappropriately qualify as a shortage county. On the other hand, if 20 were selected, the county would appear to have 120 RNs, which might prevent it from qualifying as a shortage county. These kinds of sampling errors would be random and not systematic, so less populous counties should not be consistently advantaged or disadvantaged by the method.

It is important that any method used by HRSA be easily updated using existing sources of data. Updating the decennial U.S. Census data can only be done every ten years, which creates estimation problems that grow over time, especially for counties that are rapidly growing or shrinking. Starting in 2008, another option will become available when the Census Bureau's American Community Survey (ACS) begins to provide estimates for smaller areas using three-year moving averages. Although the ACS sample will be smaller than the Census long-form data, it will be larger than any other interim data set, which increases the potential for sampling errors in small counties. Each person sampled in the ACS in one year will represent more than 100 people.

C. Adjusting for Commuting

Estimates of where RNs live were inadequate measures of supply because in some areas commuting inflows or outflows were very substantial. For example, only 16% of workers in New York County in 2000 actually resided in New York County. Using numbers of RNs living in New York County would thus substantially overestimate the degree of shortage in that county.

The U.S. Census Bureau provides data collected in the decennial census on commuting flows between every pair of counties in the U.S. From these data, commuting inflow was estimated based on the percentage of persons employed in county who lived in a different county, and commuting outflow was calculated based on the percentage of employed residents of the county who worked in a different county. These rates of county inflow and outflow were applied to RNs on the assumption that RN commuting patterns were not different from commuting patterns overall. (Preliminary analyses did not indicate that RNs were any more or less likely to work outside of their county of residence.)

D. Geography-Based Methods

There are a number of ways to conceptualize and measure RN supply at the county level, ranging from simple to sophisticated. All of the methods below were calculated using RN supply data adjusted for commuting patterns.

1. RNs to Population Ratio Method

This method is based upon the assumption that RNs should be evenly distributed across the U.S. in direct proportion to population (e.g., that 70 people in Los Angeles County, California require the same number of RNs as the 70 people who make up the entire population of Loving County, Texas). The estimated number of RNs required in a county is calculated based on population need rather than demand for RNs created by the existing health care infrastructure, and assumes that people receive nursing services where they live.

This ratio is very simple to compute (#RNs/#Population) and the data needs are also relatively clear. On the other hand, this ratio is also very crude, ignoring actual use of services (i.e., where people actually receive care), and demographic variations in health care needs (e.g., the greater needs of the older adults).

2. RNs to Adjusted Population Method

The project team explored two methods of adjusting the population. The first was based on rates of primary care utilization by gender and age (with weights based on the new primary care HPSA methodology) and the second was based on rates of utilization of multiples types of services estimated on age alone (with weights based on age-specific utilization rates for different types of services, gleaned from a variety of sources, most commonly *Health, United States, 2005*).

Because it accounts for population demographics, this method, which assumes that age-specific patterns do not vary across counties, should more accurately reflect population need than a simple RN to population ratio. However, this method, like the first, is based on estimated need for RNs rather than estimated demand for RNs.

3. RN to Physician Ratio

Both previous methods fail to account for the location of health care infrastructure. Regardless of the needs of the population, if an area has no health care employers to hire RNs there is no labor market demand for RNs and therefore no shortage. Places with more health care employers should, however, have more physicians, so physician supply can be used as a crude proxy for RN employer demand.

On the other hand, the net effect of this method is that areas that are short on both physicians and RNs appear comparable to areas that have surpluses of both physicians and RNs if the ratio is similar. This is particularly concerning because physician shortage areas may have the greatest need for RNs to provide basic primary care. This raises the RN shortage standard for exactly those counties—they must be short of RNs relative to the number of physicians when they are already short of physicians.

4. County Cluster Adjustments

All of the previous methods discussed ignore the flow of patients between adjacent counties to receive health care. An attempt was made to adjust for this by recalculating the previous ratios based on county clusters (RN, population, and/or physician counts summed for each county and its contiguous counties). The effect of this adjustment was higher shortage scores for nurse-poor counties surrounded by other nurse-poor counties, compared to nurse-poor counties surrounded by nurse-rich counties. This was theoretically appropriate in that it accounts for the unavailability of RNs in neighboring counties as well as in the county of residence.

This method showed some promise, but it still did not address some of the fundamental problems of the previous ratio methods. Furthermore, it did not account for the effects of multiple counties drawing on each others' resources. For example, it is tempting to say that County A's shortage isn't really so bad because it is bordered on the west by County B, which has a surplus of RNs. The situation of both County A and County B would be accounted for in County A's county cluster, but what would not be accounted for is the possibility that County B is bordered on the west by County C, which is also short of RNs and draws on County B's resources. County B's surplus may be sufficient to share between its own population and County A's population, but not between its own population, County B's population, *and* County C's population.

5. Cross-County Patient Flow Adjustments

Another attempt to adjust for the flow of patients between counties involved adjusting population figures based upon commuting flows. This assumed that the flows for seeking health care services were similar to those for commuting in general, and that areas that drew more commuters had more health care infrastructure and would also draw more health care consumers. Unfortunately, it was not clear that this is a reasonable assumption. It seemed likely to be true for many counties, but may not be true for some (particularly counties with large outflows of "extreme commuters" who travel more than sixty minutes to their jobs).

After reviewing the various versions of these ratio models, it was unclear whether county clusters or adjustments for cross-county patient flows were consistently an improvement on base ratios. It was concluded that an ideal method should use actual measures of health care utilization rather than attempting to estimate patient flows.

6. Factor Analysis of Nursing Shortage Indicators

A more sophisticated attempt to create a typology of counties based on the RN labor market involved factor analysis, an advanced statistical technique used to collapse a large set of characteristics of objects (counties in this case) into a smaller set of "factors" that represent different aspects of the objects. In this case, different characteristics of counties related to the supply of and demand for nurses (e.g., #RNs per capita, per capita income) load into different factors that represent different aspects of the supply and demand for nurses (e.g., a factor related to the economic conditions in the county).

This technique identified three broad factors relevant to nursing shortages at the county level: RNs relative to infrastructure (demand); RNs relative to population (need); and economic conditions. These factors are shown in Table 12.

Table 12. Standardized Factor Analysis Coefficients Related to Nursing Shortages in Counties in the U.S.

Variable	Factor 1	Factor 2	Factor 3
Metropolitan dummy variable	-0.025	-0.044	**0.188**
RNs per 1,000 individuals	0.003	**0.256**	0.012
RNs per 1,000 individuals < 5 years	-0.007	**0.259**	0.002
RNs per 1,000 individuals >=65 years	0.005	**0.109**	**0.122**
RNs per hospital bed	**0.213**	-0.052	0.048
RNs per MD	**0.136**	0.096	**-0.132**
RNs per 1,000 civilian labor force	0.020	**0.274**	-0.045
RNs per 1,000 inpatient days	**0.272**	-0.059	-0.058
RNs per 1,000 outpatient visits	**0.158**	0.007	-0.016
RNs per 1,000 emergency room visits	**0.134**	0.066	0.016
Infant mortality rate	0.028	0.019	**-0.140**
RNs per 100 Medicare inpatient days	**0.278**	-0.053	-0.038
RNs per 100 Medicaid inpatient days	**0.220**	-0.018	-0.069
Median household income ($10,000)	-0.027	-0.091	**0.310**
Percent persons in poverty	0.037	0.052	**-0.297**
Unemployment rate	0.064	-0.037	**-0.151**
Percentage of manufacturing workers	0.057	**-0.102**	0.036
Percentage of health service workers	-0.041	**0.232**	-0.168
Percentage of Blacks and Hispanics	0.010	-0.053	**-0.098**
Percentage of AIAN	0.020	0.061	**-0.119**

Note: The three factors can explain 50.3 percent of total variation of all variables

Using factor analysis, a typology of eight categories was created based upon their scores on the three dimensions (Table 13). Using this approach, the counties with the greatest shortages were low on all three factors (i.e., category 111), indicating high levels of unmet need, unmet demand, and socioeconomic disadvantage. The counties with the least shortages were high on all three factors (i.e., category 222).

Table 13. Numbers and Percentages of Counties in Factor Analysis Categories, by Census Division

Census Division	Category[(a)]									Total
	Missing	111	112	121	122	211	212	221	222	
East North Central	70	14	40	23	50	20	111	34	75	437
	16.0%	3.2%	9.2%	5.3%	11.4%	4.6%	25.4%	7.8%	17.2%	100%
East South Central	80	38	22	32	20	71	35	46	20	364
	22.0%	10.4%	6.0%	8.8%	5.5%	19.5%	9.6%	12.6%	5.5%	100%
Middle Atlantic	13	3	13	26	48	2	8	9	28	150
	8.7%	2.0%	8.7%	17.3%	32.0%	1.3%	5.3%	6.0%	18.7%	100%
Mountain	67	49	55	29	18	13	22	22	5	280
	23.9%	17.5%	19.6%	10.4%	6.4%	4.6%	7.9%	7.9%	1.8%	100%
New England	4	1	2	2	25	1	2	4	26	67
	6.0%	1.5%	3.0%	3.0%	37.3%	1.5%	3.0%	6.0%	38.8%	100%
Pacific	25	20	30	10	12	18	29	15	5	164
	15.2%	12.2%	18.3%	6.1%	7.3%	11.0%	17.7%	9.2%	3.0%	100%
South Atlantic	167	66	58	55	41	66	56	44	36	589
	28.4%	11.2%	9.8%	9.3%	7.0%	11.2%	9.5%	7.5%	6.1%	100%
West North Central	147	34	50	62	89	21	35	97	83	618
	23.8%	5.5%	8.1%	10.0%	14.4%	3.4%	5.7%	15.7%	13.4%	100%
West South Central	103	102	22	52	18	72	30	58	12	469
	22.0%	21.8%	4.7%	11.1%	3.8%	15.4%	6.4%	12.4%	2.6%	100%
Total	**676**	**327**	**292**	**291**	**321**	**284**	**328**	**329**	**290**	**3,138**
	21.5%	**10.4%**	**9.3%**	**9.3%**	**10.2%**	**9.0%**	**10.4%**	**10.5%**	**9.2%**	**100%**

Note: [(a)] An example of how to interpret the category: 121 means F1<median, F2>median, F3<median

This analysis showed promise in theory, but was based on primary care utilization, with no basis for examining long-term care, home health care, or public agency services, and no way of reflecting variations in staffing intensity across types of care. While acute care hospitals are the primary driver of RN demand, the focus on hospital care does not make this method applicable to counties without hospitals.

V. Preferred Method

Staff members of the Center for Health Workforce Studies have been working with the Lewin Group on the update of the HRSA Nurse Supply Model (NSM) and Nurse Demand Model (NDM). Although the exact analyses included in the NDM could not be replicated at the county level due to data constraints, the basic logic employed in the NDM was very useful in thinking about demand for RNs.

The project staff decided to apply a simplified version of the NDM logic to: 1) estimate health care utilization in different settings for counties (e.g., inpatient days); 2) estimate current national RN staffing by setting (e.g., RNs working in inpatient units); 3) calculate national RN staffing intensity for each setting (e.g., RNs per inpatient day); 4) apply national RN staffing intensity ratios to measures of utilization for each county; and 5) sum estimate demand for each setting to produce overall RN demand for individual counties. Each step is summarized briefly below.

A. Estimate Health Care Utilization

The data on county-level health care utilization primarily came from the Area Resource File (ARF). The ARF included data on:

- Short-term inpatient days (non-psychiatric hospitals)
- Long-term inpatient days (non-psychiatric hospitals)
- Psychiatric hospital inpatient days
- Nursing home unit inpatient days (hospitals)
- Outpatient visits (non-emergency)
- Emergency department visits

The number of (non-hospital) nursing home residents in a county was obtained from the 2000 Census. This was based on the Census short-form data, which is theoretically obtained from 100% of the U.S. population.

The number of home health patients per county was estimated using the age and gender distribution of the population, based upon national age-specific and gender-specific utilization rates from the Centers for Disease Control and Prevention (CDC).

Although this estimate was based upon population characteristics rather than actual use of services, home health patients by definition were receiving services where they live, so this was somewhat less problematic than estimating other types of utilization based upon population characteristics.

B. Estimate Current National RN Staffing

Data for current levels of RN staffing by setting were taken from the 2000 NSSRN, which included data on the number of RNs employed in the following types of care:

- Short-term inpatient (non-psychiatric hospitals)
- Long-term inpatient (non-psychiatric hospitals)
- Psychiatric inpatient (non-Federal)

- Nursing home unit (hospital)
- Outpatient (non-emergency)
- Emergency outpatient
- Non-hospital nursing home
- Home health
- Nurse education
- Public/community health
- School health
- Occupational health
- Non-hospital ambulatory care
- Other nursing care

These numbers were combined with the national utilization data described above to compute national levels of RN staffing intensity for the various types of care. The national ratios are shown in the table below.

Table 14. National RN Staffing Ratios by Type of Care

Short-Term Hospital, Inpatient	4.97 RNs / 1,000 Inpatient Days
Long-Term Hospital, Inpatient	11.66 RNs / 1,000 Inpatient Days
All InPt Units in Psychiatric Hospitals	1.45 RNs / 1,000 Inpatient Days
Nursing Home Unit in Hospital	0.48 RNs / 1,000 Inpatient Days
Other Nursing Home	0.07 RNs / NH Resident
Nurse Education Programs	0.02 RNs / Active RN
Public/Community Health RNs	5.28 RNs / 10,000 Pop
School Health (excl. college)	12.5 RNs / 10,000 Pop 5-17
Occupational Health	2.07 RN / Pop 18-64
Home Health	0.10 RNs / Home Health Patient
OutPt or Diagnostic Units in All Hosp	0.14 RNs / 1,000 Visits
EDs in All Hospitals	0.86 RNs / 1,000 Visits
Ambulatory Care	7.43 / 10,000 Pop
Other	4.1 RNs / 10,000 Pop

C. Estimating RN Demand by County

These national staffing ratios were then applied to the utilization rates for each county. For example, the national ratio was 4.97 RNs working in hospital inpatient units per inpatient day. If County A has 12,000 inpatient days per year, their demand for RNs in inpatient units is estimated at 59.6 (4.97 x [12,000/1,000]).

Overall RN demand for the county was obtained by summing RN demand in the county across all settings. (This procedure also opens the possibility of comparing setting-specific demand to setting-specific supply if data on RN supply by setting are available at the county level.)

D. Use Supply of RNs to Estimate RN Shortages

RN shortages were thus measured as follows:

RN shortage = Estimated demand for RNs in the county minus the number of RNs in the county (adjusted for commuting patterns).

Raw shortage estimates were then standardized as a percent of demand. The results are presented for all counties in the U.S. in the map in Figure 8. The counties with the greatest shortages are shaded black. The full technical report has separate maps for all 50 states, along with a table with the numerical scores.

This method has advantages over any of the other methods examined in this study, especially in relation to the guiding principles initially proposed for the study:

- It uses nationally available data that is periodically updated.
- It uses actual health care utilization patterns by county.
- It accounts for multiple types of care (including non-clinical services).
- It accounts for differences in RN staffing intensity across settings. Some limitations persist, however. It does not account for county or state variations in health systems (e.g., HMO penetration, use of LPNs), and does not account for patient acuity within types of care. Furthermore, it assumes current RN staffing levels were adequate at the national level in 2000, which may not have been the case.

The NDM uses factors such as HMO penetration and LPN staffing in regressions to adjust estimated staffing intensity and make it specific to each county rather than applying national ratios. A similar procedure might eventually be used to do the same thing here.

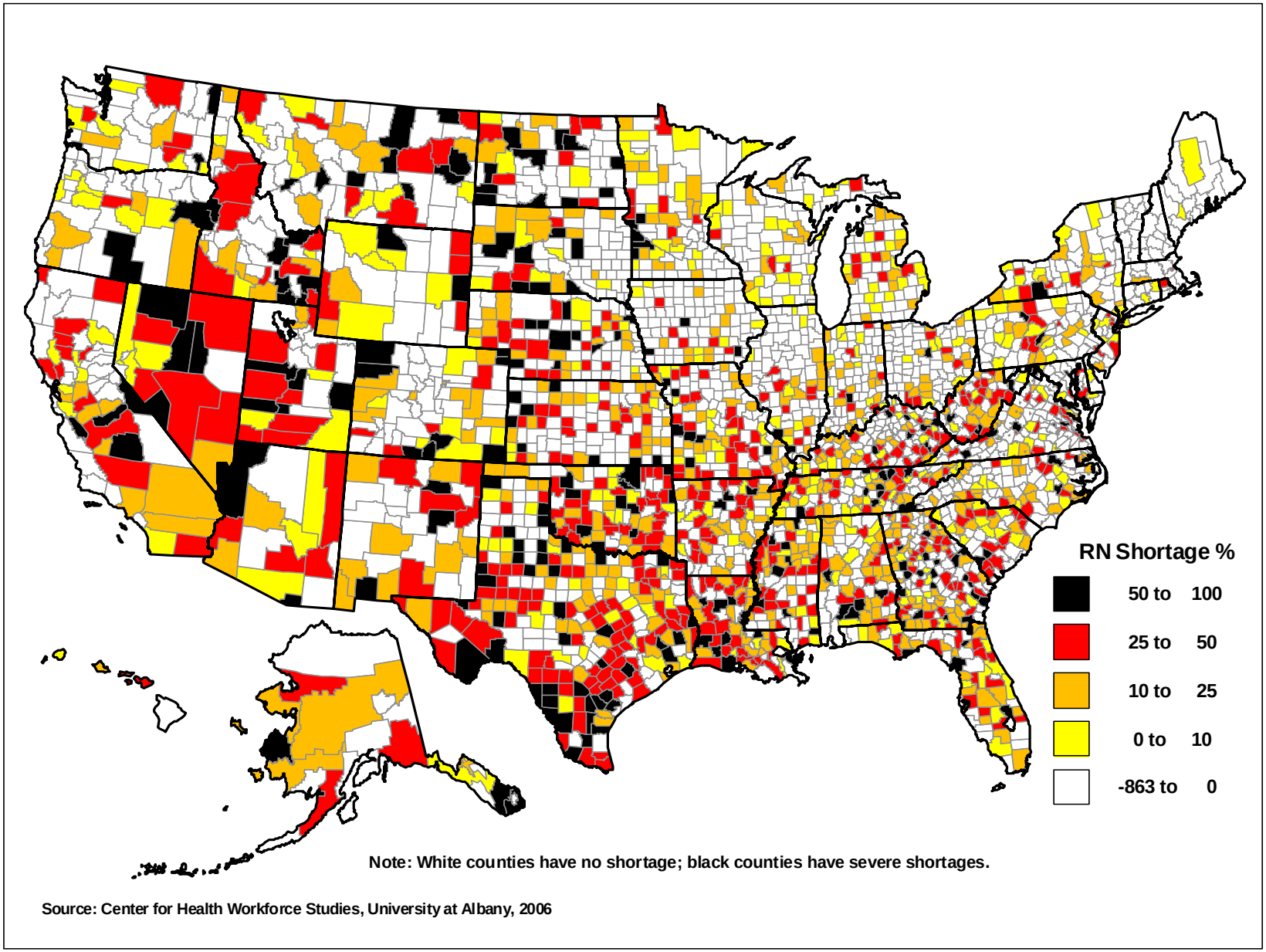
VI. Outstanding Issues Related to the Preferred Method

There were a number of outstanding issues related to the Preferred Method that required a closer look. While fully addressing some of these issues was beyond the scope and timeframe of the current study, limited analyses were performed to investigate some potential avenues for improving the Preferred Method to address these shortcomings. Two of the most important analyses are summarized below.

A. The Problem of Patient Acuity

Study staff did attempt to correct for patient acuity in hospital settings because hospital patient acuity will differ across counties in ways that may systematically disadvantage counties with major medical and trauma centers. ARF data was used to measure the number of surgeries and to estimate percent of inpatient days spent in the ICU. NSSRN data was used to estimate the number of RNs working in ICU units and operating rooms, and the steps in the earlier model were followed to obtain estimates of demand for operating room RNs and ICU RNs calculated separately from other hospital RNs. It was a problem, however, that many counties lacked accurate ICU bed data (especially large urban counties). Perhaps, as a result, this adjustment did not have the expected effects on RN demand estimates. In fact, it often resulted in lower rather than greater estimated demand for RNs.

Figure 8. Estimated RN Shortage Percentages for Counties in the U.S.



Note: White counties have no shortage; black counties have severe shortages.

Source: Center for Health Workforce Studies, University at Albany, 2006

This additional adjustment for ICU and surgical services is theoretically important in that it is one of the few possible adjustments for acuity. While the quality of currently available data may prohibit incorporating the adjustment into a national methodology, hospitals that feel that their higher patient acuity has disadvantaged them in the standard process could potentially submit ICU and surgical data through an application and appeals process.

B. RN Commuting Patterns

The original version of the Preferred Method assumed that RN commuting patterns were similar to those of the overall workforce. This is generally true in the aggregate—RNs are no more or less likely than other workers to work outside the county where they live. At the county level, however, RN commuting patterns sometimes varied dramatically from the patterns for all workers. Additional analyses summarized below revealed that RN commuting patterns depended more on county characteristics than on characteristics of RNs (e.g., gender, income level, etc.). These analyses were based on data from New York, North Carolina, and Mississippi, states for which data were available on both county of residence and county of employment.

1. Models to Predict RN Commuting Patterns

To help identify the factors related to RN commuting patterns, models based on county characteristics were developed. The commuting patterns of all workers had been, on average, a good proxy for the commuting patterns of RNs, so this was retained as one independent variable. This variable should reflect many of the primary drivers of commuting behavior (e.g., relative wages, cost of living, etc.). Another independent variable was assumed to be opportunities for RN employment in a particular county. This was measured by the extent to which the number of RNs living in a county compared to the estimated demand for RNs in that county (based on infrastructure and service use). Counties where resident RNs were in short supply relative to service use were expected to be net importers of RNs, while counties where resident RNs were more than sufficient for the county's health care needs were expected to be net exporters of RNs.

Other factors included in the analysis were whether the county was a whole-county HPSA, the county's major industry, and whether the county was a persistent poverty county. The rural-urban characteristics of the county (population size, proximity to a metropolitan area) were also accounted for, although these did not prove as crucial as expected (probably because they did not affect RN commuting any differently than overall commuting, which was already controlled for).

The intercept for the model was 0.495, indicating that if all other variables had a zero value, each resident RN would be equal to 0.495 RNs working in the county. The coefficient for overall commuting was 0.601, indicating that for every one percent increase in net incommuting, there would be a 0.601 unit increase in RN incommuting. The supply of resident RNs relative to estimated demand was negatively related to net incommuting (-0.148).

The percent of the population living in an urban area within the county was positively related to RN incommuting (0.001), but this was not statistically significant (p=0.059). For every increase of 10,000 population, RN incommuting increased by 0.0012. Whole-county HPSA status decreased net RN incommuting (-0.157), as did persistent poverty county status (-0.158) and dependence on manufacturing (-0.09).

There was an interesting interaction effect between population size and persistent poverty status: being a persistent poverty county had a greater depressant effect on RN incommuting in small population counties than in large population counties. This model had an adjusted R^2 of 0.702.

Table 15 shows the results of models estimated separately for groups of counties based on their relationship to a metropolitan area (part of a metropolitan area, adjacent to a metropolitan area, or not adjacent to a metropolitan area). Although the results show some potential to fine-tune the RN incommuting estimates for different groups of counties, the differences in the model coefficients were not dramatic. Differences in model fit were substantial, however. The model for counties not adjacent to a metropolitan area was the best fitting model (adjusted $R^2 = 0.842$). The model for metropolitan counties also fit well ($R^2 = 0.805$). The model for non-metropolitan counties adjacent to metropolitan areas, however, explained less variation ($R^2 = 0.509$).

Table 15. Ordinary Least Squares Regression Coefficients Predicting RN Incommuting, By Type of County

	All Counties		Metro Counties		Counties Adjacent to Metro Area		Counties Not Adjacent to Metro Area	
	B	S.E.	B	S.E.	B	S.E.	B	S.E.
(Constant)	0.495***	0.059	0.545***	0.089	0.357*	0.167	0.536**	0.165
All worker incommuting	0.601***	0.050	0.563***	0.057	0.559**	0.187	0.664***	0.146
RN Surplus	-0.148***	0.017	-0.221***	0.034	-0.094**	0.025	-0.227***	0.043
Pct Urban	0.001	0.001	0.0009	0.001	0.003	0.002	0.003	0.001
Whole-County HPSA (1=yes)	-0.157***	0.037	-0.117	0.059	-0.153*	0.068	-0.151**	0.046
Mfg Dependent (1=yes)	-0.009**	0.028	-	-	-	-	-0.134**	0.041
Persistent Poverty (1=yes)	-0.158**	0.053	-0.287*	0.133	-	-	-0.100	0.058
Total Population (*10,000)	0.001*	0.000	0.003	0.000	-	-	-	-
Housing Stress (1=yes)	-	-	0.050	0.065	-	-	0.100*	0.046
Service Dependent (1=yes)	-	-	-	-	0.361**	0.118	-	-
Retirement Destination (1=yes)	-	-	-	-	-0.215	0.117	-	-
Total Pop x Persistent Poverty	0.033**	0.000	0.049*	0.000	-		-	-
Total Pop x Housing Stress	-	-	-0.002	0.000	-		-	-
Pct urban x Persistent Poverty	-	-	-	-	-		-0.003	0.002
Adjusted R^2	0.702		0.805		0.509		0.842	

* p ≤ 0.05 ** p ≤ 0.01 *** p ≤ 0.001

Interestingly, the most accurate method for estimating commuting varied by county type. In metro counties, the commuting flow of all workers was the most accurate estimate of the three models 39% of the time; while in counties adjacent to metro areas, the model for all counties was the most accurate 47% of the time; and in counties not adjacent to metro areas, the best estimate was the RUCC-specific estimate 51% of the time.

In many cases, however, the "best" estimate was better than the "next best" estimate by only a point or two. When the variable used to evaluate was the percent of the time that an estimate differed by more than 10% from the actual RN commuting value, the all-county estimate was accurate more often for metro and adjacent-to-metro counties, while non-adjacent-to-metro counties did best when the RUCC-specific estimate was used. It was never preferable to use the overall commuting pattern.

Table 16. Percentage of Cases in Which Estimate Differs From Actual by More Than 10%

Group (N)	Statistic	All-Counties Regression Estimate Off By > 10%	RUCC-Specific Regression Estimate Off By > 10%	Estimate Eased on Commuting of All Workers Off By > 10%
Metro County (93)	Mean	61.3%	71.0%	61.3%
	Std Dev	49.0%	45.6%	49.0%
Adjacent to Metro County (86)	Mean	68.6%	79.1%	76.7%
	Std Dev	46.7%	40.9%	42.5%
Not Adjacent to Metro County (65)	Mean	63.1%	56.9%	66.2%
	Std Dev	48.6%	49.9%	47.7%
Total (244)	Mean	64.3%	70.1%	68.0%
	Std Dev	48.0%	45.9%	46.7%

2. Using Commuting Patterns to Estimate RN Supply

Because the evaluation of the estimates using data from the counties from which they were derived was somewhat tautological, it was decided to assess whether these corrections brought estimates of RN employment by county closer to actual employment data in other states (for which real commuting patterns were not available). The states used in this preliminary validation process were Tennessee, Texas, Pennsylvania, South Dakota, and some counties in Iowa.

When compared to actual counts of RNs working in particular counties, the revised commuting adjustments did little to improve the supply estimates. The estimated supply was closer to the actual supply on average when overall commuting was used as the adjustment factor. There was some variation by the Rural Urban Classification Code (RUCC): the estimate of commuting based on the all-county model produced somewhat lower average differentials than other estimates for counties adjacent to metro areas, and somewhat lower absolute average differentials for counties not adjacent to metro areas.

It is important to remember that the accuracy of the commuting estimates is only one source of error in estimated supply of RNs working in a county. Another source of error is in the estimated numbers of RNs living in the counties in which they work. It may be possible in future work to estimate confidence intervals around commuting estimates and estimates of supply of resident RNs, and for shortage designations to be based on the lower of the two estimates of RNs based on the two confidence intervals. This would remove some of the disadvantage potentially faced by rural areas due to sampling errors, although it would increase the likelihood of designating some counties that should not qualify.

VII. Study Recommendations

The study identified six recommendations for HRSA and other organizations to consider as they attempt to identify facilities with critical shortages of RNs accurately and reliably. Several of these recommendations are presented below.

1) Of the methods examined in this study, the Preferred Method outlined in this report is the best choice for assessing the severity of nursing shortages in counties in the U.S. It meets more of the desirable criteria identified by the study advisory panels and it can be implemented with currently available data. Additional steps outlined below could further improve the effectiveness of this method.

2) Additional review and validation of the Preferred Method is required by stakeholders who would be affected by its implementation. Ideally, this validation should take place in a representative sample of states, counties, and facilities across the U.S., and would address the following kinds of questions:

   - Are facilities and counties classified correctly by the method? Is the method biased in favor of or against a type of facility, type of community or county, or region of the country? If so, how should the bias be addressed or overcome?
   - Are the basic data required to support the method both available and accurate for all regions and states in the U.S.? How should sampling errors for small rural counties be addressed?
   - How should facilities that have nursing shortages primarily due to persistent poor management be dealt with in the method? What criteria should be used to identify facilities with poor management and should their identities be made public?
   - Should the method be supplemented by some sort of appeals process to permit a facility with a genuine shortage to qualify for NELRP and NSSP even though the method does not place it in a sufficiently severe shortage category?
   - Should the method identify just enough "severe shortage" counties and facilities to allocate all NELRP and NSSP recipients and other related funds based on nursing shortages? Or should it identify extra facilities to provide flexibility to account for other factors?

3) More accurate estimates of RN employment and supply should be developed at the county level. This may not require new data collection if appropriate refinements can be made to the sampling frames for existing datasets, especially the NSSRN.

4) More research should be conducted on factors related to the demand for RNs, including HMO penetration, alternate service delivery models, the use of LPNs and other types of staff, and new diagnostic and treatment technologies. Factor analysis may be a fruitful avenue for additional research. Another promising avenue for research will open up when the revised HRSA Nursing Demand Model becomes available sometime in 2007.

5) More research should be conducted on factors related to the supply of RNs including RN commuting patterns, how very rural communities can recruit and retain RNs, how inner-city

facilities can recruit and retain RNs, etc. The revised HRSA Nursing Supply Model is expected to be available sometime in 2007 and offer expanded research opportunities.

6) Because shortcomings in available data and extenuating circumstances might cause certain facilities to be assigned the wrong shortage designation, a formal protocol by which facilities can appeal and correct their shortage designation should be developed. The development process should consider a variety of appeal options including single facility designation changes and blanket designation changes for entire classes of facilities.

VIII. References

1. Health Resources and Services Administration. (2001). *The Registered Nurse Population: Findings from the National Sample Survey of Registered Nurses, March 2000.* Washington, DC: Division of Nursing, HRSA, USDHHS.
2. National Council of State Boards for Nursing. Nursing Regulation and Licensure Statistics. http://www.ncsbn.org/public/regulation/licensure_stats.htm https://www.ncsbn.org/index.htm
3. Center for Health Workforce Studies. (2006). *New York Registered Nursing Graduations, 1996-2004.* Rensselaer, NY: CHWS. http://chws.albany.edu
4. Berlin L., Bednash G., & Stennett J. (2001). *Enrollment and Graduations in Baccalaureate and Graduate Programs in Nursing.* Washington, DC: American Association of Colleges of Nursing.
5. National Center for Health Workforce Analysis. (2002). *Projected Supply, Demand, and Shortages of Registered Nurses: 2000-2020.* Washington, DC: Bureau of Health Professions, Health Resources and Services Administration, USDHHS.
6. American Hospital Association Commission on Workforce for Hospitals and Health Systems. (2002). *In Our Hands: How Hospital Leaders Can Build a Thriving Workforce.* Chicago, IL: AHA.
7. U.S. Government Accounting Office. (2001). *Nursing Workforce: Emerging Nurse Shortages Due to Multiple Factors.* http://www.gao.gov/new.items/d01944.pdf
8. Needleman J., Buerhaus P., Mattke S., Stewart M., & Zelevinsky K. (2002). Nurse-staffing levels and the quality of care in hospitals. *New England Journal of Medicine*; 346: 1715-1722. http://content.nejm.org/
9. Kovner C.T., Jones C., Zhan C., Gergen P., & Basu J. (2002). Nurse staffing and post-surgical adverse events: An analysis of administrative data from a sample of U.S. hospitals, 1990-1996. *Health Services Research, 37,* 611-629.
10. Kovner C.T. & Gergen P.J. (1998). Nurse staffing levels and adverse events following surgery in U.S. Hospitals. *Image: The Journal of Nursing Scholarship, 30,* 315-321.