Methods for Identifying Facilities and Communities with Shortages of Nurses

February 2007

Methods for Identifying Facilities and Communities with Shortages of Nurses

February 2007

Preface

This report summarizes the findings of a major 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 summary report is available that highlights those findings deemed most important for policy makers to consider.

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

List of Tables

List of Figures

I. Executive Summary

A. Study Background and Context

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 RNs. 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 six sections, each summarizing a different aspect of the study.

- Study Background and Context
- Methods, Models, and Analyses Using Facility Data
- Methods, Models, and Analyses Using Only Geographic Data
- Preferred Method
- Additional Analyses and Explorations
- Conclusions and Recommendations

The conclusions are designed to inform policy analysts 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 preliminary estimates of the supply and demand for RNs in counties and other jurisdictions, can also be found in the report.

1. 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 RNs 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 a standard set of indicators of critical shortages of RNs be developed and applied to all of the eligible settings 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 on the neediest facilities?
- How can true shortages of RNs at a facility be distinguished from shortages created by poor management practices?

An effective study must take all of these issues into account while researching and evaluating new methods to measure shortages of RNs. Ideally, a new method can be developed to support government programs that encourage new RNs to practice in facilities and communities with severe shortages. Such a method 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 shortage hospitals and counties throughout the United States and its territories. The Center used two separate methodologies, one to identify private, non-profit hospitals with shortages of RNs and the second to identify counties with shortages of RNs.

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 RNs.

2. 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.

a. 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 with the greatest needs.

b. 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.

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.

c. Guiding Principles

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.

d. 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.

B. Methods and Models Using Facility Data

All of the analyses using facility data are based on data sets from North Carolina and North Dakota. These datasets included a number of possible measures of nursing shortages 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 nine categories. This was an interesting variable because of in-depth discussions in the first advisory panel meeting about how true measures of a nursing shortage should be related to patient care and facility operations. Although subjective, this variable touched 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., nursing homes). 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 several 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 finding 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 responses 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.

1. Ordinary Least Squares (OLS) Regression Models

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 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-squared 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. That said, the results did reveal new insights about the supply of and demand for RNs. Thus the research findings should be of interest to students of the nursing workforce. A journal article on this aspect of the study is planned.

2. 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 NC 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 facility-specific models are summarized in detail later in the report.

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.

3. Validation of North Carolina Results

To address some of the questions regarding the adequacy of the “recruiting difficulty” variable, project staff conducted a formal validation of the “recruiting difficulty models” with a series of follow-up calls to those facilities that reported the most and least difficulty recruiting RNs. This “blinded” process was conducted with the cooperation of the North Carolina Center for Nursing (NCCN), 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. The interviewer asked for a retrospective evaluation of difficulty recruiting RNs in 2004 (the data year used in the analysis). 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.

The Spearman rank order correlation between the original data reported in 2004 and retrospective data obtained from 48 of 80 facilities through the validation process was 0.347 ($p = 0.016$), an indication that the difficulty recruiting RNs was a less than ideal measure of shortage. Not only was the difficulty recruiting in 2004 from the interviews not highly correlated with the original assessments made in 2004, but it also was not highly correlated with current difficulty.

Despite the fact that the correlation was statistically significant, the conclusion based on this validation process was 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 nursing shortage assessment and designation process.

4. Application of North Carolina Ordered Probit Coefficients in North Dakota

Another attempt to validate the recruiting difficulty models involved applying the results of the North Carolina models to another state. The coefficients from the NC ordered probit models were applied to comparable data from North Dakota to compare predicted to actual reported recruiting difficulty. The coefficients from the NC models proved to be a poor basis for predicting recruiting difficulty in ND.

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 some other state with similar demographic characteristics, interstate variations in health care and labor market environments seem to preclude nationwide use of a model constructed based on data from only one state.

5. 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 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 revealed 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 ($R^2 = 0.238$). Furthermore, these models continued to rely heavily on facility-specific data that would be difficult to obtain for a national shortage designation method.

C. Geography-Based Models

Given the practical and methodological shortcomings evident in the analyses using facility-level data, the project team shifted its attention to models based on only 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 models incorporating facility-level data.

1. 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 facilities with special circumstances would be given an opportunity to appeal disqualification based on geographic criteria alone.

2. Measuring the 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. Each person sampled in the ACS in one year will represent more than 100 people, and if three years of data are combined, one will represent about 33.

3. 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.)

4. Methods Using Only Geographic Data

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 described below were calculated using RN supply data adjusted for commuting patterns.

a. 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 healthcare 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).

b. 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 based 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.

c. 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 have shortages of both physicians and RNs may appear comparable to areas that have surpluses of both physicians and RNs if the ratios are similar. This is of particular concern because physician shortage areas may have the greatest need for RNs to help provide basic primary care services. 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.

d. 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 is theoretically appropriate in that it accounts for the unavailability of RNs in neighboring counties as well as in counties 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 really isn't 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.

e. 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 of patients seeking health care services were similar to those for commuting in general, and that areas that attracted more commuters had more health care infrastructure and would also attract more health care consumers. Unfortunately, it was not clear that this is always 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. Ultimately, it was concluded that an ideal method should use actual measures of health care utilization rather than attempting to estimate patient flows.

f. 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, a more 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 RNs (e.g., #RNs per capita, per capita income) load onto different factors that represent different aspects of the supply and demand for RNs (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. Based on the factor analysis results, a typology of eight categories was created based on a binary split of the scores on the three dimensions. 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).

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.

D. 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 develop a simplified version of the NDM model 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.

1. 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 U.S. 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.

2. 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 RN staffing for the various types of care.

3. 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.

4. Use Supply of RNs to Estimate RN Shortages.

RN shortages were then 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. A table showing the numerical results for all counties in the U.S. can be found in Appendix E. This table is presented as a series of maps for all of the states in Appendix F. The counties with the greatest shortages are shaded black.

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. The method 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. In fact, the new NDM model might be used directly to support this entire approach.

E. Additional Analyses and Explorations

Two suggestions were made at the final advisory committee meeting to improve the Preferred Method. Each is summarized briefly below.

1. Adjustments for Patient Acuity

Perhaps the greatest shortcoming of the Preferred Method is that it does not adequately account for patient acuity. This leads to underestimates of RN demand and need in counties with large medical centers with trauma units, which might be expected to have higher levels of patient acuity on average than small community hospitals. Related to this, larger hospitals may also have more patients admitted for complex surgeries and may require larger surgical staffs (including OR RNs) than their smaller counterparts.

Study staff performed a number of analyses to determine whether the Preferred Method could be improved by adjusting for patient acuity. These analyses included using more detailed categories of hospital beds, including medical and surgical intensive care beds, cardiac intensive care beds, neonatal intensive care beds, neonatal intermediate care beds, pediatric intensive care beds, burn care beds, other special care beds, and other intensive care beds. Such breakouts can be used to disaggregate inpatient days into ICU days and regular care days.

The net effect of this adjustment was to reduce the estimated nursing shortage for many counties, but to increase it for few. Unfortunately, this approach suffered from several limitations. Data on the numbers of beds in different categories were not available in the ARF for hospitals in about 10% of counties. In addition, bed type breakouts were not available for short-term non-general hospitals, which may also have ICUs and operating rooms.

Another limitation was that while RNs cannot be separated by general versus non-general short-term hospitals, so RNs in ICUs in both types of hospitals will be factored into the staffing ratio for ICU, but the inpatient days in short-term non-general hospitals cannot be adjusted down by parsing out the ICU bed days.

Despite these limitations, this adjustment has promise and should be considered as the theoretical standard, even though currently available data do not support its use in practice.

2. More Careful Analysis of 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. A number of models were developed to better understand RN commuting patterns. Among the independent, explanatory variables used in these models were:

- The commuting patterns of all workers;
- Opportunities for RN employment available in particular counties;
- Counties where resident RNs were in short supply relative to service use;
- Whether the county was a whole-county HPSA;

- The major industry in the county;
- Whether the county was a persistent poverty county; and
- The rural-urban characteristics of the county (population, proximity to a metro area).

The most accurate method for estimating RN commuting varied by county type. In metro counties, the commuting flow of all workers was the most accurate estimate of the three 39% of the time. In counties adjacent to metro areas, the model for all counties was the most accurate 47% of the time. In counties not adjacent to metro areas, the best estimate was the Rural Urban Classification Code (RUCC)-specific estimate 51% of the time.

In general, RN commuting patterns depended more on characteristics of counties than on characteristics of RNs (e.g., gender, income level, etc.). However, the "best" estimate was often better than the "next best" estimate by only a point or two.

F. 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 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. One promising avenue for research will open up when the revised Nursing Supply Model becomes available sometime in 2007.

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.

II. Study Background and Context

This report summarizes the findings of the various components of this study of methods for identifying facilities and communities with shortages of RNs. It documents the strengths and weaknesses of different methods for assessing the extent of shortages of RNs. 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.

A. Federal Initiatives to Address Nursing Shortages

In 2004, realizing that the current shortage designation process was too narrow in scope and that RN shortages were likely to worsen over the next 20 years, 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 agencies with critical shortages of RNs. The New York Center for Health Workforce Studies at SUNY Albany was selected to conduct this project.

There is growing recognition and efforts are underway to increase production of RNs and use incentives to target new graduates to facilities and agencies with the most critical shortages of RNs. However, there are issues that must be taken into account when assessing need and demand for RNs and identifying health care providers with the most critical shortages of RNs. These include:

- 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 included in this project?
- 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 there 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 on the neediest facilities?

- How can true shortages of RNs be distinguished from shortages created by poor management practices?

A careful review of the literature helped to inform the discussion of these and other related issues. Through the identification and review of existing methods and models for measuring health professional shortages, information on these issues will be obtained and shared with each of four expert panels, who are providing guidance for this project. It is unlikely that standard data sets on staffing will be available for all of the health care settings included in this project. Rather, data may be available for only some providers, e.g., the American Hospital Association nurse staffing data set for acute care facilities, or the Centers for Medicare and Medicaid Services (CMS) Online Survey, Certification and Reporting (OSCAR) data set for long-term care providers. The information on staffing for some types of health care providers may be less than adequate, or it may not be available at the national level.

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 RNs. It would also provide a better basis for monitoring RN shortages locally and nationally.

B. Study Overview

This study was conducted over a two-year period, starting in the fall of 2004, during which nine different research components were carried out. Each component is summarized in the body of this report in roughly the chronological order they were conducted during the study:

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.

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

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.

Project 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;
  - 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.

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.

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 used 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 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 RNs and the second to identify counties with shortages of RNs.

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 placement sites, based on the premise that they faced critical shortage of RNs.

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

The two demographic characteristics most relevant to shortages of RNs were gender and age. The gender mix of RNs was important because it reflected the size of the pool of potential candidates from which to recruit new RNs.

The age distribution was important because it dictated the numbers of existing RNs who will leave nursing in the future, creating a need to replace them in the workforce.

Table 1 provides estimates of the percentages of active RNs in the U.S. by gender and age group. 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. The table also reveals that 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. In fact, 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 new RNs 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	Percent
< 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%**
Percent	**6.1%**	**93.9%**	**2,396,329**

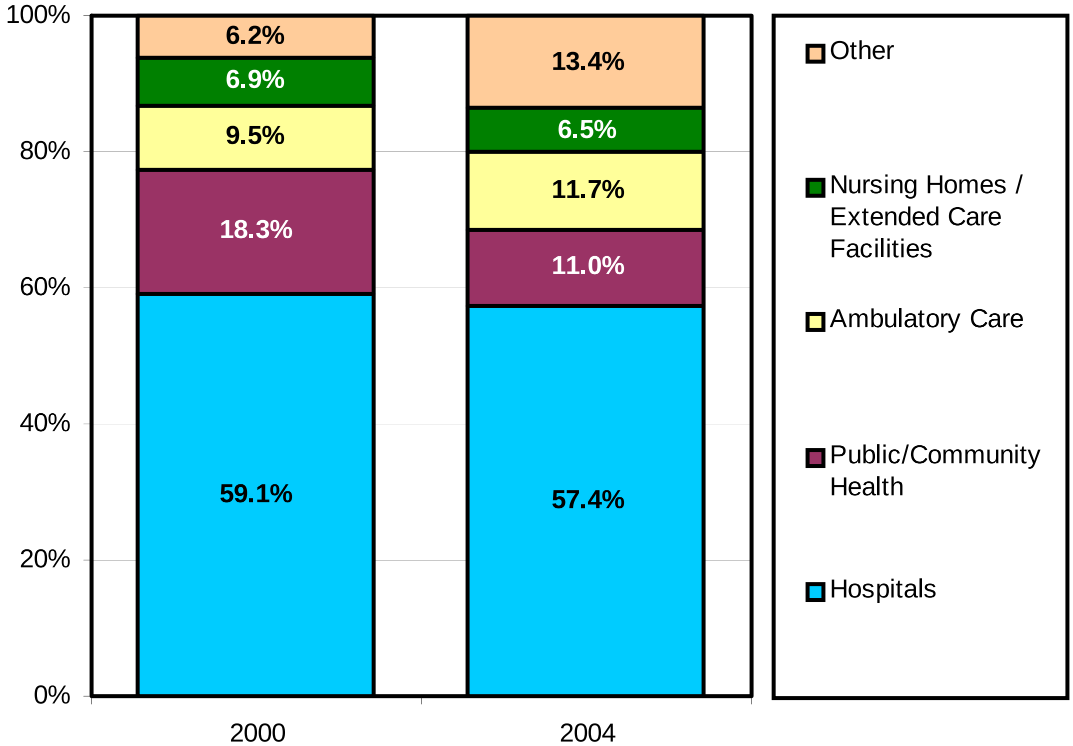
Source: 2004 NSSRN

2. Employment Settings

Figure 1 shows that hospitals continued to be the major employer of RNs in 2004, although the percentage of RNs working in hospitals declined from 59.1% in 2000 to 57.4% in 2004. The percentage working in public or community health organizations declined from 18.3% in 2000 to 11.0% in 2004.

A fact hidden in these simple employment statistics was that the day-to-day demands on many of these RNs, especially those employed in hospitals, increased dramatically over the past two decades. In fact, increases in patient acuity in hospitals and nursing homes resulted in a corresponding increase in the stress of nursing practice that caused a growing number of RNs to leave active patient care.

Figure 1. RN Employment by Setting, 2000 and 2004



Source: The Registered Nurse Population, March 2000. USDHHS, 2001.
2004 NSSRN, USDHHS, 2006

3. Trends in Supply

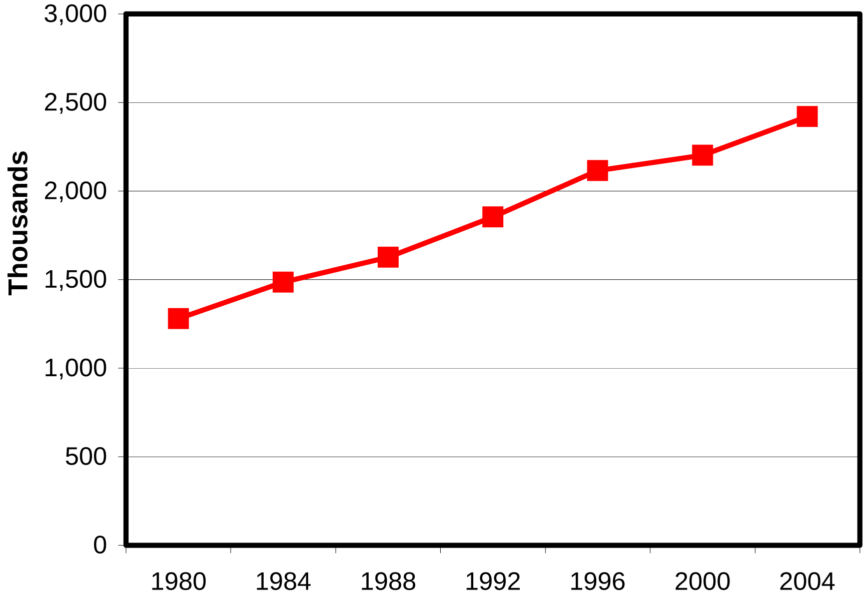
Between 1980 and 2004, the number of active RNs in the U.S. grew by nearly 90%. In 2000, there were more than 2.4 million active RNs, an increase of more than 1.1 million since 1980.

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 and later years (Figure 2). This slowdown in growth between 1996 and 2000 was 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 resumed between 2000 and 2004, more than keeping up with the growth in the population over the same period. The number of active RNs per 100,000 population nationally decreased from 798 in 1996 to 782 in 2000 (Figure 3). There was also wide variation in RNs per 100,000 population across the country. Massachusetts and South Dakota had the highest number of employed RNs per capita in 2000, 1,194 and 1,128 per 100,000 population, respectively. California and Nevada had the smallest number of employed RNs per capita, 544 and 520, respectively [1].

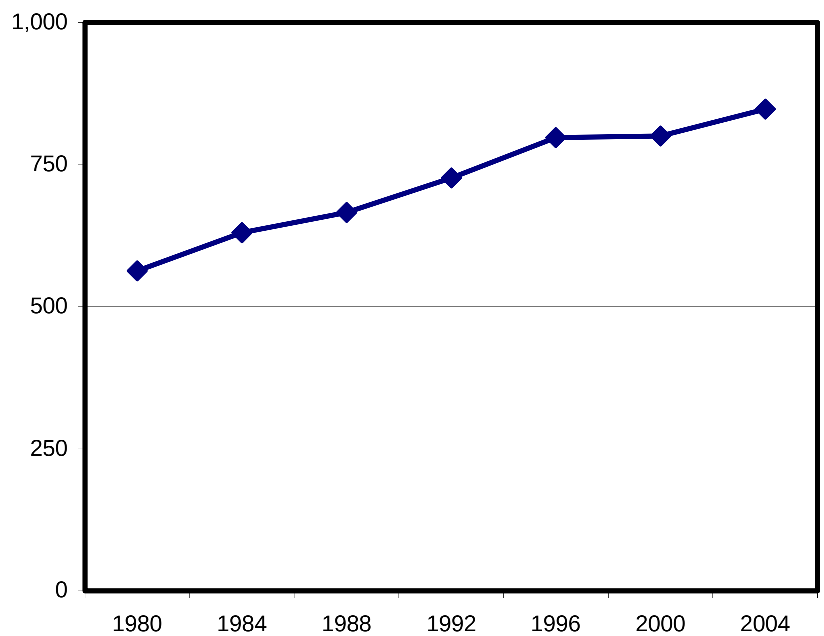
The number of candidates passing the RN licensure examination decreased steadily since 1995. Between 1995 and 2001, the number of RNs passing the licensing exam declined by nearly 28% [2].

Figure 2. Number of Active U.S. RNs, 1980 - 2004



Source: USDHHS, Findings from the National Sample Survey of RNs 2000, 2004

Figure 3. Active RNs per 100,000 Population, U.S., 1980 to 2000



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

The number of graduates of RN education programs in the country also declined between 1995 and 2001. While RN production grew steadily in the early 1990s, the total number of U.S.-

educated candidates taking the RN licensing examination dropped between 1995 and 2001, with nearly 29% fewer RNs graduating in 2001 than in 1995 [2]. Bachelor degree RN graduates (BRN) dropped by 20% while associate degree RN graduates (AND) declined by 28%. Although RN enrollments are increasing and the numbers of RN graduates in 2002 and 2003 were higher than the number of RN graduates in 2001 [3, 4], these figures are not yet back to 1995 levels.

4. Geographic Distribution

These national estimates and projections tell only part of the story. The two maps presented on the next page provide additional perspective on the supply of RNs in the U.S. in 2004. Figure 4 shows that the geographic dispersion of active RNs in 2004 was far from uniform across the country. In fact 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).

The range of ratios by county was even greater, which highlights one of the challenges for anyone interested in identifying counties or facilities with shortages of RNs. It is essential to have access to detailed data on RNs in counties in order to develop accurate estimates.

Figure 5 provides an additional perspective on this geographic variation, the change over time in the RN per capita ratios. This map shows that seven states (Connecticut, Florida, Idaho, Louisiana, Massachusetts, Maryland, and Rhode Island) experienced a decline in the number of active RNs per capita between 2000 and 2004. On the other end of the supply change spectrum were Alaska, District of Columbia, and New Hampshire, all with increases in active RNs per capita of over 25%. After discarding these three outliers, the Pearson correlation coefficient between the 2004 supply of RNs and the change in supply between 2000 and 2004 was only -0.039 (NS).

Figure 4. RNs per 100,000 Population in the U.S., 2004

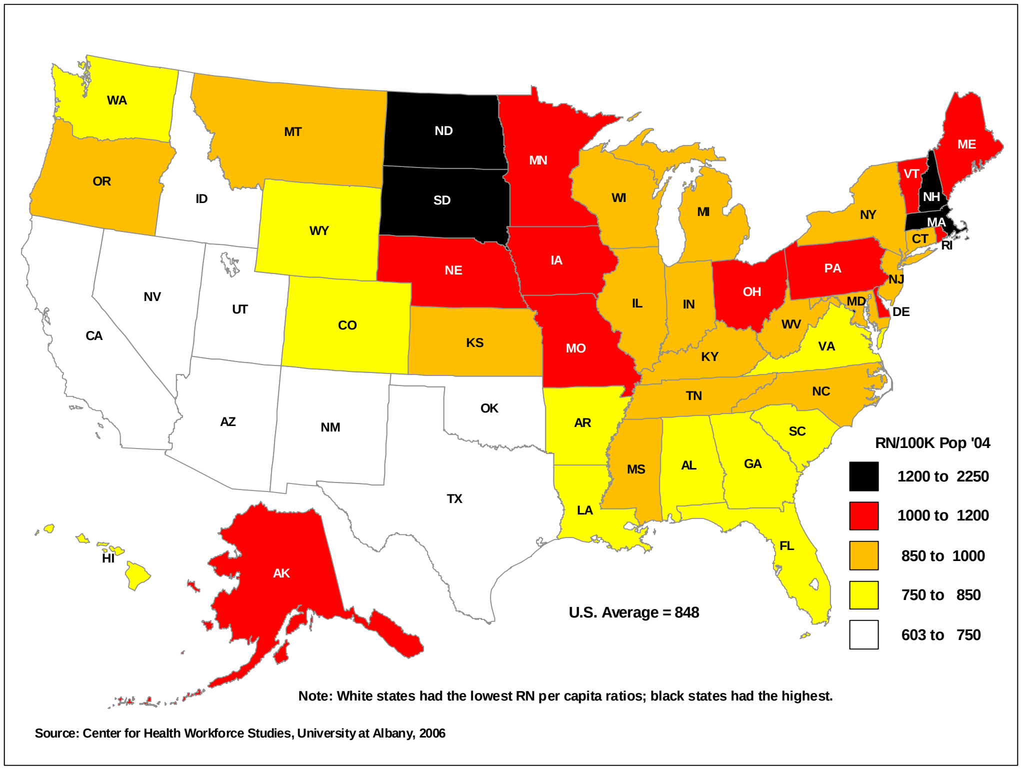
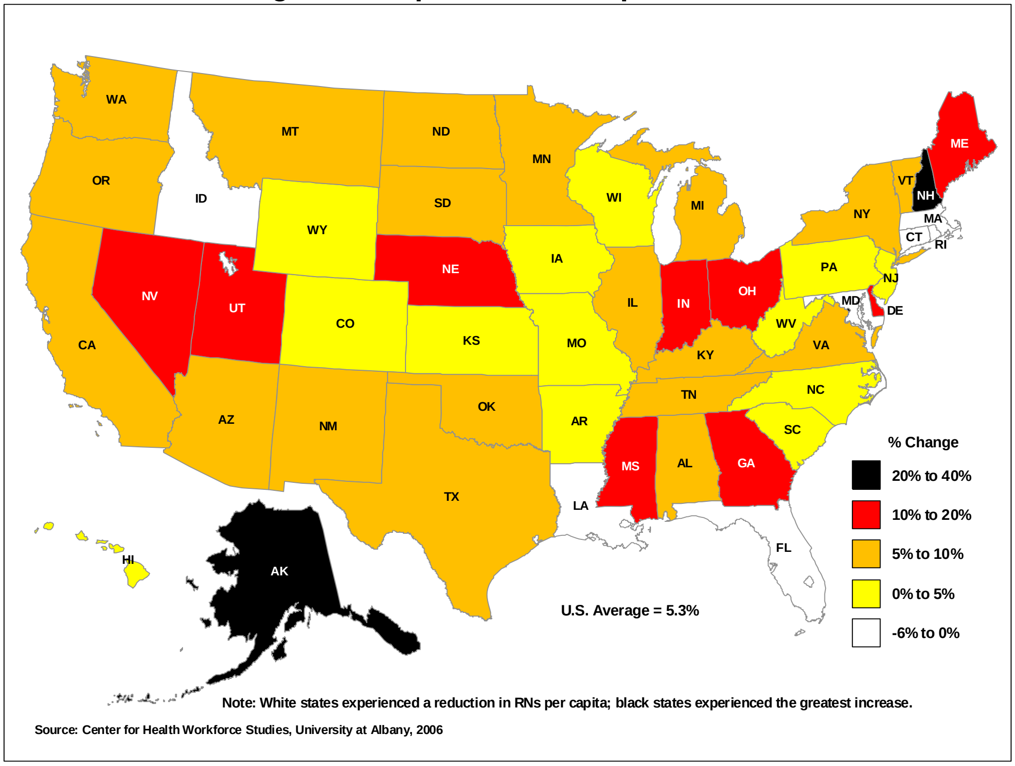


Figure 5. Percent Change in RNs per 100,000 Population in the U.S., 2000 to 2004



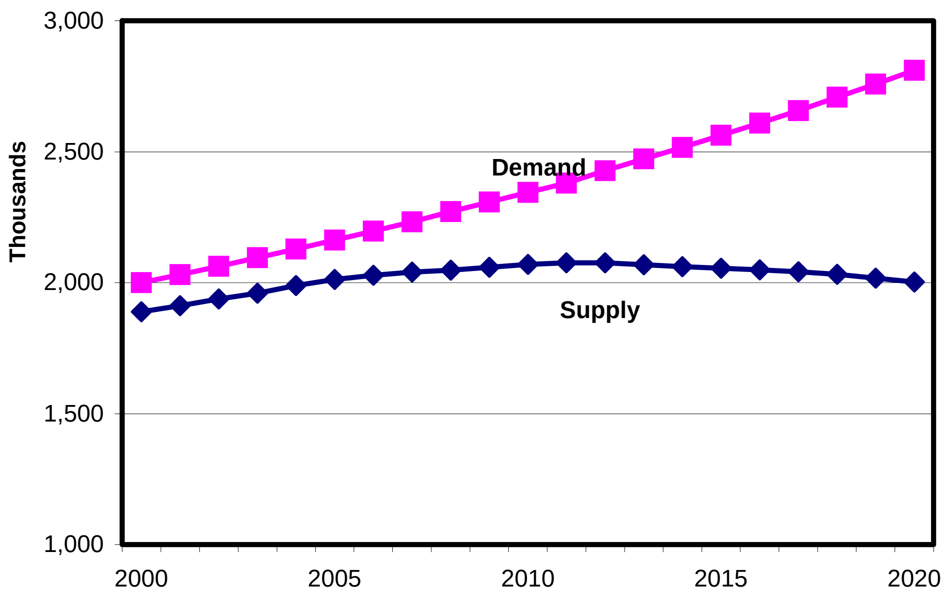
5. Projections of Future Supply

The National Center for Health Workforce Analysis at HRSA has projected a growing shortage of RNs over the next 15 years, with a 12% shortage by 2010 and a 20% shortage by 2015 (Figure 6). The projected shortage is the result of the expected increase in demand, coupled with a relatively stable supply of RNs [6].

Figure 7 updates these projections based in part on the 2004 NSSRN. Total numbers of RNs may rise until 2016 if age-specific cohorts follow patterns observed in the RNSS between 2000 and 2004. This is in large part because the sizes of birth cohorts in nursing tend to increase well into ages 50 to 55, and so a number of baby boomers (those born between 1947 and 1964) may still enter nursing as a second career over the next 10 years.

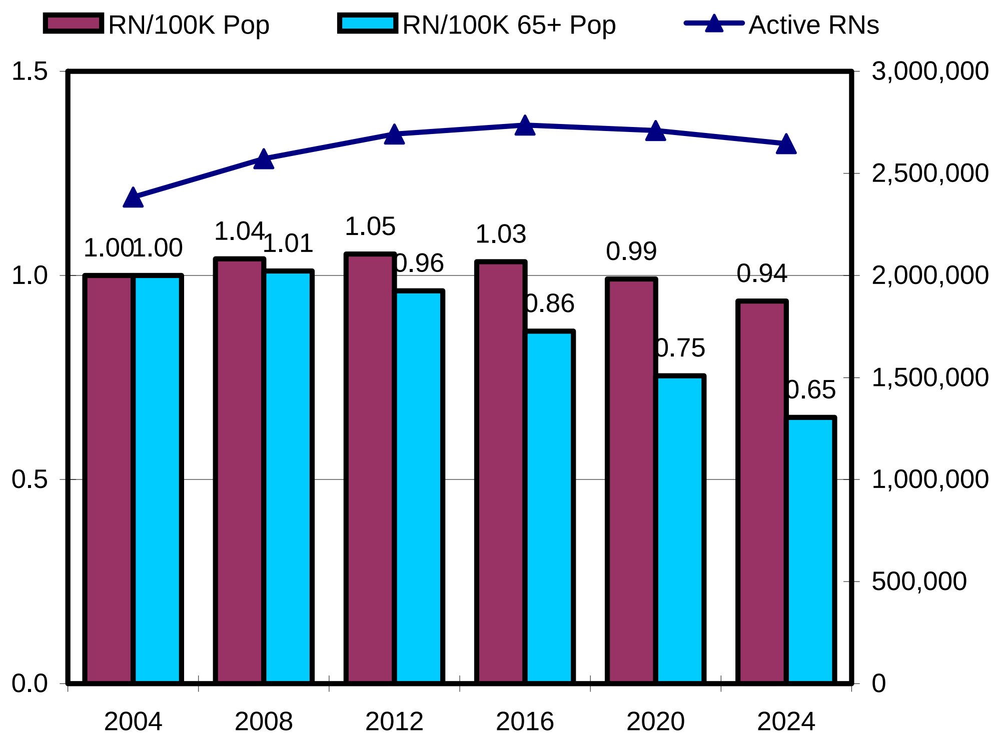
This does not mean that problems will not be felt until after 2016, however. Using these projections of numbers of RNs, total population, and the population age 65 and older from the U.S. Census Bureau, Figure 7 shows that the number of RNs per 100,000 population will peak in 2012, while the number of RNs per 100,000 population age 65 and older will peak in 2008 and decline by 5% (to below current rates) by 2012.

Figure 6. National Supply and Demand Projections for FTE RNs, 2000 to 2015



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

Figure 7. Indexed Projections of RNs per 100K Pop, RNs per 100K 65+ Pop, 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 [7, 8].
- 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. Staff RNs across all settings were less likely to be satisfied with their jobs, as were older RNs, with the exception of those employed in ambulatory care [1].
- Experienced RNs who left clinical settings identified a variety of reasons for their decision to leave, including lack of autonomy, heavy workload, too much paperwork, lack of opportunity for professional growth, inadequate staffing, and concerns about the quality of care. In some instances, these RNs went on to become advanced practice nurses (APNs) and return to clinical settings with more skills, more autonomy, and higher wages.

- There is increasing 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 [9, 10, 11].
- Several states have passed legislation prohibiting or limiting mandatory overtime for RNs and one state passed legislation mandating minimum nurse staff ratios in hospitals and nursing homes [12].

The current shortage of RNs and concerns about future shortages have led to new efforts—including this study—to address the problem of identifying facilities and communities with shortages of RNs.

D. Data Sets

Based on suggestions from the study advisory panels, four steps were implemented to develop criteria and methods to use for identifying facilities and communities with shortages of RNs. The four steps were:

- Designate data requirements, data elements, and data sets;
- Acquire data sets to use in pilot analyses;
- Perform pilot analyses for assessing different methods;
- Document the analyses for interested stakeholders.

1. Indicators and Corresponding Data Elements

These indicators were selected for inclusion based on the extent to which they were associated with facilities and agencies that have a shortage of RNs due to factors beyond their control (e.g., being located in a geographic area with few RNs). The advisory panels identified potential indicators at both the community and facility levels.

Community Indicators provide a critical context for any nursing shortage designation process. A number of community indicators identified by the expert panels seemed particularly relevant:

- Demographic Context
  - Rural or urban;
  - Age distribution of population;
  - Percent of population using Medicare or Medicaid;
  - Median population income; and
  - Percent of population in poverty.
- Nursing Context
  - RNs per 100 hospital beds;
  - Local nursing wages;
  - Numbers of nursing schools and graduates; and
  - Numbers of new RNs passing the National Council Licensure Examination for Registered Nurses (NCLEX).

Facility Indicators further refine and inform the shortage designation process. Facility indicators suggested by the panels included:

- Facility Indicators
    - Type of facility; and
    - Facility size;
- Workforce Statistics
    - Turnover rates;
    - Vacancy rates;
    - Hard-to-fill positions;
    - Staffing ratios (e.g., RNs per 100 beds, support staff per RN);
    - Poor facility outcomes (e.g., bad outcomes per 1,000 admissions);
    - Case mix and acuity;
    - Worker satisfaction; and
    - Turnover of leadership.

2. Identification and Compilation of Data

Data were compiled for two different tracks for this study: an "ideal" shortage designation methodology that incorporates all essential indicators required to identify shortages of RNs in either facilities or communities; and a "fall back" methodology that represents the best possible solution based on currently available data.

- **Indicators for an "ideal" methodology.** This step required the identification of facilities and communities that had data for all of the kinds of indicators listed above. Potential pilot sites considered by study staff were the Veterans Administration, Hospital Corporation of America, Health and Hospitals Corporation of New York City, and states such as North Dakota, North Carolina, Iowa, Pennsylvania, California, and Delaware.
- **Indicators for a "fall-back" methodology.** This step involved identifying data elements from the lists above that were available for facilities and communities of all different types across the U.S.

Two important data sources were used in this project: the Survey of Nurse Employers in North Carolina conducted by the North Carolina Center for Nursing; and the Area Resource File [ARF, 2004 release]. The facility variables were obtained from the Survey of Nurse Employers in North Carolina, and the community variables were obtained from the ARF database. The number of observations used to estimate the models was 325. There were four types of facilities estimated in this study: hospitals (65), home health facilities (79), long-term care facilities (128), and public health facilities (53).

Data were also obtained for 141 facilities in North Dakota (35 hospitals, 28 home health agencies, 45 long-term care facilities, and 33 public health agencies). These data were collected by the Center for Rural Health at the University of North Dakota, using questionnaires and definitions patterned after those used in North Carolina.

III. Methods and Models Using Facility Data

The third step in the process involved analyses of the data compiled previously to test different methods for which pertinent data currently exist. Part of this process involved experimentation with different equations and computational methods to determine which specific formulas are most appropriate for each of the four types of facilities. These activities revealed a number of interesting and important insights about nursing shortages, which are summarized below.

A. Preliminary Analyses

Figure 8 presents the distribution of the indicator of difficulty recruiting RNs based on all facilities in North Carolina. The figure shows the number of facilities that experienced difficulty recruiting RNs (indicator >3) was more than double the number of the facilities with no difficulty recruiting RNs (indicator <3). In this case, 68 facilities (20.9%) reported not having difficulty recruiting RNs compared to 155 facilities (47.7%) that reported having difficulty recruiting RNs. The figure also shows that only 17 facilities (5.2%) reported that it was very easy to recruit RNs, in contrast to 56 facilities (17.2%) that reported it was very difficult to recruit RNs.

Figure 8. Distribution of RN Recruitment Difficulty Indicator, Based on Four Types of Health Facilities in North Carolina in 2004

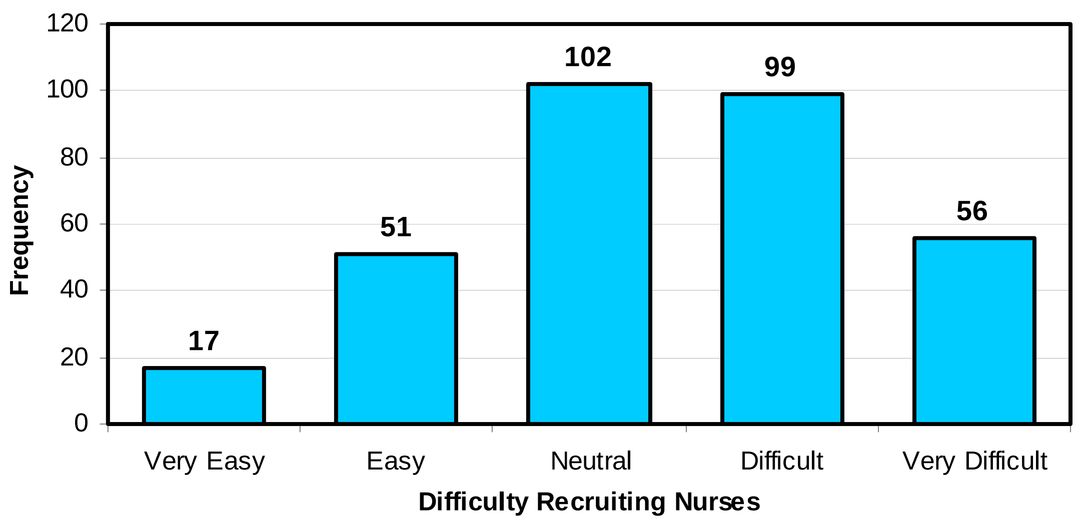


Figure 9 presents the distribution of difficulty indicator by facility type. From this figure we can see that the distributions of difficulty to recruit RNs were different among all four types of facilities. For example, 4.6% of hospitals reported it was very difficult to recruit RNs, in contrast to 26.4% of public health facilities reported very difficult to recruit RNs.

Figure 9. Nursing Recruitment Difficulty Indicators in North Carolina, by Facility Type, 2004

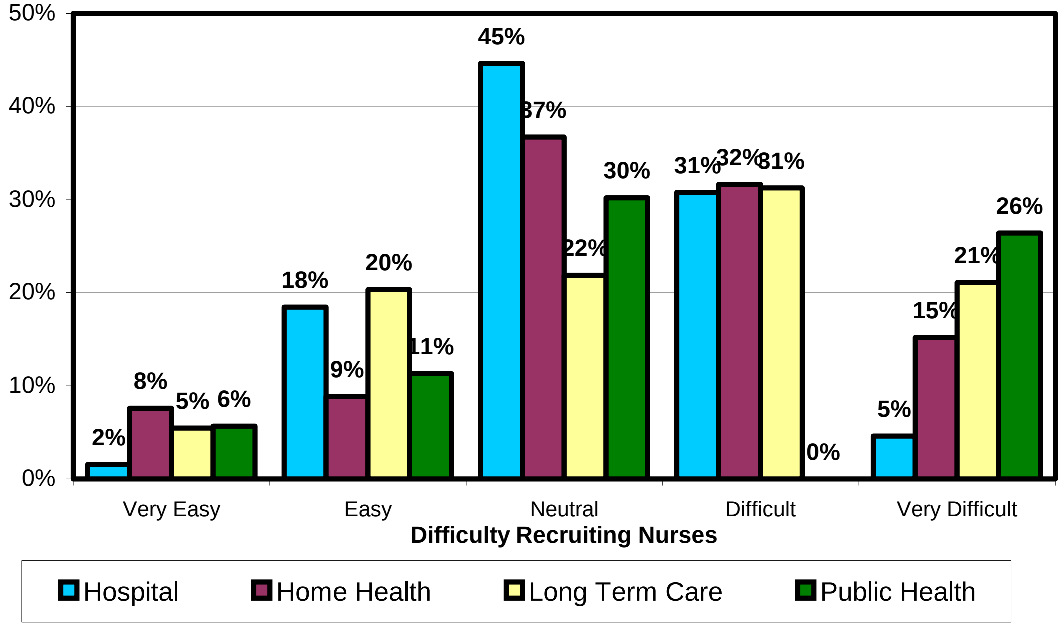


Figure 10 compares the distributions of the predicted recruiting difficulty scores for the four types of facilities in North Carolina, based on the Ordered Probit model estimated using data for 2004. The figure shows clearly that the variation in recruiting difficulty is greatest for public health agencies and least for hospitals. It also shows that on average both public health agencies and long-term care facilities have statistically significantly greater difficulty recruiting RNs than hospitals ($p \leq 0.05$, since the 95% confidence intervals do not overlap).

Figure 10. Distribution of Predicted Difficulty Recruiting RNs in North Carolina by Type of Facility, 2004

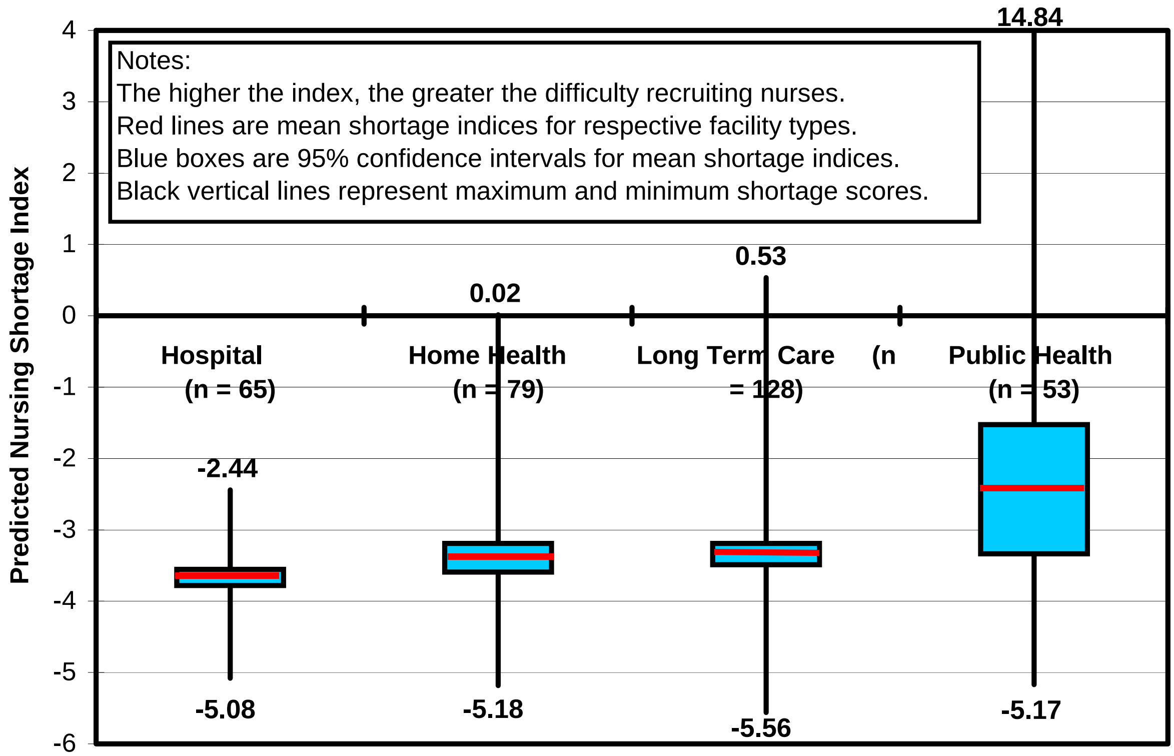


Table 2 presents the distribution of facility type by difficulty indicator and Chi-Square statistic to test the null hypothesis that there is no association between type of facility and the difficulty recruiting RNs. Based on the Chi-square statistic, the null hypothesis was rejected (p = 0.011) because different types of facilities had different levels of difficulty recruiting RNs. The implication was that different types of facilities have different behaviors in term of modeling nursing shortages.

Table 2. Distribution of Type of Facility by Nursing Recruitment Difficulty Indicator

Facility Type	Difficulty Indicator					Total
	Very Easy	Easy	Neutral	Difficult	Very Difficult	
Hospital	1.5%	18.5%	44.6%	30.8%	4.6%	65
Home Health	7.6%	8.9%	36.7%1	31.6%	15.2%	79
Long-Term Care	5.5%	20.3%	21.9%	31.2%	21.1%	128
Public Health	5.7%	11.3%	30.2%	26.4%	26.4%	53
Total	17	51	102	99	56	325

Chi-Square = 25.9 (df = 12)
Test of H_0: No association between type of facility and difficulty to recruit
H_0 is rejected with p-value = 0.011

Table 3 presents the distribution of difficulty indicator by number of adverse consequences of shortages and the Spearman correlation coefficient to test the null hypothesis that there is no relationship between difficulty indicator and number of consequences. From the Spearman correlation statistic, the null hypothesis was rejected ($p<0.0005$), meaning that on average facilities that experienced greater difficulty recruiting RNs had more bad consequences.

Table 3. Distribution of Nursing Recruitment Difficulty Indicator by Number of Bad Consequences

Difficulty Indicator	Number of Consequences				
	0	1	2	≥ 3	Total
1	86.7%	13.3%	0.0%	0.0%	15
2	61.4%	22.7%	9.1%	6.8%	44
3	56.5%	26.1%	14.1%	3.3%	92
4	32.2%	32.2%	23.0%	12.6%	87
5	26.0%	32.0%	24.0%	18.0%	50
Total	**133**	**80**	**49**	**26**	**288**

Spearman correlation coefficient = 0.343
Test of H_0: Correlation = 0
H_0 is rejected with p-value < 0.0005

B. Empirical Models for North Carolina Hospitals

A number of models were estimated for hospitals in North Carolina. The steps followed are summarized below.

1. Select Shortage Indicator (Dependent) Variable

The indicator of nursing shortage used as a dependent variable was the number of reported negative effects on operations revealed by a facility. Most facilities indicated no effects or only one effect. The mean value for all facilities was 0.89, with a standard deviation of 1.07. Based on this, we defined facilities as being needy (for test purposes only), if they presented two or more effects on operations. Under this definition, 15.5% of hospitals were needy.

2. Estimated Medical Need Based on Population Characteristics

The population was adjusted by gender and age based on average use of primary care. This weighted older adults and infants more heavily than younger people and weighted women more heavily than men. The resulting variable was an estimate of how many primary care visits the population would require in a year's time. Although the relationship between use of primary care and need for services, such as home health or long-term care, is open to debate, this variable was simply a way of standardizing the population based on characteristics known to affect medical need.

3. Select/Construct Explanatory (Independent) Variables

The following variables were selected for use in the North Carolina analyses:

1) Active RNs Employed in the County per 100,000 Adjusted Population
2) Students Enrolled in RN Programs in the County per 100,000 Adjusted Population
3) Number of Short-Term General Hospitals
4) Number of Short-Term General Hospital Beds
5) Ratio of Average RN Salary to Median Income
6) Number of Nursing and Personal Care Facilities
7) Percent of Population with Income Below Poverty Level
8) Population per Square Mile
9) Ratio of RNs to Hospital Beds
10) Number of Hours per Week Paid for Agency RNs
11) Number of Overtime RN Hours per Week
12) RN Vacancy Rate
13) RN Turnover Rate
14) Ratio of LPNs to RNs
15) Total Number of Budgeted RN Positions
16) Percent Non-Hispanic White

Average values for these variables are shown in Table 4 for three groups of hospitals in North Carolina.

Table 4. Average Values of Selected Indicators for Three Groups of Hospitals in NC

Indicator	All Hospitals		Hospitals Reporting No Nurse Staffing Problems		Hospitals Reporting Two or More Nurse Staffing Problems	
	Mean	S.D.	Mean	S.D.	Mean	S.D.
Active RNs Employed per 100K Medical Need	204.2	104.7	226.0	116.1	182.7	66.5
Students in RN Programs per 100K Medical Need	19.7	2.6	23.8	44.3	2.6	4.7
Number of Short-term Community Hospitals	2.0	1.8	2.3	2.0	1.4	0.6
Number of Short-term Community Hospital Beds	679.5	75	807.2	804.4	474.4	536.2
Ratio of Average RN Salary to Median Income	1.5	0.3	1.4	0.29	1.6	0.3
Number of Nursing and Personal Care Facilities	20.1	21.7	22.7	22.98	16.1	17.9
Percent of Population Below Poverty Income	13.0	4.2	12.5	3.88	15.25	5.1
Population per Square Mile	334.4	358.8	398.9	388.1	193.9	183.5
Ratio of RNs to Hospital Beds	0.5	0.3	0.5	0.22	0.6	0.4
Number of Hours per Week Paid for Agency RNs	2.6	3.4	2.1	3.14	2.1	2.6
Number of Overtime RN Hours per Week	4.94	8.2	3.8	2.66	4.5	3.4
RN Vacancy Rate	6.9	4.9	6.2	4.45	9.6	5.9
RN Turnover Rate	15.5	7.8	13.5	5.73	18.9	11.8
Ratio of LPNs to RNs	0.1	0.1	0.1	0.14	0.2	0.1
Total Number of Budgeted RN Positions	358.6	455.3	429.7	498.7	319.1	478.1
Percent Non-Hispanic White	70.0	16.2	70.0	15.2	65.4	18.9

Population per square mile was very highly correlated with several other variables, so a natural log transformation was applied to reduce problems of multicollinearity. There was also potential multicollinearity between the number of RNs per 100,000 adjusted population and number of general hospital beds per 100,000 adjusted population. Number of hospital beds was dropped in favor of number of hospitals.

4. Run OLS Regression Model, Full and Abbreviated

Two different OLS models were estimated to predict the number of adverse effects in hospitals in North Carolina, one for the full model that included both community and facility data and one that included only community data. These models are summarized below.

Full model

Table 5. Coefficients for Full OLS Regression Model to Predict Number of Adverse Effects of Nursing Shortages in Hospitals in North Carolina

Explanatory (Independent) Variable	Unstandardized Coefficients		Standardized Coefficients	t	p Value
	B	Std Err			
Constant	-0.683	3.020	-	-0.226	0.822
RNs per 100,000 Adjusted Need	-0.0035	0.002	-0.353	-2.002	0.052
RN Salary to Average Salary	0.518	0.707	0.132	0.732	0.468
# Nursing and Personal Care Facilities	**0.032**	**0.015**	**0.663**	**2.176**	**0.035**
% Population Below Poverty, 2000	0.078	0.065	0.308	1.202	0.236
RNs per Hospital Bed	0.265	0.445	0.082	0.596	0.555
Hours of Agency RNs	0.0025	0.043	0.008	0.058	0.954
Hours of RN Overtime	-0.0008	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 ln)	0.156	0.358	0.146	0.436	0.665
# Short-term Community Hospitals, '01	**-0.359**	**0.134**	**-0.610**	**-2.690**	**0.010**
RN Students per 100K Adjusted Need	**-0.010**	**0.004**	**-0.392**	**-2.828**	**0.007**
% Population Non-Hispanic White, 2004	-0.011	0.012	-0.167	-0.902	0.372

Dependent Variable: NUM_CONS
Selecting only cases for which FAC_TYPE = hospital
R^2 = 0.429

Abbreviated model

Because most of the variables that appeared most critical were community variables rather than facility variables, an abbreviated model was also run using only community information. Due to the constraints of data availability, the abbreviated model is one that can be used more easily in practice. The R^2, however, dropped substantially, from 0.429 in the full model to only 0.177 in the abbreviated model.

Table 6. Coefficients for Abbreviated OLS Regression Model to Predict Number of Adverse Effects of Nursing Shortages in Hospitals in North Carolina

Independent Variables	Unstandardized Coefficients		Standardized Coefficients	t	p Value
	Coefficient	Std Err			
Constant	1.295	2.374	-	0.546	0.587
RNs per 100,000 Adjusted Need	-0.0012	0.001	-0.132	-0.880	0.382
RN Salary to Average Salary	0.281	0.582	0.081	0.482	0.631
# Nursing/Personal Care Facilities	0.023	0.012	0.494	1.905	0.061
% Population Below Poverty, 2000	0.033	0.053	0.136	0.622	0.536
RNs per Hospital Bed	0.044	0.402	0.013	0.108	0.914
Persons per Square Mile (natural ln)	-0.158	0.271	-0.159	-0.582	0.563
# Short-term Community Hospitals, '01	**-0.227**	**0.109**	**-0.414**	**-2.076**	**0.041**
RN Students per 100K Adjusted Need	-0.0054	0.003	-0.226	-1.967	0.053
% Population White Non-Hispanic, '04	-0.0053	0.010	-0.086	-0.511	0.611

Dependent Variable: NUM_CONS
Selecting only cases for which FAC_TYPE = hospital
$R^2 = 0.177$

5. Compare Predicted and Actual Scores for Full and Abbreviated Models

Coefficients from the full and abbreviated regression models were used to estimate predicted number of problems in each facility. The top 16% of facilities in regard to predicted number of problems were considered to have made the test "cut" of 15.5% chosen arbitrarily based on earlier analysis (see Step 1). The facilities selected by the full model and the abbreviated model were compared to the facilities whose actual problem scores were in the top 15.5%.

Using the abbreviated model, 84% of hospitals were classified correctly based on the arbitrary value chosen earlier. Eight percent of facilities were misclassified as not needy by the abbreviated model when their actual scores qualified them as needy, while 7% were misclassified as being needy when their actual scores did not qualify them as such.

Using all the information in the full model would have increased the accuracy of prediction to 89%, with 5% of facilities erroneously classified as needy and 5% erroneously classified as not needy.

6. Conclusion

Using the information from the testing in Step 5, we conclude that using an abbreviated model with widely available community level data to assign facilities need scores would result in approximately 84% of facilities being correctly classified. Supplementing this with an appeals process requiring the additional information needed for the full model would correctly classify an additional 5% of facilities.

C. Empirical Models for North Dakota Hospitals

The coefficients estimated for North Carolina hospitals were applied to hospitals in North Dakota. The results are summarized below.

1. Assign North Carolina Predicted Need Scores to North Dakota Hospitals

When the coefficients for the abbreviated model obtained from the empirical models developed for North Carolina were applied to hospitals in North Dakota, not surprisingly the classifications were less accurate. Seventy-nine percent of North Dakota hospitals were correctly classified by this application of North Carolina data, while 10% were erroneously classified as needy and 10% were erroneously classified as not needy.

2. Conclusion

This analysis suggests that using coefficients based on models estimated in one state achieves lower accuracy when applied to facilities in another state. Additional research would be required to determine whether the decline in accuracy might be related to the extent to which general characteristics of the states are similar or different.

D. Empirical Models for North Carolina Nursing Homes

The empirical models for nursing homes in North Carolina are summarized below.

1. Select Indicator (Independent) Variable

The indicator of nursing shortage used as a dependent variable was the number of reported effects on operations reported by a facility. Most facilities reported no effects or only one effect. The mean value for all facilities was 1.0, with a standard deviation of 1.1. Based on this, we defined facilities as being needy (for test purposes only) if they reported two or more effects on operations. Under this definition, 31.3% of nursing homes were needy.

2. Estimate Medical Need Based on Population Characteristics

The population was adjusted by gender and age based on average use of primary care. This weighted older adults and infants more heavily than younger people and women more heavily than men. The resulting variable was an estimate of how many primary care visits the population would require in a year’s time. Although the relationship between use of primary care and need for services such as home health or long-term care is open to debate, this variable was simply a way of standardizing the population based on characteristics known to affect medical need.

3. Select/Construct Independent Variables

1) Active RNs employed in the county per 100,000 adjusted population
2) Students enrolled in RN programs in the county per 100,000 adjusted population
3) Number of short-term general hospitals
4) Number of short-term general hospital beds
5) Ratio of average RN salary to median income
6) Number of nursing and personal care facilities
7) Percent of the population with income below poverty level
8) Population per square mile
9) Ratio of RNs to hospital beds

10) Number of hours per week paid for agency RNs
11) Number of overtime RN hours per week
12) RN vacancy rate
13) RN turnover rate
14) Ratio of LPNs to RNs
15) Total number of budgeted RN positions
16) Percent non-Hispanic white

Table 7. Means and Standard Deviations of Selected Independent Variables Related to Nursing Shortages in North Carolina Nursing Homes

Independent Variables	All Nursing Homes		Nursing Homes Reporting No Nurse Staffing Problems		Nursing Homes Reporting Two or More Nurse Staffing Problems	
	Mean	S.D.	Mean	S.D.	Mean	S.D.
Active RNs Employed in County per 100K Medical Need	189.6	101.1	204.9	111.2	207.3	87.5
Students in RN Programs per 100K Medical Need	36.0	139.9	29.6	65.8	18.1	30.8
Number of Short-Term Community Hospitals	1.7	1.6	2.1	2.0	1.5	1.2
Number of Short-Term Community Hospital Beds	597.9	689.3	740.5	780.3	704.5	654.9
Ratio of average RN salary to median income	1.5	0.3	1.5	0.3	1.4	0.3
Number of Nursing and Personal Care Facilities	18.4	19.6	22.9	23.0	22.3	18.8
Percent of Population w/ Income Below Poverty Level	13.0	4.1	12.9	4.1	12.5	3.9
Population per Square Mile	300.0	315.0	357.4	373.2	351.6	278.9
Ratio of RNs to Hospital Beds	0.5	0.3	0.4	0.2	0.5	0.2
Hours per Week Paid for Agency RNs	2.1	5.9	2.5	7.9	3.8	9.4
Number of Overtime RN Hours per Week	6.5	9.7	12.4	11.1	14.1	13.3
RN Vacancy Rate	9.5	13.6	8.5	12.2	9.7	13.1
RN Turnover Rate	29.6	43.6	40.7	69.3	38.8	32.8
Ratio of LPNs to RNs	1.3	2.2	2.2	1.5	3.1	4.3
Total Number of Budgeted RN Positions	79.1	240.8	7.2	4.6	7.0	4.9
Percent Non-Hispanic White	70.9	16.4	70.4	17.5	70.2	15.1

Population per square mile was very highly correlated with several other variables, and so a log transformation was applied to avoid problems with multicollinearity. There was also potential multicollinearity between the number of RNs per 100,000 adjusted population, and number of general hospital beds per 100,000 adjusted population. Number of hospital beds was dropped in favor of number of hospitals.

4. Run OLS Regression Model

The following regression was run for nursing homes in North Carolina:

Table 8. Coefficients for OLS Regression Model to Predict Number of Adverse Effects of Nursing Shortages in Nursing Homes in North Carolina

Independent Variable	Unstandardized Coefficients		Standardized Coefficients	t	p Value
	B	Std Err	Beta		
(Constant)	-2.395	2.872	-	-0.834	0.407
RNs per 100,000 Adjusted Need	-0.0007	0.002	-0.063	-0.379	0.706
RN Salary to Average Salary	-1.307	0.779	-0.338	-1.677	0.098
# Nursing/Personal Care Facilities	-0.00338	0.012	-0.060	-0.271	0.787
% Population Below Poverty, 2000	0.114	0.067	0.393	1.690	0.095
RNs per Hospital Bed	0.585	0.637	0.110	0.919	0.361
Hours of Agency RNs	0.0051	0.016	0.040	0.314	0.754
Hours of RN Overtime	0.0073	0.012	0.070	0.599	0.551
RN Vacancy Rate	-0.0014	0.011	-0.014	-0.131	0.896
RN Turnover Rate	0.0002	0.002	0.012	0.095	0.925
Persons per Square Mile (Natural ln)	0.632	0.330	0.494	1.914	0.059
# Short-Term Commun Hospitals, '01	**-0.344**	**0.119**	**-0.485**	**-2.881**	**0.005**
RN Students per 100,000 Adjusted Need	0.0010	0.003	0.047	0.385	0.701
% Population White Non-Hispanic, 2004	0.014	0.012	0.199	1.160	0.250

Dependent Variable: NUM_CONS
Selecting only cases for which FAC_TYPE = long-term care
$R^2 = 0.20$

This model had little predictive value, perhaps because the chosen dependent measure of nursing shortage was inappropriate for nursing homes, which rely heavily on LPNs. The question about the effects of a nursing shortage on facility operations did not specify RN shortages, and so it seemed plausible that significant relationships were not emerging based on RN variables because respondents answered this question primarily thinking of LPNs.

Therefore, in estimating this model, the decision was made to revert to RN vacancy rates, acknowledging that the facilities reporting the highest vacancy rates are not necessarily the facilities suffering the most from the RN shortage. Several variables relating to the LPN job market were also included in this second version of the model. The mean RN vacancy rate for nursing homes was 10.6, with a standard deviation of 15.8. On this basis, we classified any facility with a RN vacancy rate of more than 26.4 as "needy" as a test value (11.9% of facilities).

5. Run Alternate Model

An alternate OLS regression model was estimated for RN Vacancy Rates in nursing homes in North Carolina (Table 9). It focused more on LPNs and less on RNs, which better reflects the actual staffing patterns at nursing homes.

Table 9. Coefficients for Alternate OLS Regression Model to Predict RN Vacancy Rates in Nursing Homes in North Carolina

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

Dependent Variable: RNVacRate
Selecting only cases for which FAC_TYPE = long-term care
$R^2 = 0.35$

6. Compare Predicted and Actual Scores for NC Nursing Home Models

Coefficients from the regression model were used to estimate predicted number of problems in each facility. The top 31.5% of facilities in regard to predicted number of problems were considered to have made the test "cut" of 31.3% chosen arbitrarily based on earlier analysis (see Step 1). The facilities selected by the full model were compared to the facilities whose actual problem scores were in the top 31.3%.

Using the full model, only 73% of nursing homes were classified correctly based on the arbitrary value chosen earlier. Fourteen percent of facilities were misclassified as not needy by the model when their actual scores qualified them as needy, while 12% were misclassified as being needy when their actual scores did not qualify them as such.

The alternate model, however, proved very effective in identifying facilities with the highest RN vacancy rates. Eighty-eight percent of facilities were correctly classified as "needy" based on the arbitrary value chosen earlier. Seven percent were misclassified as not needy by the model when their actual scores qualified them as needy, while 6% were misclassified as being needy when their actual scores did not qualify them as such.

7. Conclusion

Although there are several reliable indicators of high RN vacancy rates in nursing homes, there is little that predicts need in terms of the problems facilities report in their operations as a result of the nursing shortage. This is problematic because the facilities reporting the highest vacancy rates are not necessarily the facilities suffering the most from nursing shortages. Indeed, RN vacancy rates were unrelated to reports of shortage problems. The facilities the majority of facilities defined as needy on the basis of reported problems were not the same facilities defined as needy on the basis of RN vacancy rates. This may be due to the prominence of LPNs in long-term care, however, causing most people to answer the question about problems based on LPN shortages rather than RN shortages. Given this ambiguity, RN vacancy rates may be the better indicator of long-term care shortages.

Another shortcoming of the analyses is that population is standardized based on primary care utilization rates estimated by age and gender. This formula may be inappropriate for estimating long-term care need in the population, and perhaps a new formula for standardization based on long-term care utilization rates should be introduced. A standardization of the population that is tailored to long-term care might produce more useful models and more reliable estimates of community need. Number of long-term care beds and beds per older adults would also be useful information to include in future attempts to model.

E. Tailoring for Long-Term Care

As stated in the Conclusion section of Part I, the initial analyses were based on a general model tested for four types of facilities: hospitals, home health agencies, public health agencies, and long-term care facilities. For the former three types of facilities, indicators of general medical need and availability of general medical services may be relevant indicators in judging adequacy of the RN supply. Long-term care, however, is a more specific type of care provided to a narrower segment of the population.

1. Assign North Carolina Predicted Need Scores to North Dakota LTC Facilities

When the alternate model obtained in Part I of the pilot testing was applied to nursing homes in North Dakota, the classifications were considerably less accurate. This is the same result as observed for hospitals.

2. Conclusion

As was the case for hospitals, this analysis suggests that using coefficients based on long-term care models estimated in one state achieves lower accuracy when applied to facilities in another state. Additional research would be required to determine whether the decline in accuracy might be related to the extent to which general characteristics of the states are similar or different.

F. Empirical Models for North Carolina Home Health Agencies

The steps used to estimate the empirical models for home health agencies in North Carolina are summarized below.

1. Selection Indicator (Dependent) Variable

The indicator of nursing shortage used as a dependent variable was the number of reported effects on operations reported by an agency. Most agencies reported no effects or only one effect. The mean value for all agencies was 0.8, with a standard deviation of 1.0. Based on this, we defined agencies as being needy (for test purposes only) if they reported two or more effects on operations. Under this definition, 19.4% of home health agencies were needy.

2. Estimating Medical Need Based on Population Characteristics

The population was adjusted by gender and age based on average use of primary care. This weighted older adults and infants more heavily than younger people and weighted women more heavily than men. The resulting variable was an estimate of how many primary care visits the population would require in a year's time. Although the relationship between use of primary care and need for services such as home health or long-term care is open to debate, this variable was simply a way of standardizing the population based on characteristics known to affect medical need.

3. Select/Construct Independent Variables

1) Active RNs employed in the county per 100,000 adjusted population
2) Students enrolled in RN programs in the county per 100,000 adjusted population
3) Number of short-term general hospitals
4) Number of short-term general hospital beds
5) Ratio of average RN salary to median income
6) Number of nursing and personal care facilities
7) Percent of the population with income below poverty level
8) Population per square mile
9) Ratio of RNs to hospital beds
10) Number of hours per week paid for agency RNs
11) Number of overtime RN hours per week
12) RN vacancy rate
13) RN turnover rate
14) Ratio of LPNs to RNs
15) Total number of budgeted RN positions
16) Percent non-Hispanic white

Table 10. Means and Standard Deviations of Selected Independent Variables Related to Nursing Shortages in North Carolina Home Health Agencies

Independent Variable	All Home Health Agencies		Agencies Reporting No Nurse Staffing Problems		Agencies Reporting Two or More Nurse Staffing Problems	
	Mean	S.D.	Mean	S.D.	Mean	S.D.
Active RNs in County per 100K Medical Need	184.6	98.0	187.2	102.4	187.5	95.4
Students in RN Programs per 100K Medical Need	48.7	195.8	54.6	230.8	39.5	75.6
Number of Short-Term Community Hospitals	1.7	1.4	1.7	1.4	1.8	1.7
Number of Short-Term Community Hospital Beds	555.8	655.4	590.2	693.9	617.2	679.9
Ratio of Average RN Salary to Median Income	1.5	1.4	1.5	0.3	1.5	0.3
Number of Nursing and Personal Care Facilities	16.8	17.3	17.3	17.8	20.1	20.4
Percent of Population w/ Income Below Poverty Level	12.6	4.1	12.6	4.0	13.5	4.8
Population per Square Mile	287.6	283.5	292.3	287.8	312.6	331.8
Ratio of RNs to Hospital Beds	0.5	0.3	0.4	0.3	0.5	0.2
Number of Hours per Week Paid for Agency RNs	1.7	4.9	1.2	3.4	4.7	9.2
Number of Overtime RN Hours per Week	2.7	4.3	2.3	3.7	5.0	6.6
RN Vacancy Rate	10.1	15.9	7.6	14.1	21.5	21.2
RN Turnover Rate	28.3	37.4	19.1	24.9	60.4	52.9
Ratio of LPNs to RNs	0.3	0.3	0.3	0.3	0.3	0.2
Total Number of Budgeted RN Positions	12.0	10.5	12.5	11.5	10.7	7.9
Percent Non-Hispanic White	72.7	16.4	73.5	15.8	63.2	19.0

Population per square mile was very highly correlated with several other variables, and so a log transformation was applied to avoid problems with multicollinearity. There was also potential multicollinearity between the number of RNs per 100,000 adjusted population, and number of general hospital beds per 100,000 adjusted population. Number of hospital beds was dropped in favor of number of hospitals.

4. Run OLS Regression Models, Full and Abbreviated

The following regression was run for home health agencies in North Carolinas:

Table 11. Coefficients for 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

Dependent Variable: NUM_CONS
Selecting only cases for which FAC_TYPE = home health
R^2 = 0.44

An abbreviated model was also estimated. It appeared to have little value for home health agencies because most of the variables that appeared most critical were facility variables rather than community variables, and would have to be collected directly from facilities. Variables that were "optional," and were able to be dropped for an abbreviated model were the variables most widely available.

5. Compare Predicted and Actual Scores for Full Home Health Agency Model

Coefficients from the full regression model were used to estimate predicted number of problems in each agency. The top 19.2% of agencies in regard to predicted number of problems were considered to have made the test "cut" of 19.4% chosen arbitrarily based on earlier analysis (see Step 1). The agencies selected by the full model were compared to the agencies whose actual problem scores were in the top 19.4%.

Using the full model, 85% of home health agencies were classified correctly based on the arbitrary value chosen earlier. Seven percent of agencies were misclassified as not needy by the model when their actual scores qualified them as needy, while 8% were misclassified as being needy when their actual scores did not qualify them as such.

6. Conclusion

Using the information from the testing in Step 5, we can conclude that using the full model with both widely available community level data and data collected directly from agencies to assign need scores would result in approximately 85% of agencies being correctly classified. The importance of the facility-level variables in the model, however, means that any effective strategy for classifying home health agencies will require the collection of data on factors such as turnover and vacancy rates.

As with long-term care facilities, however, there was an issue in using a model designed to incorporate measures of general medical need. Home health is not primary care, and patients tend to be predominantly older while both the oldest and the youngest segments of the population disproportionately consume primary care. A standardization of the population that is tailored to long-term care utilization might produce more useful models and more reliable estimates of community need. While reliable community-level data on home health capacity will not be obtainable, number of long-term care beds and beds per older adult might also be useful information to include in future attempts to model, both because long-term care serves similar populations to home health, and because long-term care and home health may compete for the same pool of RNs. Incorporation of such variables may make community-level indicators more useful in evaluating home health shortages, possible enabling the construction of a reliable abbreviated model as was done for hospitals.

G. Empirical Models for North Dakota Home Health Agencies

The coefficients estimated for North Carolina home health agencies were applied to home health agencies in North Dakota. The results are summarized below.

1. Estimate North Dakota Values Based on North Carolina Coefficients

When the model obtained in Part I of the pilot testing was applied to home health agencies in North Dakota, the classifications were considerably less accurate. This is the same result as observed for hospitals and long-term care facilities.

2. Conclusion

As was the case for hospitals, this analysis suggests that using coefficients based on home health agencies models estimated in one state achieves lower accuracy when applied to facilities in another state. Additional research would be required to determine whether the decline in accuracy might be related to the extent to which general characteristics of the states are similar or different.

H. Empirical Models for North Carolina Public Health Agencies

The steps used to estimate the empirical models for public health agencies in North Carolina are summarized below.

1. Select Indicator (Dependent) Variable

The indicator of nursing shortage used as a dependent variable was the number of reported effects on operations reported by an agency. Most agencies reported no effects or only one effect. The mean value for all agencies was 1.09, with a standard deviation of 1.03. Based on this, we defined agencies as being needy (for test purposes only) if they reported two or more effects on operations, or more than one standard deviation above the mean. Under this definition, 26.5% of public health agencies were needy.

2. Estimate Medical Need Based on Population Characteristics

The population was adjusted by gender and age based on average use of primary care. This weighted older adults and infants more heavily than younger people and weighted women more heavily than men. The resulting variable was an estimate of how many primary care visits the population would require in a year's time. Although the relationship between use of primary care and need for services such as home health or long-term care is open to debate, this variable was simply a way of standardizing the population based on characteristics known to affect medical need.

3. Select/Construct Independent Variables

1) Active RNs employed in the county per 100,000 adjusted population
2) Students enrolled in RN programs in the county per 100,000 adjusted population
3) Number of short-term general hospitals
4) Number of short-term general hospital beds
5) Ratio of average RN salary to median income
6) Number of nursing and personal care facilities
7) Percent of the population with income below poverty level
8) Population per square mile
9) Ratio of RNs to hospital beds
10) Number of hours per week paid for agency RNs
11) Number of overtime RN hours per week
12) RN vacancy rate
13) RN turnover rate
14) Ratio of LPNs to RNs
15) Total number of budgeted RN positions
16) Percent non-Hispanic white

Table 12. Means and Standard Deviations of Selected Independent Variables Related to Nursing Shortages in North Carolina Public Health Agencies

Independent Variables	All Public Health Agencies		Agencies Reporting No Nurse Staffing Problems		Agencies Reporting Two or More Nurse Staffing Problems	
	Mean	S.D.	Mean	S.D.	Mean	S.D.
Active RNs in County per 100K Medical Need	155.6	92.2	148.5	82.5	162.6	109.6
Students in RN Programs per 100K Medical Need	61.9	220.1	90.9	277.8	16.8	26.24
Number of Short-Term Community Hospitals	1.2	1.1	1.2	1.2	1.1	0.6
Number of Short-Term Community Hospital Beds	343.1	488.3	242.2	378.8	380.0	386.1
Ratio of Average RN Salary to Median Income	1.6	0.3	1.6	0.3	1.6	0.3
Number of Nursing and Personal Care Facilities	10.9	13.3	8.8	11.3	12.3	10.4
Percent of Population w/ Income Below Poverty	14.0	4.1	14.1	3.9	14.8	4.5
Population per Square Mile	184.2	218.2	158.8	210.8	200.3	198.3
Ratio of RNs to Hospital Beds	0.5	0.4	0.4	0.3	0.5	0.5
Number of Hours per Week Paid for Agency RNs	0.9	2.9	0.9	3.2	1.7	3.4
Number of Overtime RN Hours per Week	0.8	2.4	0.7	2.2	1.6	3.4
RN Vacancy Rate	9.0	11.6	8.6	12.2	10.2	10.2
RN Turnover Rate	15.5	18.7	16.9	22.3	15.2	11.4
Ratio of LPNs to RNs	0.1	0.2	0.1	0.2	0.1	0.1
Total Number of Budgeted RN Positions	26.6	27.6	22.9	26.4	26.4	17.1
Percent Non-Hispanic White	71.1	17.1	73.3	63.9	63.9	16.8

Population per square mile was very highly correlated with several other variables, and so a log transformation was applied to avoid problems with multicollinearity. There was also potential multicollinearity between the number of RNs per 100,000 adjusted population, and number of general hospital beds per 100,000 adjusted population. Number of hospital beds was dropped in favor of number of hospitals.

4. Run OLS Regression Model, Full and Abbreviated

The following regression was run for public health agencies.

Table 13. 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**

Dependent Variable: NUM_CONS
Selecting only cases for which FAC_TYPE = public health
R^2 of 0.34

Because most of the variables that appeared most critical were community variables rather than facility variables, an abbreviated model was also run using only community information. Due to general constraints of data availability, the abbreviated model is one that can be used more realistically in practice. The following regression was run for public health agencies, with an R^2 of 0.30, which is only slightly smaller than the R^2 for the full model.

Table 14. Coefficients for Reduced 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**

Dependent Variable: NUM_CONS
Selecting only cases for which FAC_TYPE = public health
R^2 of 0.30

5. Check Predicted versus Actual Scores for Model

Coefficients from the full regression model were used to estimate predicted numbers of problems in each agency. The top 27.2% of agencies in regard to predicted numbers of problems were considered to have made the test cut of 26.5%, chosen arbitrarily, based on earlier analysis (see Step 1). The agencies selected by the full model were compared to the agencies whose actual problem scores were in the top 26.5%.

Using the full model, 25% of public health agencies were not classified correctly based on the arbitrary value chosen earlier. About 14% of agencies were misclassified as not needy by the model when their actual scores qualified them as needy, while about 12% were misclassified as being needy when their actual scores did not qualify them as such.

The full model provided relatively poor predictive value, suggesting that an abbreviated version of the full model was not worth pursuing for public health agencies.

6. Conclusion

Although there are significant predictors of problems related to nursing shortages in public health agencies, the full regression model has a high degree of error in predicting which agencies report the greatest problems. This model does not seem effective to estimate RN shortages in public health agencies. More information may be needed to assess the roles of RNs in public health and the consequences of inabilities to fill RN positions.

I. Empirical Models for North Dakota Public Health Agencies

The coefficients estimated for North Carolina public health agencies were applied to public health agencies in North Dakota. The results are summarized below.

1. Estimate North Dakota Values Based on North Carolina Coefficients

When the abbreviated model obtained in Part I of the pilot testing was applied to public health agencies in North Dakota, the classifications were considerably less accurate. This was the same result observed for hospitals, long-term care facilities, and home health agencies.

2. Conclusion

As was the case for hospitals, long-term care facilities, and home health agencies, this analysis suggested that using coefficients based on long-term care models estimated in one state achieves lower accuracy when applied to facilities in another state. Additional research would be required to determine whether the decline in accuracy might be related to the extent to which general characteristics of the states are similar or different.

J. Ordered Probit Models for North Carolina

Although it is possible (as demonstrated in the analyses in the previous section) to use OLS regression to estimate the relationships between a set of independent explanatory variables and an ordinal dependent variable like "difficulty recruiting RNs," the fact that the dependent variable was ordinal and not Gaussian violates one of the underlying assumptions of OLS regression. One way to address this violation is to use an alternate regression technique, ordered probit analysis. This technique is similar in concept to OLS regression, but uses very different computational procedures. Most important, however, it is designed to work effectively with ordinal dependent variables.

Two different ordered probit models were developed to identify the factors related to difficulty recruiting RNs in the four types of facilities in North Carolina. The first analyzes all four types of facilities simultaneously. The second analyzes the four different types of facilities separately; i.e., hospitals, home health facilities, long-term care facilities, and public health facilities. Both models included variables that represent community characteristics and facility characteristics. The community variables were divided into three groups – demographic, economic, and nursing variables. For each type of facility, the variables included in the model were based on p-values. The lower the p-value of a variable, the stronger the influence the variable had on the nurse recruiting. In other words, lower p-values meant better prediction of difficulty recruiting RNs; therefore variables with lower p-value were included in the model. If p-value was lower than 0.10 then the variable was statistically significant in explaining the shortage at the 10% level of significance.

1. Single Model for Four Facility Types

In this technique, dummy variables for the types of facilities reflect the effects of facility type. By creating interaction variables (which are the products of the dummy variables with other independent variables), this technique provides coefficient estimates for all four types of facilities (hospitals, home health care, long-term care, and public health). The coefficient for an independent variable for one type of facility may be different from the coefficients for the other facility types. In addition, an independent variable may be statistically significant in explaining the recruiting difficulty for one type of facility, but not for another.

The advantage of estimating the model based on all facilities together was that the predicted recruiting difficulty scores were comparable not only within the same type of facility, but also across facilities of different types. The variables included in the model are shown in Table 15. Each variable in the table was statistically significant for at least one type of facility.

Table 16 presents the coefficient estimates for the simultaneous model. The table shows that different types of facilities had different sets of independent variables and therefore different sets of coefficient estimates. For example, the variables selected for hospitals were: metropolitan area, proportion of American Indian and Alaska Native (AIAN), income per capita, number of hospices per 10,000 individuals, a dummy if the county had a hospital with a nursing school, number of hospital full time persons per 10 individuals, facility type, total number of budgeted RN positions, RN vacancy rate, total number of budgeted LPN positions, and RN turnover rate.

The coefficient estimates were used to calculate a predicted nursing recruitment difficulty score for each facility type (as similar to the OLS models). These predicted nursing recruitment difficulty scores were used to create groups of facilities with different predicted levels of difficulty (Table 17).

2. Goodness of Fit of the Model

At least three indicators can be used to measure the goodness of fit of the estimated model in explaining the difficulty recruiting RNs. The first is based on the significance levels of the independent variables included in the model. Lower p-values mean a better estimated model. The fact that many of the p-values for many of the variables in the model are less than 0.10 (bolded values) means the model is a good one (Table 16).

A second indicator of goodness of fit is based on a cross tabulation of the *actual* recruiting difficulty indicator for facilities obtained from the original survey data by the recruiting difficulty indicator based *predicted* by the model (Table 17). If all off-diagonal values in this table were zero, the model would perfectly explain the difficulty in recruiting RNs. A statistical test of goodness of fit can be computed based on this cross tabulation based on the Spearman Rank Order Correlation. This tests the null hypothesis that there is no correlation between actual recruitment difficulty and the predicted recruitment difficulty. Table 17 shows that the Spearman Correlation coefficient is 0.53, which is statistically significantly different from 0 ($p < 0.0005$). This is a second reason to trust this model, although a higher correlation coefficient would make the model even stronger.

Table 15. Variables, Source of Data, and Year of Independent Variables in Ordered Probit Model for North Carolina for 2004

Facility variables

- Type of facility (hospital setting, home health setting, long-term care setting, and public health setting). [North Carolina Center for Nursing (NCCN), 2004]
- Total number of budgeted RN positions/100, representing the size of a facility. (NCCN, 2004)
- Number of RN vacant FTE/100. [NCCN, 2004]
- Total number of budgeted LPN positions/100, representing other profession as a substitute for RNs in a facility. [NCCN, 2004]
- Number of LPN vacant FTE/100. [NCCN, 2004]
- RN turnover/100, representing the quality of management of a facility. [NCCN, 2004]

Community variables:

Demographic conditions in the county where the facility is located

- Indicator of metropolitan area representing the rural/urban. [ARF, 2003]
- Proportion of population age less than 5 years*10. [ARF, 2000]
- Proportion of population age 20 to 65 years. [ARF, 2000]
- Proportion of population older than 65 years. [ARF, 2000]
- Proportion of non-Hispanic White population. [ARF, 2002]
- Proportion of Hispanic population*10. [ARF, 2002]
- Proportion of non-Hispanic Black population. [ARF, 2002]
- Proportion of AIAN population*10. [ARF, 2002]

Economic conditions in the county where the facility is located

- Income per capita/10000. [ARF, 2001]
- Percentage of population in poverty. [ARF, 2001]
- Total Medicaid inpatient days per population. [ARF, 2002]
- Total Medicaid inpatient days per population. [ARF, 2002]

Nurse-related conditions in the county where the facility is located

- Number of RNs per 100 individuals. [ARF, 2000]
- Number of medical records and health information technologists per 1,000 individuals as a proxy for market conditions of other health professionals. [ARF, 2000]
- Number of hospital per 10,000 individuals. [ARF, 2002]
- Number of Hospices per 10,000 individuals. [ARF, 2002]
- Indicator for county having a hospital with nursing school. [ARF, 2002]
- Number of hospital full time personals per 10 individuals. [ARF, 2002]
- Number of nursing home full time personals per 1,000 individuals. [ARF, 2002]
- Ratio of average RN salary to median income. [Census, 2000]

The third goodness of fit indicator is pseudo-R^2, the McKelvey-Zavoina R^2. The higher the value of this pseudo-R^2, the better the accuracy of the model. The value of 0.71 for this statistic shown in Table 16 is high for this kind of model, another indicator that this model is a good one.

Figure 11 presents the distribution of predicted shortage scores for all facilities. The range of the nursing shortage scores for facilities facing difficulty in recruiting RNs was much higher than those for facilities not facing difficulty in recruiting RNs. These predicted values showed that the number of facilities facing difficulty in recruiting RNs was 141 (43.4%), and the number of facilities not facing difficulty in recruiting RNs was 30 (9.2%).

Figure 11. Distribution of the Predicted Nursing Recruitment Difficulty Score Based on All North Carolina Facilities

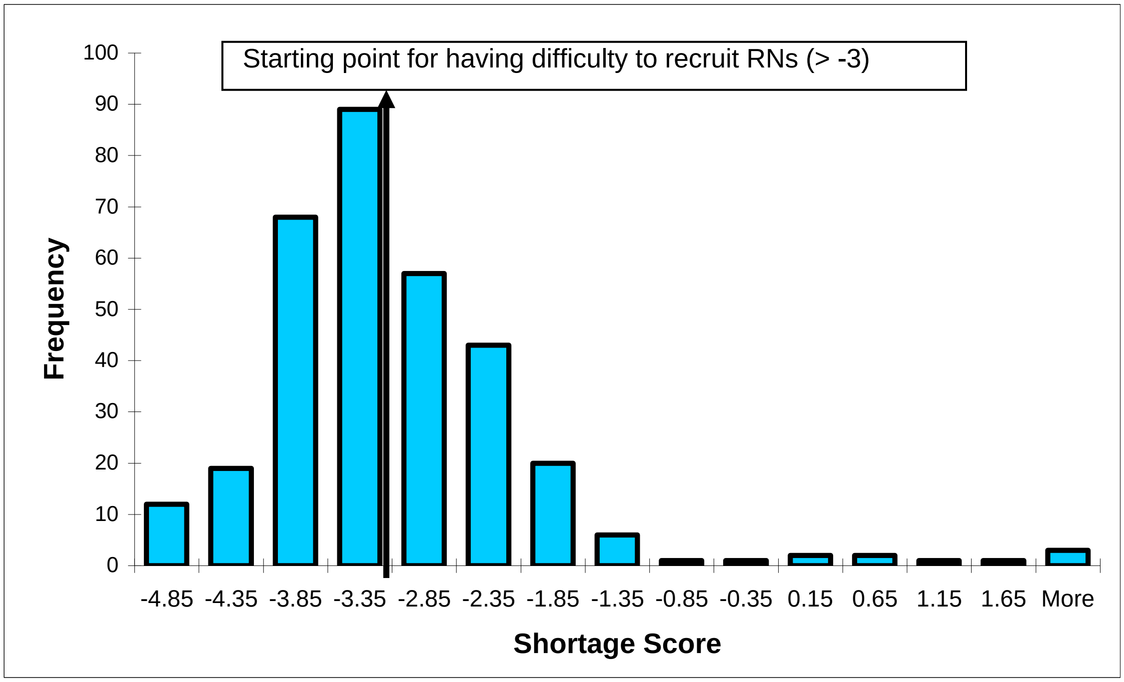


Table 16. Coefficient Estimates of the Ordered Probit Nursing Shortage Model Based on All Facilities in North Carolina

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**

Table 16, continued

Recruiting Difficulty Thresholds
Very easy (1) to recruit if score ≤ -5.494008
Easy (2) to recruit if score ≤ -4.429288
Not difficult (3) to recruit if score ≤ -3.348048
Difficult (4) to recruit if score ≤ -2.158602
Very difficult (5) to recruit if score > -2.158602

McKelvey-Zavoina $R^2 = 0.71$

Table 17. Cross Tabulation of Actual Nursing Recruitment Difficulty Indicator by Predicted Nursing Recruitment Difficulty Indicator

Actual	Predicted					Total
	1	**2**	**3**	**4**	**5**	
1	1	9	7	0	0	17
2	0	10	32	9	0	51
3	0	3	66	31	2	102
4	0	7	40	47	5	99
5	0	0	9	26	21	56
Total	1	29	154	113	28	325

Note: Values on the diagonal are shaded

Spearman correlation coefficient = 0.53

Test of H_0: Correlation = 0

H_0 is rejected with p-value < 0.0005

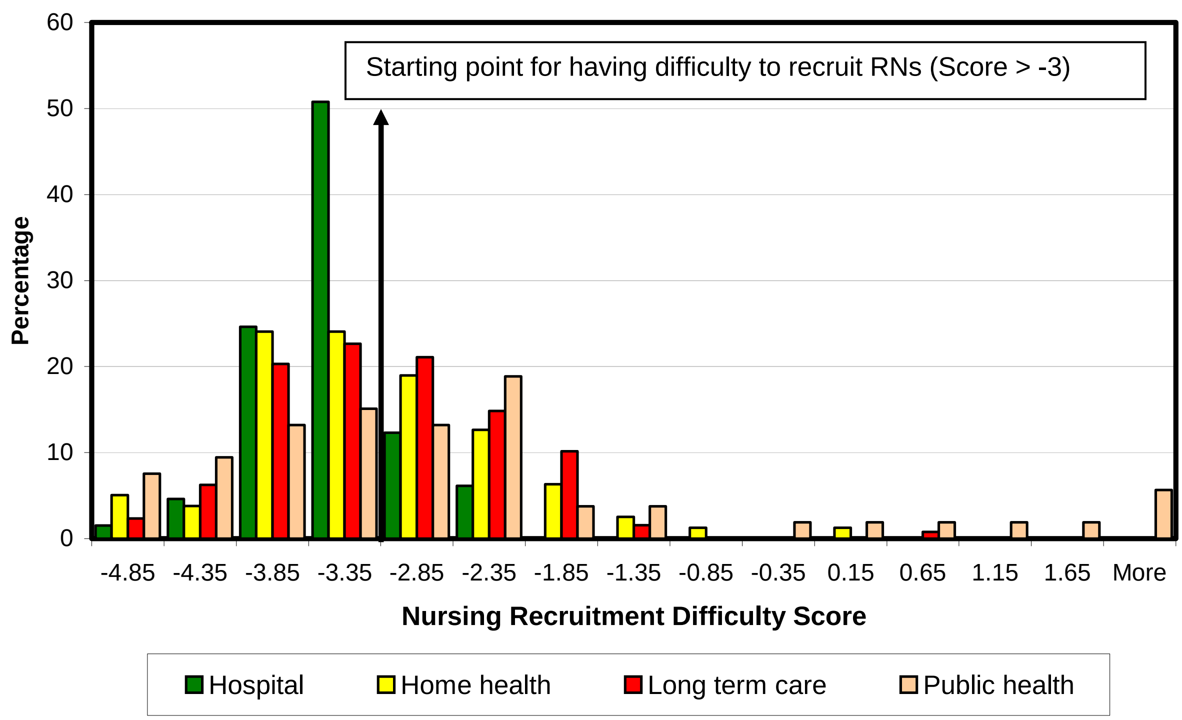
The descriptive statistics of predicted nursing shortage scores by type of facility are presented in Table 18, which shows that on average the shortage was highest for public health and lowest for hospitals. This means that on average public health facilities faced the most nursing recruitment difficulty and hospitals faced the least.

Table 18. Descriptive Statistics of For Predicted Nursing Recruitment Difficulty Score Based on Ordered Probit Model Using North Carolina Data for 2004

Facility Type	Predicted Shortage Score			
	Mean	SD	Minimum	Maximum
Hospital	-3.668	0.466	-5.0797	-2.4393
Home health	-3.391	0.901	-5.1829	0.0163
Long term care	-3.342	0.848	-5.5617	0.5345
Public health	-2.432	3.298	-5.1676	14.8448

Figure 12 shows more clearly the differences in recruiting difficulty among the four types of facilities. The figure presents the distribution of the predicted nursing recruitment difficulty by type of facility. From the figure we can see that a relatively high proportion of public health agencies have high scores (the right side of the figure). This confirms the finding presented above in Table 18.

Figure 12. Predicted Nursing Recruitment Difficulty Scores, by Facility Type in North Carolina, 2004



3. Separate Models for the Four Facility Types

Tables 19 to 22 present the coefficient estimates for hospitals, home health facilities, long-term care facilities, and public health, respectively, based on separate ordered probit models for each type of facility. Similar to Technique 1, using these coefficients one can calculate predicted nursing recruitment difficulty scores for each facility. The key difference is that one cannot compare the predicted nursing shortage scores of different types of facilities. For example, a score for a hospital cannot be compared to a score for a nursing home.

In general, the p-values obtained from separate models were lower than those obtained from the simultaneous model (i.e., the results were more significant statistically). This implied that the number of significant variables obtained from the separate models was greater than the number of significant variables obtained from the simultaneous model. These lower p-values also tell us that the estimation using separate models gave more efficient results.

Both techniques provided very similar patterns of predicted nursing recruitment difficulty scores for each type of facility. The strength of the relationship between predicted nursing shortage scores obtained from the two models can be measured using a correlation coefficient. The Spearman correlation coefficient between the two predicted scores was 0.9985 for hospitals; 0.9911 for home health agencies; 0.9991 for long-term care facilities; and 0.9853 for public health agencies. This meant both techniques gave very similar ranks of predicted nursing recruitment difficulty scores across facilities in North Carolina

Table 19. Coefficient Estimates of the Nursing Recruitment Difficulty Model Based on Ordered Probit Analysis of North Carolina Hospital Data, 2004

Variable	**Coeff.**	**p-value**
Dummy for metropolitan area	-0.47	0.195
Proportion of AIAN population	**1.72**	**0.012**
Income per capita ($10,000)	**0.89**	**0.042**
# of hospital full time personals per 10 individuals	1.44	0.392
# of hospices per 10,000 individuals	-1.24	0.383
Dummy for county having hospital with nursing school	**-1.54**	**0.022**
Total number of budgeted RN positions	**-0.16**	**0.041**
RN vacancy rate	**2.52**	**0.014**
Total number of budgeted LPN positions	**-1.13**	**0.047**
RN turnover rate	2.28	0.209
Threshold 1	-0.41	0.712
Threshold 2	1.25	0.240
Threshold 3	**2.67**	**0.013**
Threshold 4	**4.19**	**<0.0005**

McKelvey-Zavoina R^2 = 0.362

Table 20. Coefficient Estimates of the Nursing Recruitment Difficulty Model Based on Ordered Probit Analysis of North Carolina Home Health Agency Data, 2004

Variable	Coeff.	p-value
Proportion of population >65 years	6.75	0.234
Proportion of population age 20 - 65 years	**22.99**	**0.002**
Proportion of Black population	1.64	0.241
Proportion of Hispanic population	0.88	0.128
Percentage of population in poverty	**-0.18**	**0.022**
# of hospitals per 10,000 individuals	**2.58**	**0.016**
# of nursing home full time personals per 1,000 individuals	**-0.57**	**0.027**
# of hospices per 10,000 individuals	0.37	0.681
Ratio of average RN salary to median income	**2.04**	**0.036**
Total number of budgeted RN positions	-1.91	0.120
RN vacancy rate	**48.77**	**<0.0005**
Threshold 1	**14.58**	**0.005**
Threshold 2	**15.20**	**0.004**
Threshold 3	**16.39**	**0.002**
Threshold 4	**17.61**	**0.001**

McKelvey-Zavoina R^2 = 0.406

Table 21. Coefficient Estimates of the Nursing Recruitment Difficulty Model Based on Ordered Probit Analysis of North Carolina Long-Term Care Facility Data, 2004

Variable	Coeff.	p-value
Dummy for metropolitan area	**-0.69**	**0.026**
Proportion of population < 5 years	**-6.30**	**0.019**
Proportion of population >65 years	**-18.39**	**0.001**
Proportion of Hispanic population	**-1.68**	**0.001**
Proportion of AIAN population	**0.54**	**0.033**
Proportion of population using Medicare	**1.37**	**0.074**
Income per capita ($10,000)	-0.58	0.304
Percentage of population in poverty	-0.10	0.149
# of medical Records & Health Info Techs per 1,000 individuals	**1.68**	**0.020**
# of hospital full time personals per 10 individuals	-2.52	0.151
# of RN's per 100 individuals	**-0.98**	**0.020**
Ratio of average RN salary to median income	**-1.78**	**0.024**
Dummy for county having hospital with professional nursing school	0.38	0.449
Total number of budgeted RN positions	1.98	0.406
RN vacancy rate	**29.88**	**0.032**
LPN vacation rate	13.01	0.148
RN turnover rate	0.18	0.332
Threshold 1	**-14.10**	**<0.0005**
Threshold 2	**-12.86**	**<0.0005**
Threshold 3	**-12.14**	**0.001**
Threshold 4	**-11.11**	**0.002**

McKelvey-Zavoina R^2 = 0.364

Table 22. Coefficient Estimates of the Nursing Recruitment Difficulty Model Based on Ordered Probit Analysis of North Carolina Public Health Agency Data, 2004

Variable	Coeff.	p-value
Dummy for metropolitan area	**-0.65**	**0.083**
Proportion of population >65 years	**26.18**	**0.001**
Proportion of White population	**-41.84**	**<0.0005**
Proportion of Black population	**-33.75**	**0.003**
Proportion of Hispanic population	**-3.01**	**0.012**
Proportion of population using Medicaid	**1.82**	**0.029**
Income per capita ($10,000)	**-2.35**	**0.025**
Percentage of population in poverty	**-0.33**	**<0.0005**
# of hospitals per 10,000 individuals	**-4.32**	**<0.0005**
# of hospices per 10,000 individuals	1.50	0.152
Ratio of average RN salary to median income	**-3.06**	**0.015**
Dummy for county having hospital with professional nursing school	-0.40	0.666
Total number of budgeted RN positions	**-1.59**	**0.054**
RN turnover rate	**5.60**	**0.001**
Threshold 1	**-52.36**	**<0.0005**
Threshold 2	**-51.37**	**<0.0005**
Threshold 3	**-50.34**	**<0.0005**
Threshold 4	**-49.21**	**<0.0005**

McKelvey-Zavoina R^2 = 0.830

4. Models Without Facility Variables

One of the objectives of this study was to assess the importance of facility-specific variables for predicting the difficulty of recruiting RNs and other measures of nursing shortages. Figures 13 through 16 present the results of a series of four comparisons of models, one for each of the four facility types. The figures revealed, based on OLS analysis of data from North Carolina in 2004, that predictions of nurse recruiting difficulty with and without facility data were positively and significantly correlated for all four types of facilities. Similar results were obtained using ordered probit models based on the same data. This result was encouraging for subsequent studies of nursing shortages and related topics because it suggested that, although some predictive accuracy was lost when facility data were not available, at least some helpful insights could be obtained from community data alone.

Figure 13. Comparison of OLS Nursing Recruitment Difficulty Models for Hospitals in North Carolina, With and Without Facility Variables, 2004

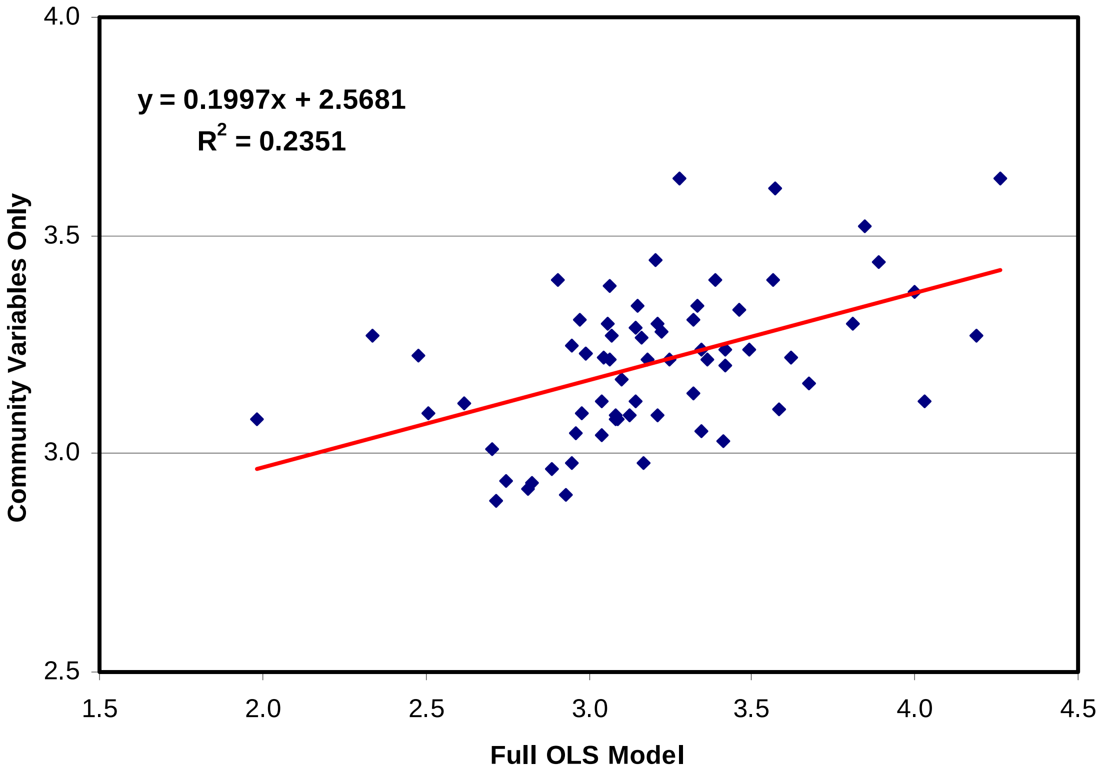


Figure 14. Comparison of OLS Nursing Recruitment Difficulty Models for Home Health Agencies in North Carolina, With and Without Facility Variables, 2004

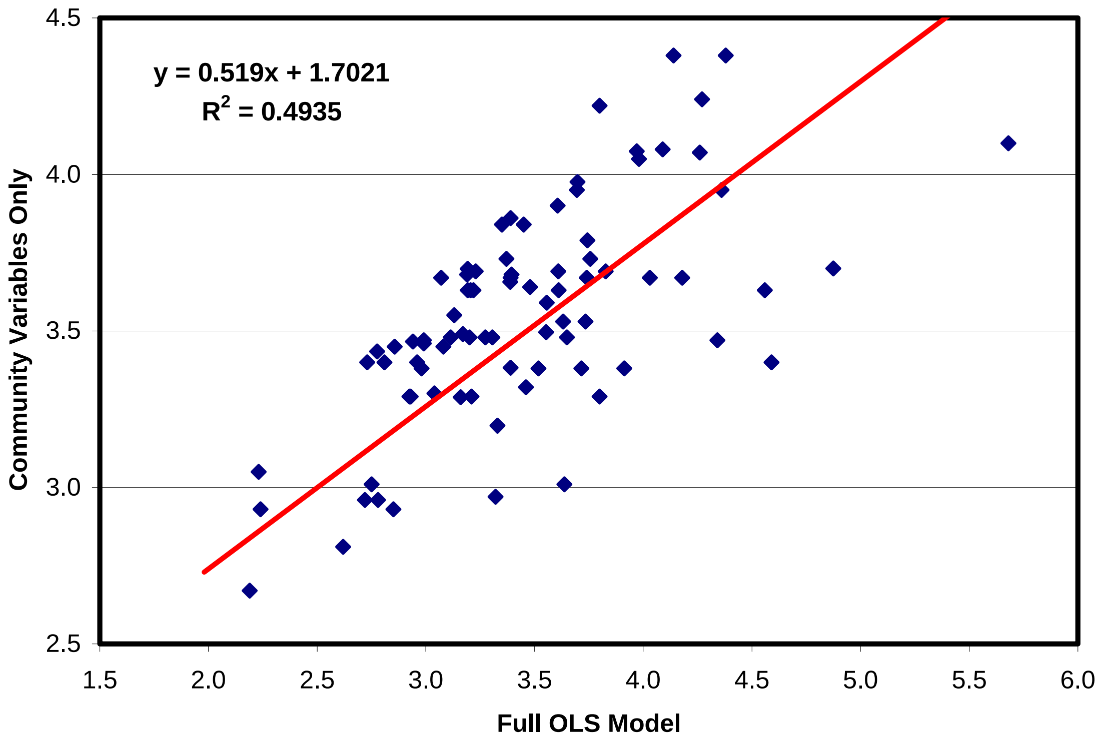


Figure 15. Comparison of OLS Nursing Recruitment Difficulty Models for LTC Agencies in North Carolina, With and Without Facility Variables, 2004

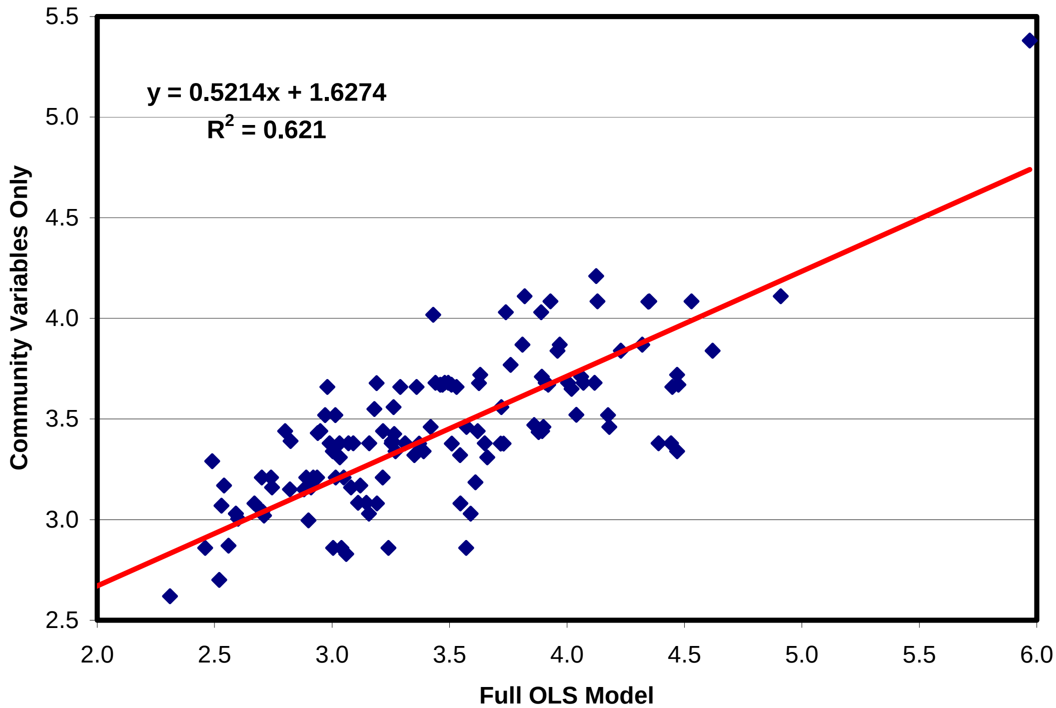
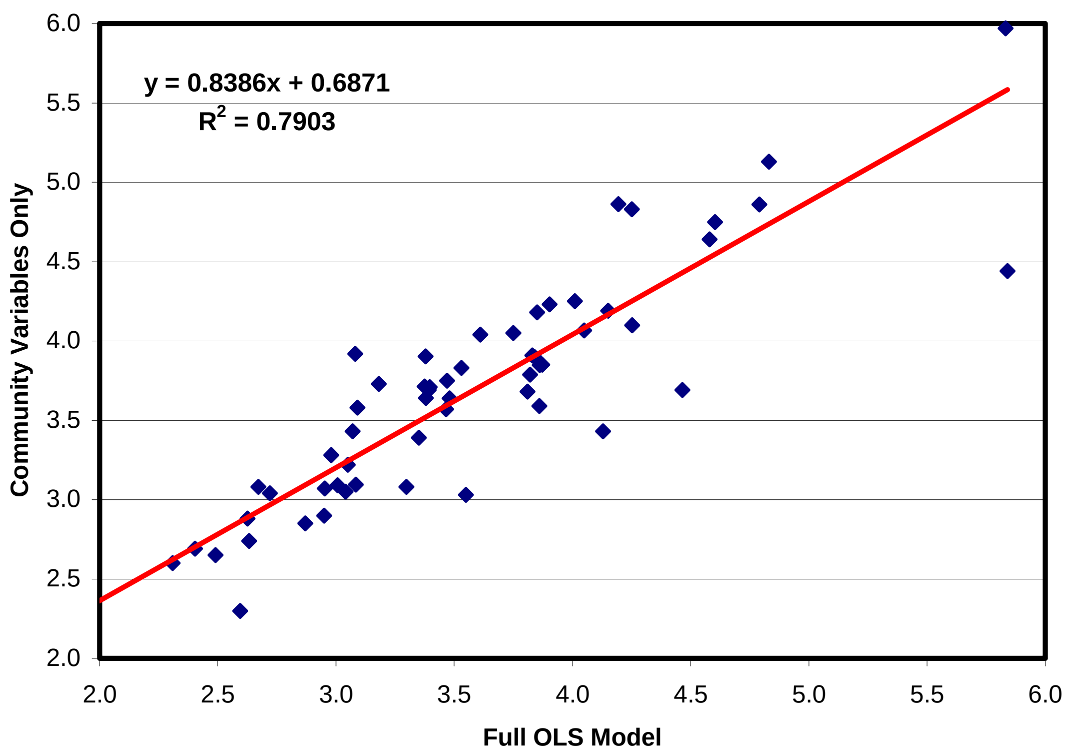


Figure 16. Comparison of OLS Nursing Recruitment Difficulty Models for Public Health Agencies in North Carolina, With and Without Facility Variables, 2004



5. Validation of North Carolina Models

As part of the process of developing and refining the North Carolina ordered probit models, a special validation process was devised to confirm that the values of the nursing recruitment difficulty index predicted by the statistical models were realistic. This process was made more difficult by the requirement of anonymity of the facilities by the NCCN.

The validation procedure used involved sending the anonymous facility ID Codes back to the NCCN for the 10 facilities of each type that had the highest and lowest nursing recruitment difficulty index scores. NCCN staff then attached to each ID Code the name and contact information for each facility on the list. These individuals were then surveyed over the telephone (see Appendix C) asking for insights about the difficulty experienced by the facility in recruiting RNs at the time, six months earlier, and in 2004 (when the original survey data were collected).

When the survey responses were returned, the data were entered into a separate file for analysis. The primary analysis used in this validation was based on a Spearman Rank-Order Correlation coefficient between the variable indicating that the facility was in the top 10 or bottom 10 for its type, and the 5-point scale from the questionnaire rating difficulty of recruiting RNs in 2004. Based on the 48 (out of a possible 80) facilities that responded to the survey questionnaire, the Spearman's Rho was 0.347, $p = 0.016$. Although the correlation between the original rating of recruiting difficulty and the retrospective rating obtained in the validation process was statistically significant, the low value of the correlation coefficient gave little support for the use of these kinds of subjective measures in a formal shortage designation process.

Although this statistical test (that the correlation coefficient = zero) was not particularly stringent, it did provide an indication that the independent variables in the ordered probit model helped to explain variations in nursing recruitment difficulty. Based on this conclusion, project staff moved forward with plans to examine the possibility of using a model estimated in one state to predict nursing recruitment difficulty in another state (in this case, North Dakota).

6. Analysis of North Dakota Data

Data were shared with project staff by two states (North Carolina and North Dakota). Although the North Dakota (ND) data were based on a survey instrument identical in many respects with the North Carolina (NC) questionnaire, the ND survey did not ask the same question about difficulty recruiting RNs that was asked in NC. ND did ask a question about vacancy rates for RNs in the facilities, but unfortunately, the question was answered by only 20% of the respondents. The net result was that the ND data did not provide a sound dependent variable to use in an independent modeling effort similar to that conducted for the NC data.

K. Models for North Carolina and North Dakota Combined

The characteristics of counties in North Carolina and counties in North Dakota differ considerably. For example, using the ARF database, 40% of counties in North Carolina were metropolitan compared to only 8% of counties in North Dakota; 83% of counties in North Carolina had a hospital compared to 64% of counties in North Dakota. In addition, 63% of counties in North Carolina had a hospice compared to 25% of counties in North Dakota.

1. Average Values of Selected Variables in NC and ND

The averages of community variables of counties in North Carolina and in North Dakota are presented in Table 23. The average percentage of Whites in the population was higher in North Dakota than in North Carolina. The average per capita number of hospital beds in North Dakota was more than three times higher than in North Carolina. Although the average per capita number of hospital beds was much higher in North Dakota than in North Carolina, the average per capita number of full time RNs was slightly lower in North Dakota. Moreover, the average percentage of the population in poverty in North Carolina was slightly higher than that of North Dakota, while average per capita income was slightly higher in North Carolina.

Table 23. Means of Community Variables for NC and ND

Variable	State	
	NC	ND
Dummy for metropolitan area	0.400	0.076
Income per capita	$23,520	$22,820
Proportion of White population	0.715	0.920
Proportion of AIAN population	0.016	0.061
Proportion of Black population	0.216	0.003
Proportion of Hispanic population	0.038	0.008
# Hospital beds per 100 individuals	0.295	0.718
# Hospices per 10,000 individuals	0.147	0.276
Hospital full time personals per 10 individuals	0.082	0.068
Total Medicaid inpatient days per population	0.193	0.473
Total Medicare inpatient days per population	0.270	0.429
# Med Records & Health Info Techs per 1,000 individuals	0.158	0.030
Nursing home full time personals per 1,000 individuals	0.310	1.001
Proportion of population < 5 years	0.061	0.055
Proportion of population >65 years	0.141	0.201
Proportion of population age 20 - 65 years	0.579	0.541
Percentage of population in poverty	13.7%	12.2%
# Full time RNs per 100 individuals	0.247	0.227

The differences between North Carolina and North Dakota were not only in terms of the community characteristics, but also in terms of facility characteristics. For example, 76% of facilities in North Dakota reported zero vacancy rates compared to 37% in North Carolina; 53% of facilities in North Dakota reported zero turnover rates compared to 13% in North Carolina. Table 24 presents the averages of facility variables in North Carolina and North Dakota and shows that characteristics of the states' facilities differ considerably. The average number of budgeted RN positions of facilities in North Carolina is almost four times the average in North Dakota. In addition, the average RN vacancy rate of facilities in North Carolina is almost three times the average in North Dakota.

Table 24. Means of Facility Variables for NC and ND

Variable	State	
	NC	ND
Number of budgeted RN positions	79.28	20.59
Number of budgeted LPN positions	9.94	6.14
RN vacancy rate	9.51	3.64
RN turnover rate	29.02	12.29
LPN vacancy rate	7.41	2.46
LPN turnover rate	28.23	6.01

2. OLS Regression Analysis

Using data from both North Carolina and North Dakota together, OLS regression was run for RN vacancy rates on a combination of community variables and facility variables. Vacancy rate was chosen because it was the dependent variable collected using the same definitions in both states. A state dummy variable was included, defined as 1 if a facility was located in North Dakota and 0 if it was located in North Carolina. The model was estimated separately for each type of facility. The coefficient estimates are presented in Tables 25 to 28 for hospitals, home health agencies, long-term care facilities, and public health agencies, respectively. Variables included in the model were selected based on their p-values. Variables with smaller p-values can explain variation in dependent variable better than variables with higher p-values. In addition, adjusted-R^2 was also considered when selecting variables to be included in the model. The higher the adjusted-R^2, the better the model.

The table shows that each type of facility yielded different sets of independent variables that were statistically significant. For example, dummy for North Dakota was <u>not</u> significant in both hospital and long-term care models, while it <u>was</u> significant in both home health and public health models.

The dependent variable in the models was RN vacancy rate. The higher the value of RN vacancy rate the bigger was the shortage. Thus, a positive coefficient revealed that a facility with a higher value of the corresponding independent variable faced a bigger shortage compared to a facility with a lower value of the variable. A negative coefficient revealed that a facility with a higher value of the corresponding independent variable faced less shortage compared to a facility with a lower value. For example, the coefficient estimate of dummy for North Dakota in the home health model was negative. This indicated that on average home health facilities in North Dakota faced less shortage than home health facilities in North Carolina.

Table 25. OLS Coefficient Estimates for Hospital Setting for Combined NC & ND Model (Dependent variable is RN Vacancy Rate)

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 population	**-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 26. Coefficient Estimates for Home Health Setting for Combined NC & ND Model (Dependent variable is RN Vacancy Rate)

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/ hospital w/ professional nursing school	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 *10	-0.0966	0.0606	-1.593	0.114
# Hospitals per 10,000 individuals	-0.0240	0.0217	-1.105	0.272
# Med records and health info techs per 1,000 individuals	0.0607	0.0579	1.0491	0.297
Proportion of population < 5 years *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 27. 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	0.3090
Dummy for North Dakota	0.0392	0.0337	1.1655	0.2457
Income per capita ($10,000)	-0.0324	0.0380	-0.8517	0.3957
Proportion of Hispanic population *10	0.0567	0.0443	1.2798	0.2026
Proportion of population < 5 years *10	-0.2825	0.1789	-1.5792	0.1164
Proportion of population >65 years	-0.5939	0.3957	-1.5007	0.1355
# Full time RNs per 100 individuals	0.0343	0.0459	0.7470	0.4562
Ratio of average RN salary to median income	0.0389	0.0569	0.6834	0.4954
Number of budgeted RN positions	-0.1725	0.1351	-1.2765	0.2037
RN turnover rate	0.0223	0.0199	1.1191	0.2648
LPN vacancy rate	**0.3508**	**0.0754**	**4.6521**	**0.0000**
Number of budgeted LPN positions	**0.4843**	**0.1735**	**2.7920**	**0.0059**

$R^2 = 0.238$

Table 28. 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	0.4859
Dummy for North Dakota	**-0.0848**	**0.0288**	**-2.9455**	**0.0042**
Dummy for County w/ Hospital w/ Prof Nursing School	0.0622	0.0708	0.8792	0.3818
Proportion of AIAN population *10	**0.0652**	**0.0226**	**2.8856**	**0.0050**
Proportion of Black population	0.1231	0.0937	1.3133	0.1927
# Hospitals per 10,000 individuals	**0.0466**	**0.0239**	**1.9535**	**0.0541**
# Hospices per 10,000 individuals	-0.0541	0.0293	-1.8447	0.0686
Total Medicaid inpatient days per population	**-0.0910**	**0.0378**	**-2.4066**	**0.0183**
Proportion of population < 5 years *10	-0.2130	0.1284	-1.6591	0.1008
Percentage of population in poverty	**-0.0081**	**0.0041**	**-1.9653**	**0.0527**
Ratio of average RN salary to median income	**0.1357**	**0.0428**	**3.1690**	**0.0021**
RN turnover rate	0.0710	0.0421	1.6854	0.0956

$R^2 = 0.389$

III. Methods and Models Using Geographic Data Only

The rationale for using a geography-based method to identify facilities with critical shortages of RNs is that recruiting and retention difficulties at the facility level will be strongly influenced by geographic context (e.g., availability of RNs in the immediate geographic area). Certain types of facilities (e.g., long-term care facilities, publicly sponsored facilities) will have greater relative difficulty in obtaining and retaining adequate numbers of RNs in the presence of geographic shortages, but when numbers of RNs available at the local level are adequate to meet the needs of all facilities, inter-facility competition should be a less important factor. Facilities in communities with an adequate supply of RNs may face difficulties in attracting and retaining RNs due to issues related to organizational culture and management practices, but the NELRP program is not intended to address these difficulties.

Potential shortage areas were primarily analyzed at the county level due in large part to data constraints. An obvious shortcoming of county-based analysis was that people often cross county (or even state) lines to seek health care. There were many counties that had no hospitals, for example, but their residents presumably obtained care in other counties.

On the other hand, facilities were likely to draw RNs from the same geographic areas from which they draw patients, and so a shortage of RNs in residence relative to the estimated needs of the population may indicate problems even if both RNs and patients commute to an adjacent county to give or receive care. Clearly, however, the use of counties was inferior to the use of service areas based on actual patterns of health care access, but existing service area designations were badly dated or based on zip codes (to which the necessary data at county or census tract levels did not easily correspond).

Another shortcoming of counties as the unit of analysis relates to shortages in large metropolitan areas. In New York City, for example, all of Manhattan is included in a single county, but neighborhoods within Manhattan vary widely in their economic and demographic characteristics and in their health care infrastructure. Although neighborhoods with high and low levels of resources may even be contiguous to one another, physical and social barriers can prevent both RNs and patients from traveling into other neighborhoods to give or receive care. Therefore, in the largest metropolitan areas, we attempted to replicate the county-level methodology at the level of census tracts. Most of the necessary data were available at the census tract level.

The methodology for defining geographic areas with shortages of RNs was inspired in large part by the Nurse Demand Model (NDM) and Nurse Supply Model (NSM) used by HRSA to project nursing supply and demand. Facilities within shortage areas were then prioritized based upon facility characteristics.

A. National Models Based on County Data

There are several methods used in the literature to estimate the demand or need for RNs. For example, the most commonly used measure is the ratio of RNs to population. In addition, it is also common in the literature to use the ratio of RNs to MDs as a measure of the need for RNs. In this study, we focused on the ratio of RNs to population as a measure of the need for RNs. When calculating this measure, we needed to adjust population size by the compositions of gender and age groups in the population. The next section describes how gender- and age-adjusted population estimates were calculated.

The purpose of this component of the research was to estimate relative need of RNs across counties of the U.S. based on RNs per gender- and age-adjusted population. In addition, as a comparison, the ratio of RNs to MDs was also estimated. Both of these measures were used as a dependent variable in an OLS regression analysis. Data used in this study came from the ARF of 2005 and NHIS of 2003-2004.

1. Model Based on RN to Age-Adjusted Population Ratios

Assumption: RNs should be evenly distributed across the U.S. population adjusting for age.

Assumption: Age-specific patterns of health care utilization do not vary substantially across counties.

Assumption: Need for RNs (as distinct from demand for RNs) is based on population characteristics rather than existing health infrastructure.

Assumption: RN commuting patterns are similar to the commuting patterns of other workers in terms of county inflow and outflow.

This method was an effort to improve upon the basic RN-to-population ratio by applying weights to adjust for the age distribution of the population. The essential idea was similar to that employed in Method 1, but was more limited in what was accounted for to enable the same methodology to be applied to different units of geography so long as basic population data was available. Because older Americans use more health care than the general population, and younger adults use less, this methodology applied a greater weight to older adults than to those ages 18-44. However, the differences in utilization rates by age differed by type of health care, and types of health care differed in their demand for RNs (e.g., about 45% of the demand for RNs was inpatient hospital demand, while only 6% of RN demand was nursing home demand). The weights took into account both estimated use of various forms of health care and the influence of these types of health care on the national demand for RNs.

Table 29. National Estimates of RNs per Unit of Service

RNs	RNs	Units of Care	RNs per Unit	Total RNs	% RNs
Inpatient units	1,058,242	168,846,928	0.0062675	2,375,792	45%
Outpatient units	145,118	83,715	1.7334767	2,375,792	6%
Physician offices	278,093	951,214	0.2923559	2,375,792	12%
Emergency Department	117,381	107,490	1.0920179	2,375,792	5%
Long-term hospitals	139,091	22,402,741	0.0062087	2,375,792	6%
Extended care	153,366	1,469,500	0.1043661	2,375,792	6%
Home health agency	106,690	1,355,290	0.0787212	2,375,792	4%
Nursing education	63,833	2,375,792	0.0268681	2,375,792	3%
Public/community health	87,952	280,836,834	0.0003132	2,375,792	4%
School health	78,539	49,036,764	0.0016016	2,375,792	3%
Occupational health	22,569	173,907,572	0.0001298	2,375,792	1%
Other	124,918	280,836,834	0.0004448	2,375,792	5%

Source: 2004 NSSRN

Utilization rates[1] for specific age groups were then standardized against the average for the population overall to obtain ratios of how many persons an individual from a specific age group should count as in calculating utilization. For example, overall use of inpatient days in the U.S. in 2002[2] was 541.3 per 1,000 population, while those ages 0-5 averaged 601.0 days of care (a ratio of 1.11). Therefore, each person ages 0-5 would be weighted as 1.11 persons for the purposes of determining demand for inpatient care. In contrast, those ages 6-17 averaged 110.3 days of inpatient care per 1,000 (a ratio of 0.20). Therefore, each person ages 6-17 would be weighted as 0.20 persons in determining demand for inpatient care.

Utilization of each type of care was further adjusted for the relative influence of that type of care on demand for RNs overall. For example, the greatest driver of demand for RNs was demand for inpatient services (where 45% of RNs are employed), while non-emergency hospital outpatient services influenced overall demand for RNs much less (as they only employ 6% of RNs). Each age group's weight for inpatient services was then multiplied by 0.45; each age group's weight for hospital outpatient services was multiplied by 0.06, etc. Adjusted weights for each type of care for each group were then summed to produce the group's total weight, which should be a reflection of how many people each individual "counts as" in determining overall RN demand. A constant adjustment factor was then applied to the adjusted weight for each group so that the weighted population totals equaled the actual U.S. population.

Final weights for each group are shown below in Table 30. As shown, the age group that exerted the least influence on demand for RNs (ages 6-17) was weighted at about half a person, while the group that exerted the most influence (ages 85 and up) was weighted at about five persons.

Table 30. Final Population Weights by Age Group

Age group	Final Weight
0 to 5	0.890
6 to 17	0.511
18 to 44	0.690
45 to 54	0.897
55 to 64	1.078
65 to 74	1.947
75 to 85	3.367
85 and up	5.024

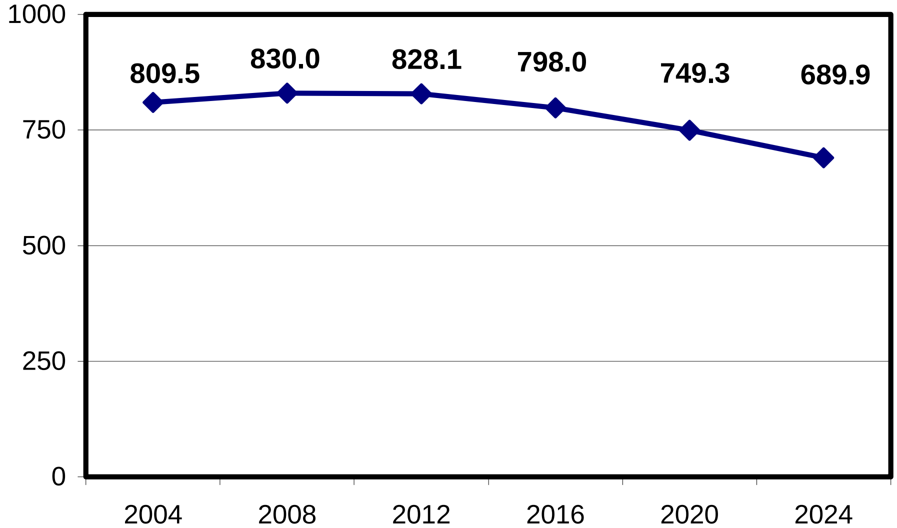
These weights were applied to the population of each county to produce a weighted population count that reflected demand for RNs more accurately than simply an unweighted population ratio.

Figure 17 below shows the projected number of RNs per 100,000 age-adjusted population (see Table 44 for the weights used to adjust the population for patterns of health care use by age).

[1] Taken from *Health,United States, 2005*
[2] Data were not published for 2000.

As the figure illustrates, the supply of RNs relative to the age-adjusted population will peak in 2008, decline very slightly by 2012, and decline further by 2016. By 2024, the relative supply of RNs is estimated to be 15% less than the 2004 level.

Figure 17. Projected RNs per 100,000 Age-Adjusted Population



2. Age-Gender Adjusted Population

When comparing RNs per population across counties, it is important to consider the age-gender distribution in each county, because this is an important determinant of health care utilization. A county with a higher proportion of older adults needs more RNs compared to a county with a lower proportion of older adults, even if the counties have the same other characteristics. However, numerous other factors could affect the need for RNs. For example, a county with higher morbidity rate needs more RNs compared to counties with a lower rate.

Health care utilization rate is commonly used in adjusting population to calculate the ratio of physicians to population. For example, the Sheps Center for Health Service Research, at the University of North Carolina at Chapel Hill, has used adjusted population to estimate physician per population ratios. A similar procedure was used in this study, but it focused on RNs per population rather than physicians per population.

Health care utilization rate was estimated based on the number of nights in hospital (inpatient days) and the number of visits to health care professionals including emergency department (outpatient visits). The utilization rates were estimated based on a sample of non-Hispanic Whites who had health plans obtained from NHIS 2003-2004. Table 31 presents the health care utilization rates by age and gender categories for inpatient days and outpatient visits.

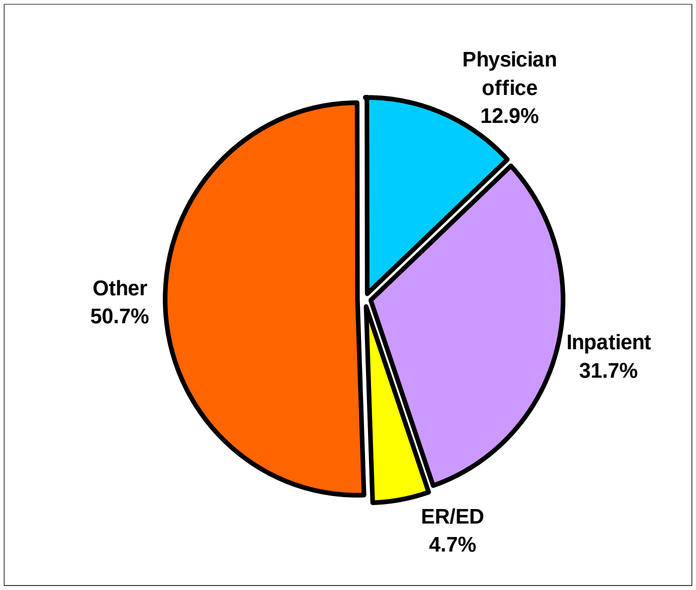
Figure 18 shows the distribution of RN hours spent in direct patient care consisting of physician office, hospital inpatient, and emergency department (ED), and in non-direct patient care. These proportions were obtained from the 2000 NSSRN. The distribution presented in the figure was used to aggregate inpatient days, outpatient visits, and ED visits estimated from NHIS data. Please note that in the NHIS dataset physician office visits and emergency room visits were combined into one variable. Thus, in this study the percentage of RN hours spent in physician office and emergency room was consolidated into one value, 17.6%.

Table 31. Inpatient and Outpatient Health Care Utilization by Age and Gender, 2003-04

Age Group	Inpatient Days / Year		Outpatient Visits / Year	
	Male	Female	Male	Female
0 – 4	0.741	0.759	0.255	0.240
5 - 9	0.054	0.053	0.153	0.132
10 – 14	0.064	0.040	0.124	0.119
15 – 19	0.127	0.225	0.125	0.178
20 – 24	0.296	0.634	0.120	0.238
25 – 29	0.344	0.606	0.121	0.283
30 – 34	0.173	0.581	0.134	0.273
35 – 44	0.207	0.424	0.156	0.279
45 – 54	0.477	0.387	0.209	0.314
55 – 59	0.762	0.685	0.298	0.369
60 – 64	0.880	0.843	0.347	0.389
65 – 74	1.152	1.112	0.354	0.405
75 – 84	1.592	1.963	0.483	0.435
≥ 85	3.567	2.159	0.512	0.398
Average	**0.746**	**0.748**	**0.242**	**0.289**

Source: NHIS 2003-2004

Figure 18. Distribution of RN Hours Spent in Direct Patient Care (Physician Office, Inpatient, ER/ED) and Other Activities



Source: Calculated based on data from the National Sample Survey of RNs, March 2000

The next step was to calculate the weight corresponding to each age-gender group. To illustrate, the weight for males ages 5-9 years was computed as follows:

$$\textbf{Weight} = 0.317 \times (0.054/0.746) + 0.176 \times (0.153/0.242) + 0.507 = 0.6410$$

The final weights for all age-gender groups are presented in Table 32.

Table 32. Weights for Age-Gender Adjusted Population

Age group (year)	Weight	
	Male	Female
0 - 4	1.0072	0.9746
5 - 9	0.6410	0.6098
10 - 14	0.6244	0.5962
15 - 19	0.6516	0.7106
20 - 24	0.7199	0.9208
25 - 29	0.7410	0.9356
30 - 34	0.6783	0.9195
35 - 44	0.7083	0.8563
45 - 54	0.8621	0.8619
55 - 59	1.0474	1.0216
60 - 64	1.1333	1.1010
65 - 74	1.2541	1.2248
75 - 84	1.5354	1.6036
≥ 85	2.3960	1.6637

Note: Estimated based on NHIS 2003-2004 data and the National Sample Survey of RNs of 2000

The final step was calculating age-gender adjusted population using the weights presented in Table 32. The age-gender adjusted population for County C was calculated as the weighted sum of populations of all age-gender groups, formulated as follows:

$$\textbf{Adjusted Pop} = 1.0072 \times (\text{\# Males 0-4}) + \ldots + 2.3960 \times (\text{\# Males} \geq 85) + 0.9746 \times (\text{\# Males 0-4}) + \ldots + 1.6637 \times (\text{\# Males} \geq 85)$$

This method is similar to the method commonly used to calculate the base population to estimate the need for physicians in a specific county, state, region, or other geographic area.

In the first model specification, a dependent variable defined as the ratio of RNs per 1,000 age-gender adjusted population was generated. In addition, as a comparison, a model with RNs per MD as the dependent variable was also developed. The distributions of these two dependent variables are presented in Table A-1 to A-3 in Appendix A for states, regions, and rural and urban areas, respectively.

3. OLS Regression Analysis

RNs per Age-Gender Adjusted Population as the Dependent Variable

The dependent variable in the first specification model was the ratio of RNs to age-gender adjusted population. Explanatory variables used in the analysis are as follows:

1. Dummies for 9 census divisions with the Pacific region used as reference (8 dummies: *dr1 - dr8*)
2. Dummies for metropolitan counties, with Non-Metropolitan County used as reference (3 dummies: dm1 – dm3)
3. Percentage of population ages 5 years or younger(pp_5)
4. Percentage of population ages 65 years or older (pp65_)
5. Percentage of Black and Hispanic population (blck_hsp)
6. Percentage of American Indian and Alaska Native population (AIAN)
7. Percentage of population in poverty (pvrtypct)
8. Infant mortality rate (infmortr)
9. Percentage of agriculture/forest/fish/hunt/mine workers (agricpct)
10. Percentage of manufacturing workers (manufpct)
11. Percentage of health and social service workers (healthpct)
12. Percentage of white collar workers (whcollar)
13. Dummy for the number of hospital in the county is more than one (*dhsp2*)
14. MDs per 1,000 individuals (md_pop)
15. Medicare inpatient days per 100 individuals (*mdicr_pop*)

The descriptive statistics for each of these variables are presented in Tables A-4 to A-8 in Appendix A.

Table 33 presents the coefficient estimates for the first model based on county data from the ARF of 2005. The table shows that the coefficient estimates of the dummies for regions were significant and positive. These tell us that the regions represented by the eight dummy variables had significantly higher RNs per age-gender adjusted population than the Pacific region. The coefficient estimates of the regions varied considerably ranging from 0.515 for Mountain to 2.378 for East South Central. The coefficient estimates of dummies for metropolitan counties were positive and significant indicating that counties of metropolitan areas had higher RNs per gender-adjusted population than non-metropolitan counties with similar other characteristics.

Table 33. Estimates of Impact of Selected Factors on RNs per Age-Gender Adjusted Population

Independent Variable	Coefficient	Std Err	t-stat	p-value
Intercept	-0.137	0.841	-0.163	0.870
New England	1.682	0.353	4.765	0.000
Middle Atlantic	1.367	0.281	4.867	0.000
East North Central	1.770	0.235	7.536	0.000
West North Central	1.708	0.228	7.505	0.000
South Atlantic	1.617	0.226	7.164	0.000
East South Central	2.378	0.251	9.456	0.000
West South Central	1.024	0.228	4.482	0.000
Mountain	0.515	0.239	2.156	0.031
Counties of metro areas of 1 million population or more	0.369	0.172	2.143	0.032
Counties in metro areas of 250,000 - 1,000,000 pop.	0.864	0.161	5.368	0.000
Counties in metro areas of fewer than 250,000 pop.	0.697	0.149	4.678	0.000
Percentage of population age 5 or younger	0.139	0.058	2.376	0.018
Percentage of population age 65 years or older	0.032	0.015	2.095	0.036
Percentage of Black and Hispanic population	-0.019	0.004	-5.335	0.000
Percentage of AIAN population	-0.035	0.007	-5.032	0.000
Percentage of population in poverty	-0.158	0.013	-11.835	0.000
Infant mortality rate	-0.013	0.011	-1.214	0.225
Percentage of agriculture/forest/fish/hunt/mine workers	-0.041	0.008	-5.011	0.000
Percentage of manufacturing workers	0.020	0.008	2.481	0.013
Percentage of health and social service workers	0.196	0.013	15.019	0.000
Percentage of white collar workers	0.064	0.010	6.356	0.000
Dummy for county having 2 or more hospitals	0.221	0.108	2.043	0.041
Number of MDs per 1,000 individuals	0.203	0.040	5.107	0.000
Medicare inpatient days per 100 individuals	0.263	0.128	2.060	0.039

Note: Estimated using OLS regression based on data from Area Resource File of 2005
$R^2 = 0.43$

The coefficient estimate of proportion of population age 5 years or younger was positive and significant. This revealed that the higher the proportion of population age 5 or younger, the higher was the RNs age-gender adjusted population. The coefficient estimate of proportion of population age 65 or older was positive indicating that the higher the proportion of population age 65 years or older, the higher was the RNs per age-gender adjusted population. The coefficient estimate of proportion of Black and Hispanic populations was negative and significant indicating that the ratio of RNs to age-gender adjusted population was lower in counties with higher proportion of Black and Hispanic populations. Similar to the proportion of Black and Hispanic populations, the proportion of AIAN population was negatively associated with the RNs per age-gender adjusted population. Thus, the higher the proportion of AIAN population in counties, the lower was the RNs per age-gender adjusted population.

The economic condition of a county was represented by the percentage of population in poverty. Its coefficient estimate was negative and significant, which indicated that lower the economic condition of a county, the lower was the RNs per age-gender adjusted population. It was noteworthy that the economic condition had a negative correlation with the percentage of minority populations (Black, Hispanic, and AIAN). Therefore, the higher the number of minority population, especially Black, Hispanic, and AIAN, the lower the economic condition. The other variable which had a high negative correlation with economic condition was infant mortality rate. The coefficient estimate of infant mortality rate was negative, but not significant.

The variables representing the structure of labor markets were percentage of agriculture/forest/fish/hunt/mine workers, manufacturing workers, health and social service workers, and white collar workers. Table 33 shows that all coefficient estimates were statistically significant, except the coefficient estimate for infant mortality rate. The highest coefficient estimate of the percentage of health and social service workers was the highest compared to the others. This indicated that the percentage of health and social service workers in a county was the most influential factor in attracting people to enter RNs as their profession.

The number of hospitals in a county also affected the RNs per age-gender adjusted population. When the number of hospitals was included in the model, the coefficient was insignificant. In addition, the dummy for a county having at least one hospital was also insignificant. When a dummy variable defined as a county with two or more hospitals was included in the model, its coefficient was significant and positive. A county with one hospital and a county without a hospital, with the same other characteristics, tended to have the same RNs per age-gender adjusted population. But if a county had two or more hospitals, the number of RNs per age-gender adjusted population was higher compared to a county without a hospital or with only one. This indicated that in a county with more than one hospital, the demand for RNs was more competitive compared to a county without a hospital or only one. The more competitive the market from the demand side the higher was the salary; in subsequence it would attract more people to enter the nursing profession.

The MDs per 100 individuals and Medicare inpatient days per 100 individuals were also included as explanatory variables. Both variables had positive coefficient estimates and were statistically significant. The more MDs per individual in a county, the greater the RNs per age-gender adjusted population. In addition, the more Medicare inpatient days per individual, the greater the RNs per age-gender adjusted population.

4. RNs per MD as the Dependent Variable

Assumption: RNs should be evenly distributed according to locations of physicians.

Assumption: RN commuting patterns are similar to the commuting patterns of other workers in terms of county inflow and outflow.

The RN to physician ratio was expected to produce an estimate that was closer to that based on actual utilization data, as physician counts were likely to be in part a proxy for health care infrastructure. This was more useful than utilization rates in that it could be adapted to geographies or time periods where utilization rates were not available, but was a less precise measure and could bias RN shortage estimates against areas that were physician-short.

Example: Albany County

It was estimated that 4,942 RNs and 1,578 physicians were working in Albany County. If we assumed that national RN to physician ratios should be distributed evenly throughout the country, we would expect 3.11 RNs to every physician in Albany County. This would require 4,907 RNs in Albany County.

Although this method still resulted in a slight estimated oversupply of RNs in Albany County, this was only 1% more RNs than needed, rather than the 110% oversupply indicated by Method #2. This method accounted for the greater health care infrastructure in Albany relative to surrounding areas, which demanded more RNs per capita than a simple RN to population ratio would indicate.

As a comparison, we also estimated a model with RNs per MD as dependent variable. The explanatory variables for this specification were the same as those for the first specification (RNs per age-gender adjusted population as the dependent variable). Table 34 presents the coefficient estimates for this specification. The R^2 was 0.25 for this model, which was much lower than that of the first specification ($R^2 = 0.43$). In addition, some of the coefficients were not significant which indicated that the specification with RNs per age-gender adjusted population as the dependent variable was better than the specification with RNs per MD as the dependent variable.

It should be noted that just because a county had high RNs per MD did not mean the county had enough RNs. This could be explained by a low number of MDs in the county. For example, HPSA counties, which had low numbers of MDs, had higher RNs per MD. Specifically, on average, the average RNs per MD among HPSA counties was 16.9, which was about twice the average of the non-HPSA and partial-HPSA counties, at 8.9 and 8.3, respectively. In contrast, the average RNs per age-gender adjusted population among HPSA counties was 6.6 compared to 8.8 for non-HPSA counties and 8.2 for partial-HPSA counties. Thus, one must be careful when interpreting and using RNs per MD as an indicator of nursing shortage.

Table 34. Estimate of Impact of Selected Factors on RNs per MD

Independent Variable	Coefficient	Std Err	t-stat	p-value
Intercept	18.502	3.901	4.743	0.000
New England	1.311	1.512	0.867	0.386
Middle Atlantic	1.892	1.215	1.558	0.119
East North Central	3.438	1.021	3.368	0.001
West North Central	6.760	0.995	6.797	0.000
South Atlantic	2.881	0.981	2.938	0.003
East South Central	4.100	1.085	3.779	0.000
West South Central	2.746	0.994	2.762	0.006
Mountain	0.734	1.044	0.702	0.482
Counties of metro areas of 1 million pop. or more	4.811	0.743	6.478	0.000
Counties in metro areas of 250,000 - 1,000,000 pop.	4.302	0.689	6.242	0.000
Counties in metro areas of fewer than 250,000 pop.	4.557	0.644	7.077	0.000
Percentage of population age 5 years or younger	-0.661	0.266	-2.485	0.013
Percentage of population age 65 years or older	0.059	0.069	0.854	0.393
Percentage of Black and Hispanic population	-0.012	0.016	-0.754	0.451
Percentage of AIAN population	-0.015	0.032	-0.472	0.637
Percentage of population in poverty	-0.085	0.061	-1.391	0.164
Infant mortality rate	-0.003	0.056	-0.060	0.952
Percentage of agriculture/forest/fish/hunt/mine workers	0.182	0.043	4.256	0.000
Percentage of manufacturing workers	0.079	0.035	2.221	0.026
Percentage of health and social service workers	0.252	0.058	4.322	0.000
Percentage of white collar workers	-0.226	0.046	-4.911	0.000
Dummy for county having 2 or more hospitals	-3.006	0.463	-6.494	0.000
Number of MDs per 1,000 individuals	-2.064	0.172	-12.009	0.000
Medicare inpatient days per 100 individuals	-1.630	0.569	-2.862	0.004

Note: Estimated using OLS regression based on data from Area Resource File of 2005
$R^2 = 0.25$

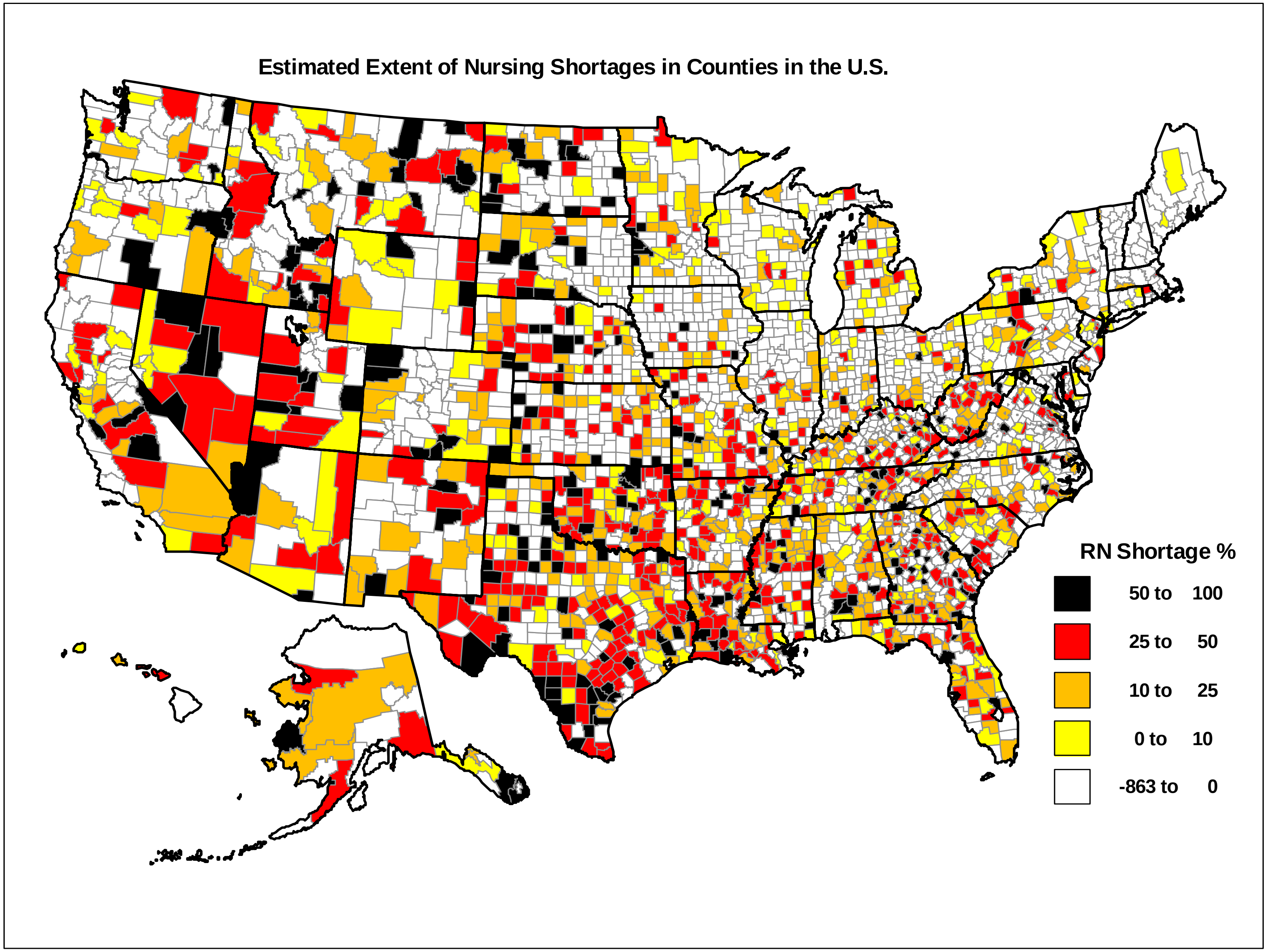
5. Distribution of Residuals

Table 35 presents the distribution of the percentage of counties with negative residual by states. (The residual was defined as the actual value of the dependent variable less its predicted value, so that a negative value indicated that a state has fewer RNs than the model predicts.) Based on the first specification, the table shows that—apart from District of Columbia—Utah had the highest percentage of counties with negative residual (83%). In the other words, 83% of counties in Utah had lower RNs per age-gender adjusted population than predicted by the model. In contrast, Hawaii and Montana had the lowest percentage of counties with negative residuals (25% each).

Table 35. Percentages of Counties in the U.S. with Negative Residuals

FIPS	State	% of Counties with Negative Residuals	
		Model 1	Model 2
1	Alabama	52%	63%
2	Alaska	59%	33%
4	Arizona	53%	40%
5	Arkansas	33%	66%
6	California	57%	54%
8	Colorado	38%	49%
9	Connecticut	63%	13%
10	Delaware	33%	33%
11	District of Columbia	100%	0%
12	Florida	63%	73%
13	Georgia	50%	67%
15	Hawaii	25%	0%
16	Idaho	73%	61%
17	Illinois	30%	49%
18	Indiana	58%	79%
19	Iowa	36%	48%
20	Kansas	44%	69%
21	Kentucky	57%	70%
22	Louisiana	44%	55%
23	Maine	56%	81%
24	Maryland	38%	50%
25	Massachusetts	36%	50%
26	Michigan	75%	79%
27	Minnesota	72%	87%
28	Mississippi	28%	52%
29	Missouri	67%	59%
30	Montana	25%	60%
31	Nebraska	58%	75%
32	Nevada	77%	67%
33	New Hampshire	50%	50%
34	New Jersey	67%	43%
35	New Mexico	42%	53%
36	New York	55%	50%
37	North Carolina	31%	65%
38	North Dakota	62%	81%
39	Ohio	46%	78%
40	Oklahoma	57%	45%
41	Oregon	56%	65%
42	Pennsylvania	45%	67%
44	Rhode Island	40%	80%
45	South Carolina	44%	83%
46	South Dakota	41%	89%
47	Tennessee	70%	79%
48	Texas	62%	70%
49	Utah	83%	62%
50	Vermont	64%	71%
51	Virginia	55%	59%
53	Washington	46%	46%
54	West Virginia	60%	60%
55	Wisconsin	71%	86%
56	Wyoming	52%	74%

Figure 19. Estimated Extent of Nursing Shortages in Counties in the U.S.



B. Model Based on RN to Population Ratios

Assumption: RNs should be evenly distributed across the U.S. population.

Assumption: Need for RNs (as distinct from demand for RNs) is based on population characteristics rather than existing health infrastructure.

Assumption: RN commuting patterns are similar to the commuting patterns of other workers in terms of county inflow and outflow.

Method 2 uses a simple, RN-to-population ratio and is based upon the assumption that RNs should be evenly distributed across the U.S. population. Method 2 is a very crude measure because it does not take into account either the age structure of the population at the county level or the health care infrastructure in the county. Like Method 1, it adjusts RN supply based on inter-county commuting patterns.

Example: Albany County, New York

As calculated in Step 4 of Methodology #1, 4,942 RNs were estimated to work in Albany County in 2000. The population of Albany County in 2000 was estimated to be 294,565. Applying national ratios of 0.0080 RNs per population, we would expect Albany County to need a total of 2,357 RNs (294,565 x 0.0080). The actual supply of RNs was estimated to be 110% more than what the population of the county required.

Albany County is a good illustration of the shortcomings of this method. Because it is an urban center with many hospitals and other health care facilities, many residents of surrounding counties come to Albany County for care. Even though there are facilities in most of the surrounding counties, Albany Medical Center is a Level I trauma center and a teaching hospital, and both Albany Medical Center and St. Peter's Hospital (also in Albany) score highly on national rankings of patient care.

C. Models Based on County Clusters

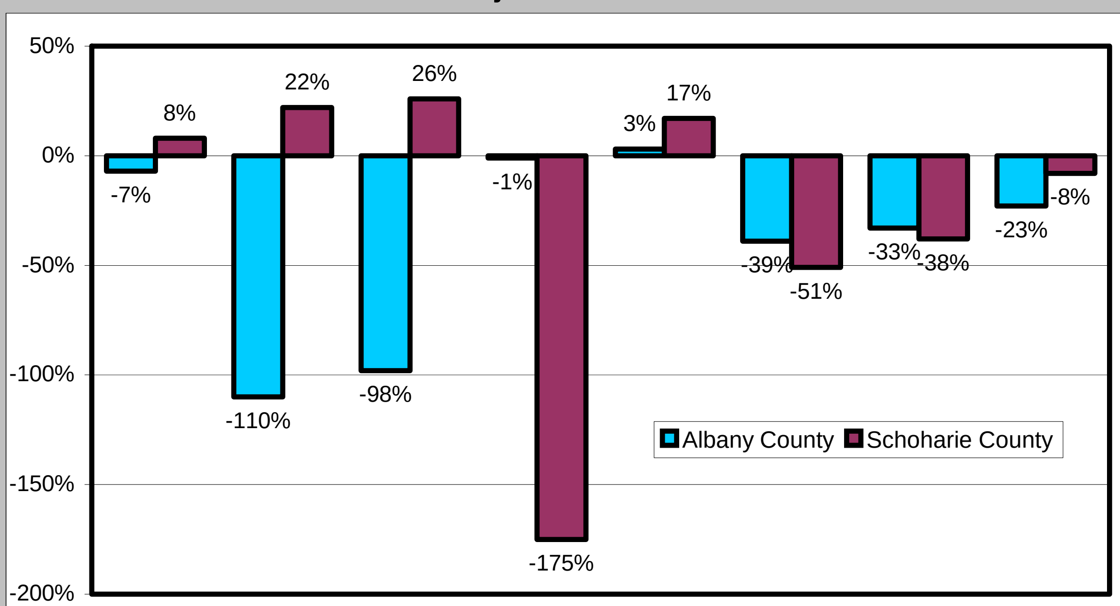
One of the obvious biases when Methods 1 to 4 were compared was that a county in which health care facilities drew many patients from outside the county, the county was shown to have more severe shortages than counties in which patients presumably traveled to other counties for health care. This was a clear problem in any methodology based solely on population. In an attempt to assess the impact of cross-county patient flow, Methods 1 to 4 were recalculated at the level of "county clusters," where population counts, nurse counts, and demand estimates at the county level were summed for a core county and its contiguous counties. This was an imperfect measure, as contiguous counties will have a patient flow to and from the core county in the cluster, but also to and from their own other contiguous counties. For example, if County A has a contiguous County B to the west, County B's population is considered part of County A's county cluster. However, if County B is bordered on the west by County C, which is part of a major metropolitan area, County B's population may be primarily going to County C for health care with very little flow to County A. Counting the population of County B as part of County A's county cluster will therefore result in an overestimate of the pool of people who may be using health services in County A.

On the other hand, the use of county clusters was expected to have a smoothing effect across the various types of estimates, which was generally observed. For example, in Albany County, estimates of RN supply ranged from a supply that was 1% greater than demand to a supply that was 110% greater than demand. In the Albany county cluster, however, estimates ranged from a supply that was a 3% shortage to a supply that was 39% more than estimated demand.

Example: Albany and Schoharie Counties in Upstate New York

Figure 20 below summarizes how the various measures of shortage differ for a feeder county and a receiver county that are contiguous to one another. Schoharie County was a rural county adjacent to Albany County. None of its other contiguous counties hosted major medical centers comparable to those in Albany County, so persons in Schoharie County were more likely to go to Albany County than to any other contiguous county for care. In the first four measures of shortage, at the individual county level, Albany County was seen as having a surplus while Schoharie County was seen as having a shortage. When county clusters were used, however, estimates for the two neighboring counties were much more similar.

Figure 20. Comparison of Selected Measures of Nursing Shortage in Adjacent Counties



D. Models Based on Adjusting for Cross-County Patient Flow

Another method to adjust for cross-county patient flow more precisely than using county clusters was to adjust population figures based upon commuting flow. In one respect it made sense that the distances and directions in which it was convenient for people to travel to work would also be convenient for them to travel for health care, and that counties with more job opportunities relative to their neighbors would also have more health care facilities. On the other hand, the nature of health care need dictates that some areas may have health facilities but few other major employers.

Furthermore, there were sometimes additional inducements to commute out for work that did not exist in commuting for health care. For example, in Monroe County, Pennsylvania, 7% of workers who lived in the county commuted to one of the five counties of New York City for work (a distance of approximately 80 miles that cannot be traveled without crossing through at least three other counties) due to the great differences in salaries (favoring working in New York City) and the great differences in cost of living (favoring living in Pennsylvania). Yet Monroe County has a medical center, and is contiguous to several other counties with major medical centers (including some with trauma centers) that are not nearly as far as New York City. Therefore, it was doubtful that 7% of the population of Monroe County traveled to New York City for health care, despite commuting patterns for work. Areas with such extreme commuting patterns to counties that were not contiguous were certainly the exception rather than the rule, but may be more common than believed, especially near major metropolitan areas with very high costs of living.

Adjustments for patient flow were similar to those made to RN supply. This produced the same RN-to-population ratio as using the unadjusted RN numbers and unadjusted population together, but produced different raw estimates.

Using this methodology, we found that Albany County, New York was estimated to need 3,634 RNs to treat its own population and incoming patients from other areas, while 4,942 RNs were estimated to work there. This estimated a supply that exceeded demand by 36%, which was a more moderate oversupply estimate than most others using Albany County as a single county, and somewhat comparable to those using the county cluster.

Schoharie County, using this methodology, was estimated to need 191 RNs, and had 197 (a shortage of 3%). This also appeared to be a moderate number compared to other estimates.

Monroe County, Pennsylvania was found to require 922 RNs, and had an estimated 834 working (a shortage of 9.5%). It was not surprising that this was lower than the other shortage estimates based on population ratios (25% and 23%), but it *was* surprising that it was so close to estimates based on actual health care use (11%).

New York County was found to need 34,126 RNs while there were an estimated 22,711 working there. This shortage (33%) was also very close to that based on actual health care use (29%).

Except for Albany County, adjustments of both population and RN supply based on commuting patterns to produce a ratio seemed to offer close approximations of estimates based on actual health care use in each of the test counties, including New York County and Monroe County, Pennsylvania, both of which experienced unusual levels of commuter flow.

E. Models Based on Sub-County Analyses

As the work on the county-level analyses described above progressed, concerns arose that counties were too large to study and understand the nursing needs of communities in the largest metropolitan areas, where very disadvantaged communities may exist in close proximity to very advantaged communities. Disadvantaged communities in urban areas may have a more difficult time recruiting RNs for two reasons:

1) RNs may be reluctant to work in communities where there is a perceived fear of crime or a large population with which they do not feel culturally competent, and
2) 2) a large percentage of the services offered in disadvantaged urban communities are provided by publicly operated facilities, which may not be able to offer salaries and benefits competitive with nearby non-public facilities that tend to serve more advantaged communities.

For this reason, some sub-county analyses were performed at the Census tract level using New York County (Manhattan) as a test case. These analyses were largely exploratory in nature, to try to determine what data might be available and what methods might be appropriate for sub-county analyses in the largest metropolitan areas across the U.S.

Census tract-level analyses posed many challenges. Demographic and economic population data were available at the Census tract level, and some RN supply data was available from the 2000 Census as well. Utilization data, however, was not available, nor were data on commuting patterns between Census tracts. There was also a question of the accuracy of RN supply estimates from sample data at such a small level of analysis. Ultimately, utilization rates were imputed based on the demographic characteristics of the tract population, the utilization data for the county, and the distribution of the county population between Census tracts.

It became very clear, however, that RNs in the population were not an adequate measure of available supply at the Census tract level. Some of the poorer neighborhoods had relatively high numbers of RNs per capita, but there was no basis for estimating how many of them worked in the neighborhoods where they lived. Similarly, many wealthy neighborhoods had relatively few RNs per capita (who presumably would not be able to afford to live in the most expensive neighborhoods of Manhattan), but there was no basis for assuming that the residents of these neighborhoods necessarily had difficulty obtaining nursing care. Estimates of service use in the population were also deemed suspect because it was impossible to estimate how many residents obtained health care within their own Census tract.

Subsequent reflection on the nature of labor markets and discussions with providers in New York County led study staff to believe that RN supply was not necessarily a correlate of difficulty recruiting at the local level. In large metropolitan areas, the pool of available labor tended to be geographically very broad, as illustrated by the fact that 70% of workers employed in New York County did not reside in New York County and that 16% of the employed residents of New York County did not work in New York County. It would thus appear that the supply of RNs within the Census tract where a health care facility was located was of limited relevance to the overall supply of RNs from which that facility may draw. Factors such as the ability to offer competitive compensation packages and the perceived environment of the neighborhood were likely to be much greater predictors of difficulty recruiting RNs in large metropolitan counties.

It may be best to establish guidelines specific to facilities in the largest metropolitan counties that would address the specialized problems of high-needs facilities. Possibilities include giving automatic eligibility to facilities in HPSA-designated areas or to those meeting certain criteria (e.g., public facilities), regardless of the eligibility of the overall county. To implement such a policy, it would be necessary to define a threshold for counties in which these automatic qualifications would apply (perhaps counties with populations of more than 1 million).

F. Factor Analysis of Nursing Shortage Indicators

The purpose of the factor analysis was to construct a smaller number of underlying common factors that could explain a large number of observed variables. This analysis was performed primarily mainly due to the lack of a single independent variable that could be used to measure nursing shortage that was comparable for all U.S. counties. The data used in this analysis came from the ARF 2005 release. In this analysis we chose three factors to describe the characteristics of counties in the U.S. based on 20 observed variables. The three factors explained 50.3% of total variation of the observed variables.

The list of the observed variables and the corresponding standardized scoring coefficients for each factor are presented in Table 36. Shadowed numbers were the highest coefficient for the corresponding variable, which revealed what variables were the primary bases for each factor. Note that 21.54% of U.S. counties were excluded from the analysis, mainly due to missing values or no hospital in those counties. Also, counties without hospitals were excluded from the analysis because a hospital was an important factor in analyzing nursing shortages. Most RNs were employed in hospital settings, which implied that hospitals drive the market for RNs. The counties without a hospital could be analyzed separately, but this had not been done at the time of this writing.

Table 36. Standardized Scoring Coefficients

Variable	Factor 1	Factor 2	Factor 3
Metropolitan dummy	-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

The standardized scoring coefficients suggested that Factor 1 consisted of high positive loadings on RNs per hospital bed, RNs per MD, RNs per 1,000 inpatient days, RNs per 1,000 outpatient visits, RNs per 1,000 emergency room visits, RNs per 100 Medicare inpatient days, and RNs per 100 Medicaid inpatient days. These loadings indicated that Factor 1 represented the ratio of RNs to health care utilization, especially in hospitals. A county with a high value for Factor 1 indicated that the county had a high number of RNs relative to health care utilization compared to other counties. On the other hand, a county with a low value for Factor 1 indicated that the county faced a nursing shortage problem, especially a shortage related to health care utilization in hospitals. Note that a county might score high on Factor 1 just because the county has low health care utilization due to underdeveloped health care infrastructure. Conversely, a county might score low on Factor 1 just because the county has high number of health care facilities which attracts many people from other counties for health care services. So, one must be cautious when interpreting Factor 1, and in particular, it should be interpreted in the context of the other two factors.

Factor 2 consisted of high positive loadings on RNs per 1,000 population, RNs per 1,000 individuals younger than 5 years, RNs per 1,000 individuals age 65 years or older, RNs per 1,000 civilian labor forces, and the percentage of health service workers; and a high negative loading on the percentage of manufacturing workers. These patterns suggested that Factor 2 represented the ratio of RNs to age-adjusted population. In addition, this factor also represented the supply of RNs. The lower the percentage of the manufacturing workers in a county, the more likely an individual was to enter the health care industry, including nursing profession. A county with high value for Factor 2 would generally have more RNs per capita than other counties. This factor was clearer in describing the nursing shortage than was Factor 1.

Factor 3 consisted of high positive loadings on the metropolitan dummy variable, RNs per individuals age 65 years and older, median household income (x $10,000); and high negative loadings on RNs per MD, infant mortality rate, unemployment rate, the percentage of individuals in poverty, the percentage of Black and Hispanic populations, and the percentage of American Indian and Alaska Native population. These patterns suggested that Factor 3 represented the economic condition of a county, including the percentage of minority populations. The percentage of minority population and quality of health were highly correlated with economic condition. A county with a high value for Factor 3 indicated that the county was in a metropolitan area with good economic conditions and lower percentage of minority populations compared to other counties.

The three factors above can be combined to describe a nursing shortage condition of each county in the U.S. To illustrate how this might work, suppose we divide each of the factors into two categories based on its median: lower than median and higher than median. (The threshold is arbitrarily chosen and could be replaced with other values, e.g., using the first quartile or other statistics.) Based on the three factors, each divided into two categories, all counties in the U.S. can be grouped into eight categories. Note that it was very common that nursing shortage was measured using the ratio of RNs to population (or age-adjusted population). As described before, Factor 2 represented the ratio of RNs to population. Based on this common criterion, Factor 2 was considered to be the most obvious factor in characterizing nursing shortage condition. So the categories were constructed based on the combinations of Factor 2, Factor 1, and Factor 3 which resulted in 8 categories: “111,” “112,” “121,” “122,” “211,” “212,” “221,” and “222”. The interpretations of these categories are described as follows.

- **Category 111.** Counties in this category had low values of the three factors. Intuitively, they were counties with a low number of RNs relative to population, low number of RNs relative to health care utilizations, and low economic conditions including a high proportion of minority populations and low quality of health. In general, counties in this category were counties with a nursing shortage problem and low economic conditions, so they needed to be supported by the government to increase the number of RNs in those counties.
- **Category 112.** Counties in this category had a low number of RNs relative to population, low number of RNs relative to health care utilization, and good economic conditions. Also, they were counties with a rich population. In addition, the health care industry in these counties was less attractive compared to other industries, suggesting not many people in these counties were interested in entering the nursing profession.
- **Category 121.** Counties in this category had a low number of RNs relative to population, high number of RNs relative to health care utilization, and low economic conditions. The high number RNs relative to health care utilization may have been due to the small number of health care infrastructures (e.g., one hospital). People from these counties may have gone to other counties for health care services because the amount of health care utilization in those counties was low. In subsequent, the ratio of RNs to health care utilization was high. Therefore, the high value of Factor 2 was not necessarily because of a high number of RNs but probably because of the limited health care infrastructure.
- **Category 122.** Counties in this category had a low number of RNs relative to population, high number of RNs relative to health care utilization, and good economic conditions. The high number of RNs relative to health care utilization may have been due to the low number of health care infrastructures, therefore people in these counties went to other counties for health care services. These counties were similar to those in category 121, except for the economic conditions.
- **Category 211.** Counties in this category had a high number of RNs relative to population, low number RNs relative to health care utilization, and low economic conditions. One possible reason for the low number of RNs relative to health care utilization was a highly developed health care infrastructure, therefore people from other counties came to these counties for health care services.
- **Category 212.** Counties in this category had a high number of RNs relative to population, low number of RNs relative to health care utilization, and good economic conditions. Counties in this category were similar to counties in category 211, except for the economic condition. They may not have had a nursing shortage problem because people from other counties came to these counties for health care utilization which suggested a low ratio of RNs to health care utilization. In addition, the counties in this category did not have economic problems.
- **Category 221.** Counties in this category had a high number of RNs relative to population, high number of RNs relative to health care utilization, and low economic conditions. These counties did not have nursing shortage problems, but had economic problems which included a high proportion of minority populations and a low quality of health.

- **Category 222.** Counties in this category had a high number of RNs relative to population, high number of RNs relative to health care utilization, and good economic conditions. These counties did not have nursing shortage problems and were without economic problems.

Now let us look at the distribution of counties by the categories for each Census division region as presented in Table 37. Among the nine regions, West South Central had the highest percentage of counties in category "111," which was 22% of the counties in the region. The second highest was Mountain (18%), followed by South Atlantic (11%), and East South Central (10%). On the other hand, New England had the highest percentage of counties in category "222," which was 39% of counties in the region. Those counties did not have nursing shortage problem and had good economic conditions. The second highest was Middle Atlantic (19%), followed by East North Central (17%), and West North Central (13%).

Table 37. Distribution of Counties by Categories for each Census Division

Census Division	Category[a]									Total
	Missing	111	112	121	122	211	212	221	222	
East North Central	70 16.0%	14 3.2%	40 9.2%	23 5.3%	50 11.4%	20 4.6%	111 25.4%	34 7.8%	75 17.2%	437 100%
East South Central	80 22.0%	38 10.4%	22 6.0%	32 8.8%	20 5.5%	71 19.5%	35 9.6%	46 12.6%	20 5.5%	364 100%
Middle Atlantic	13 8.7%	3 2.0%	13 8.7%	26 17.3%	48 32.0%	2 1.3%	8 5.3%	9 6.0%	28 18.7%	150 100%
Mountain	67 23.9%	49 17.5%	55 19.6%	29 10.4%	18 6.4%	13 4.6%	22 7.9%	22 7.9%	5 1.8%	280 100%
New England	4 6.0%	1 1.5%	2 3.0%	2 3.0%	25 37.3%	1 1.5%	2 3.0%	4 6.0%	26 38.8%	67 100%
Pacific	25 15.2%	20 12.2%	30 18.3%	10 6.1%	12 7.3%	18 11.0%	29 17.7%	15 9.2%	5 3.0%	164 100%
South Atlantic	167 28.4%	66 11.2%	58 9.8%	55 9.3%	41 7.0%	66 11.2%	56 9.5%	44 7.5%	36 6.1%	589 100%
West North Central	147 23.8%	34 5.5%	50 8.1%	62 10.0%	89 14.4%	21 3.4%	35 5.7%	97 15.7%	83 13.4%	618 100%
West South Central	103 22.0%	102 21.8%	22 4.7%	52 11.1%	18 3.8%	72 15.4%	30 6.4%	58 12.4%	12 2.6%	469 100%
Total	**676** **21.5%**	**327** **10.4%**	**292** **9.3%**	**291** **9.3%**	**321** **10.2%**	**284** **9.0%**	**328** **10.4%**	**329** **10.5%**	**290** **9.2%**	**3138** **100%**

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

Table 38 presents the distribution of counties by the categories for each rural/urban code. More than 50% of counties in the completely rural areas (Codes 8 and 9) had missing values or did not have a hospital. Apart from the two areas (8 and 9), the higher the codes (more rural the county), the higher was the percentage of counties in category "111." Less than 5% of counties in metro areas were categorized as "111." In contrast, more than 14% of counties in the non-metro areas were categorized as "111." On the other hand, the percentage of counties categorized as "222" was lower as the code increased (more rural the county). Twenty-two percent of counties of metro areas of 1 million population or more (Code=1) were categorized as "222." In contrast, only 2.5% of counties of completely rural areas were categorized as "222."

Table 38. Distribution of Counties by Categories for each Rural/Urban Code

Rural/Urban Codes[b]	Category									Total
	Missing	111	112	121	122	211	212	221	222	
1	72 17.4%	17 4.1%	72 17.4%	3 0.7%	47 11.4%	8 1.9%	100 24.2%	2 0.5%	92 22.3%	413 100%
2	63 19.4%	10 3.1%	37 11.4%	11 3.4%	72 22.2%	14 4.3%	52 16.0%	10 3.1%	56 17.2%	325 100%
3	82 23.4%	13 3.7%	46 13.1%	33 9.4%	85 24.2%	14 4.0%	30 8.6%	13 3.7%	35 10.0%	351 100%
4	3 1.4%	32 14.7%	18 8.3%	38 17.4%	15 6.9%	28 12.8%	41 18.8%	23 10.6%	20 9.2%	218 100%
5	1 1.0%	17 16.2%	12 11.4%	26 24.8%	18 17.1%	8 7.6%	10 9.5%	7 6.7%	6 5.7%	105 100%
6	73 12.0%	101 16.6%	30 4.9%	56 9.2%	20 3.3%	123 20.2%	68 11.2%	94 15.5%	43 7.1%	608 100%
7	39 8.7%	84 18.7%	52 11.6%	71 15.8%	36 8.0%	48 10.7%	21 4.7%	78 17.3%	21 4.7%	450 100%
8	124 52.8%	18 7.7%	7 3.0%	13 5.5%	5 2.1%	23 9.8%	4 1.7%	35 14.9%	6 2.6%	235 100%
9	219 50.6%	35 8.1%	18 4.2%	40 9.2%	23 5.3%	18 4.2%	2 0.5%	67 15.5%	11 2.5%	433 100%
Total	**676 21.5%**	**327 10.4%**	**292 9.3%**	**291 9.3%**	**321 10.2%**	**284 9.0%**	**328 10.5%**	**329 10.5%**	**290 9.2%**	**3138 100%**

Notes: [b] 1. Counties of metro areas of 1 million population or more

2. Counties in metro areas of 250,000 - 1,000,000 population
3. Counties in metro areas of fewer than 250,000 population
4. Urban population of 20,000 or more, adjacent to a metro area
5. Urban population of 20,000 or more, not adjacent to a metro area
6. Urban population of 2,500-19,999, adjacent to a metro area
7. Urban population of 2,500-19,999, not adjacent to a metro area
8. Completely rural or less than 2,500 urban population, adjacent to a metro area
9. Completely rural or less than 2,500 urban population, not adjacent to a metro area

Table 39 presents the distribution of counties by the categories for each HPSA designation code. Almost 50% of whole-HPSA counties had missing values or did not have a hospital. The percentage of counties categorized as “111” was almost equal in non-HPSA counties and whole-HPSA counties, at 9% each. On the other hand, 14% of non-HPSA counties were categorized as “222,” in contrast to 3% of whole-HPSA counties, and 10% of partial-HPSA counties.

Table 39. Distribution of Counties by Categories for each HPSA Code (Primary Care)

HPSA	Category									Total
	Missing	111	112	121	122	211	212	221	222	
None	105	74	78	84	93	44	129	85	110	802
	13.1%	9.2%	9.7%	10.5%	11.6%	5.5%	16.1%	10.6%	13.7%	100%
Whole County	392	74	31	43	4	101	37	100	23	805
	48.7%	9.2%	3.8%	5.3%	0.5%	12.6%	4.6%	12.4%	2.9%	100%
Part County	179	179	183	164	224	139	162	144	157	1531
	11.7%	11.7%	12.0%	10.7%	14.6%	9.1%	10.6%	9.4%	10.2%	100%
Total	**676**	**327**	**292**	**291**	**321**	**284**	**328**	**329**	**290**	**3138**
	21.5%	**10.4%**	**9.3%**	**9.3%**	**10.2%**	**9.0%**	**10.4%**	**10.5%**	**9.2%**	**100%**

IV. Preferred Method

Assumption: Current staffing patterns at the national level reflect a balance of supply and demand.

Assumption: Differences within types of care in factors such as patient acuity do not vary substantially across counties.

Assumption: RN commuting patterns are similar to the commuting patterns of other workers in terms of county inflow and outflow.

This method reflects an effort to create a more "realistic" model with which to assess the extent of nursing shortages in counties across the U.S. It incorporates elements of several of the models described above.

A. Estimating Health Care Utilization

Demand for RNs was estimated for 14 different settings. In eight of these settings (short-term inpatient, long-term inpatient, hospital outpatient, emergency department, psychiatric inpatient, hospital nursing home unit, other nursing home, and home health), demand was estimated based on actual or estimated use of services at the county level. In the other six settings (nurse education, public/community health, school health, occupational health, ambulatory care, and all other settings), estimates of demand were based on the size of the population.

<u>Short-term inpatient days (non-psychiatric hospitals)</u>: Data on inpatient days in short-term general hospitals by county was available from the ARF.

The ARF does not separate inpatient days in short-term non-general and long-term hospitals, but does separate several specific types of short-term non-general and long-term hospitals: short-term psychiatric, short-term rehabilitation, short-term children's psychiatric, long-term general medical and surgical, long-term psychiatric, long-term rehabilitation, and long-term children's psychiatric. This allowed the division of many of the most common hospital types into short-term non-general and long-term. Short-term non-general and long-term inpatient days that fell outside of these categories were categorized as short-term non-general and long-term based upon whether hospitals in the county that did not fall into any of the specific categories were short-term non-general or long-term hospitals.

Only 34 counties had both short-term non-general and long-term hospitals that fell outside the seven categories above, so in most cases it was easy to determine whether the remaining inpatient days were either short-term non-general or long-term. In the remaining 34 counties, the unidentified inpatient days were assigned as either short-term non-general or long-term based upon the proportion of hospital beds in the county falling into either of those categories.

This produced reasonable estimates of short-term non-general inpatient days by county, but these estimates included inpatient days spent in nursing home units. ARF provided estimates of nursing home unit inpatient days for short-term non-general and long-term hospitals, but did not separate the two. Once again, however, nursing home unit beds were separated into short-term non-general versus long-term, and this proportion was used to assign nursing home unit inpatient days to the two categories of hospitals. Inpatient days spent in nursing home units in short-term non-general hospitals were subtracted from the total number of inpatient days in short-term non-general hospitals, and are dealt with separately. The same was done for short-term general hospitals.

Days in short-term psychiatric and children's psychiatric hospitals were also then subtracted from total short-term non-general hospital inpatient days. These will also be treated separately. In a few cases, it was apparent that nursing home unit days in short-term non-general hospitals were being reported by short-term psychiatric hospitals (e.g., because the only short-term non-general hospital in the county was psychiatric, leaving this as the only explanation[3]), and in those cases subtracting both nursing home inpatient days and psychiatric inpatient days would have resulted in double subtraction and negative values. This was handled by subtracting only short-term psychiatric inpatient days in the counties where this occurred.

Long-term inpatient days (non-psychiatric hospitals): As described above, inpatient days in long-term hospitals were separated from those in short-term non-general hospitals using information about inpatient days in specific types of long-term hospitals and information about other hospitals in the county. Once again, however, these estimates contained nursing home unit days, which were subtracted as described for short-term non-general inpatient days. Days in long-term psychiatric and children's psychiatric hospitals were also then subtracted from total long-term hospital inpatient days.

As with short-term non-general inpatient days, it was evident that a small number of long-term psychiatric hospitals had reported nursing home unit days[4]. Because subtracting both nursing home inpatient days and psychiatric inpatient days would result in double subtraction and negative values, this was handled by only subtracting long-term psychiatric inpatient days in the counties where this occurred.

Psychiatric hospital inpatient days: Because inpatient days in both short-term and long-term psychiatric and children's psychiatric hospitals were separated out in the ARF for all counties, psychiatric hospital inpatient days were not difficult to count. The only complexity was that 18 of these hospitals, as discussed above, appeared to report nursing home unit days. Because this seemed improbable, the decision was made to ignore the nursing home days rather than subtracting them from the totals for psychiatric inpatient days and adding them to the total for hospital nursing home unit days[5].

Nursing home unit inpatient days: Nursing home unit inpatient days were presented in ARF for both short-term general and short-term non-general and long-term hospitals. The data was clear except for the issue discussed above of a small number of psychiatric hospitals (both short-term and long-term) apparently reporting nursing home unit days. These nursing home days were removed from the nursing home days total for short-term non-general and long-term hospitals.

[3] It was not clear why psychiatric hospitals would report nursing home unit days. As this occurred in only 10 of the 3,140 counties it was possible that this was simply due to a reporting error.
[4] This occurred in eight of the 3,140 counties, and may be due to reporting errors.
[5] The potential result of this, if these reports were not in error, could be to overestimate demand for RNs in psychiatric hospitals if some of these hospitals did indeed contain nursing home units. RN staffing in psychiatric hospitals is typically more intensive than in nursing home units, so misclassifying nursing home days as psychiatric days could inflate the demand figures for RNs.

Example: Tuscaloosa, Alabama

Short-term general days were 187,432. Short-term non-general and long-term days were 393,627. The county had two long-term hospitals (both psychiatric) and no short-term non-general hospitals, and all of the 393,627 days were in long-term rather than short-term non-general hospitals. In total, 61,861 nursing home unit days in long-term hospitals were reported for this county, and by definition had to be reported for one of the long-term psychiatric hospitals. These days were treated as long-term psychiatric days rather than nursing home days. The total number of short-term days for Tuscaloosa County was 187,432, and the total number of long-term non-psychiatric days was 0. The number of psychiatric inpatient days was 393,627, and the number of nursing home inpatient days was counted as 0.

Example: Pima County, Arizona

Short-term general days were 555,167. Short-term non-general and long-term days were 75,844. The county had four short-term non-general hospitals (totaling 251 beds) and one-long-term hospital (totaling 51 beds). Inpatient days in short-term non-general and long-term hospitals (75,844) were apportioned according to the ratio of beds (approximately 83% short-term general and 17% long-term), to produce 63,036 short-term non-general days and 12,808 long-term days.

Outpatient visits (non-emergency): Outpatient visits to hospital non-emergency departments for short-term hospitals and short-term non-general and long-term hospitals were available in ARF. The sum of these two figures was used to produce the figure for total non-emergency hospital outpatient visits.

Emergency department visits: Visits to hospital emergency departments for short-term hospitals and short-term non-general and long-term hospitals were available in ARF. The sum of these two figures was used to produce the figure for total hospital emergency department visits.

Non-hospital nursing home population: The 2000 U.S. Census has data by county for those living in specific types of group quarters, including nursing homes.

Home health patients: 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 taken from a CDC report available online at: www.cdc.gov/nchs/data/nhhcsd/curhomecare00.pdf.

Example: Albany County, New York

Table 40, below, illustrates how national age- and gender-specific rates were applied to the population of Albany County to obtain estimates of 528 male and 1,155 female home health patients in the county (total home health patients = 1,683).

Table 40. Illustrative Application of Age- and Gender-Adjusted Utilization Rates Are Applied for a County

Age Group	Male	Patients per 10,000 Pop	Male Home Health Patients	Female	Patients per 10,000 Pop	Female Home Health Patients
0-17	34,074	9.1	31	32,004	8.6	28
18-44	58,186	9.3	54	60,733	13.2	80
45-64	32,013	35.6	114	34,655	33.9	117
65-69	4,571	98	45	5,687	107.5	61
70-74	4,512	103.8	47	6,013	203.7	122
75-79	3,576	216.8	78	5,550	377.9	210
80-84	2,324	358.5	83	4,376	432.3	189
85+	1,387	553.9	77	4,596	754.9	347
Total	140,643	35.1	528	153,614	61.8	1,155

Other nursing care: The use of other types of nursing care (nurse education, public and community health, school health, occupational health, non-hospital ambulatory care, and other) was estimated based upon population ratios as described below.

B. Estimating Current National RN Staffing

Using data from the NSSRN, it was possible to estimate RN staffing by setting at the national level.

Short-term inpatient (non-psychiatric hospitals): The RNs included as employed in this category of care were all those working in hospital units other than emergency department, outpatient, home health, radiologic, or dialysis in non-federal, non-psychiatric short-term hospitals, federal government hospitals[6], and other types of hospitals.

Long-term inpatient (non-psychiatric hospitals): The RNs included as employed in this category of care were all those working in hospital units other than emergency department, outpatient, home health, radiologic, or dialysis in non-federal non-psychiatric long-term hospitals.

Psychiatric inpatient (non-federal): The RNs included as employed in this category of care were all those working in hospital units other than emergency department, outpatient, home health, radiologic, or dialysis in non-federal psychiatric hospitals.

[6] Federal government hospitals include some long-term and psychiatric hospitals, but these were not distinguished in the NSSRN. Most federal hospitals are VAs, which tend to provide short-term general care, and so RNs in federal government hospitals will be included in this category rather than another.

Nursing home unit inpatient: The RNs included as employed in this category of care were all those who reported working in a "nursing home unit in hospital."

Outpatient (non-emergency): The RNs included as employed in this category of care were all those who reported working in outpatient, radiologic, or dialysis units in any type of hospital.

Emergency outpatient: The RNs included as employed in this category of care were all those who reported working in emergency departments in any type of hospital.

Non-hospital nursing home: The RNs included as employed in this category of care were all those who reported working in a nursing home other than a hospital nursing home unit.

Home health: The RNs included as employed in this category of care were all those who reported working in a home health unit in a hospital of any type or any type of home health agency.

Nurse education: The RNs included as employed in this category of care were all those who reported working in any type of nursing care education program, including LPN and CNA programs.

Public and community health: The RNs included as employed in this category of care were all those who worked in state or local health departments, community mental health and substance abuse facilities, any kind of community health clinic (CHC, family planning clinic, RHC), or a day care, hospice, or other community health setting.

School health: The RNs included as employed in this category of care were all those who worked in public or private school health services, elementary through high school. Those working in college or university health services were not included.

Occupational health: The RNs included as employed in this category of care were all those who worked in private, government, or other occupational health services.

Non-hospital ambulatory care: The RNs included as employed in this category of care were all those who worked in physician or nurse practices, clinics, HMOs, or other non-hospital ambulatory settings.

Other nursing care: The RNs included as employed in this category of care were all those who worked in any setting not included in the above, including facilities for the mentally retarded, college health services, insurance companies, state boards of nursing, and professional associations.

C. Estimating RN Demand by Setting

The national estimates for utilization and current RN staffing were combined to produce ratios of RNs to units of care, as shown in Table 41 below. These ratios were then applied to utilization and population counts at the county level to estimate how many RNs would be needed to achieve these ratios.

Table 41. RNs per Unit of Care in Fourteen Health Care Settings in Selected Years

RNs by Setting		Units of Care by Setting (Year of Estimate)		Ratio of RNs to Units of Care
All inpatient Units in Short-Term Hospitals (2004)	861,113	ST Inpatient Days (2003)	173,161,615	4.97 RNs per 1,000 Inpatient Days
All inpatient Units in Long-Term Hospitals (2004)	84,662	LT Inpatient Days (2003)	7,261,248	11.66 RNs per 1,000 Inpatient Days
All inpatient Units in Psychiatric Hospitals (2004)	36,651	Psychiatric Inpatient Days (2003)	25,313,077	1.45 RNs per 1,000 Inpatient Days
Nursing Home Unit in Hospital (2004)	12,090	Total Nursing Home Hospital Inpatient Days (2003)	25,374,490	0.48 RNs per 1,000 Inpatient Days
Other Nursing Home (2000)	118,898	Nursing Home Resident Population (2000)	1,720,500	0.07 RNs per NH Resident
Nurse Education Programs (2000)	46,301	Estimated Active RNs (2000)	2,233,864	0.02 per Active RN
Public/Community Health RNs (2000)	148,507	Total Population (2000)	281,421,906	5.28 RNs per 10,000 Pop
School Health (excl. college)	66,587	Population Age 5-17	53,089,688	12.5 RNs per 10,000 Pop
Occupational Health (2000)	36,099	Population Age 18-64	174,294,950	2.07 RN per Pop
Home Health (2000)	131497	Estimated Home Health Patients (2000)	1,365,940	0.10 RNs per Patient
Outpatient or Diagnostic Units in All Hosp (2000)	85,433	Outpatient Visits - Other Than ED (All hospitals) (2000)	600,155,715	0.14 RNs per 1,000 Visits
EDs in All Hospitals (2000)	91,732	ED Visits (All Hospitals) (2000)	107,293,419	0.86 RNs per 1,000 Visits
Ambulatory Care (2000)	209,165	Total Population (2000)	281,421,906	7.43 per 10,000 Pop
Other	114,958	Total Population (2000)	281,421,906	4.1 RNs per 10,000 Pop

D. Estimating Supply of RNs

The only nationally available figures for RNs by county were from the 2000 Census, and were based on county of residence. For a substantial portion of the RN workforce, however, county of residence was different from county of employment. To adjust for this, county-to-county commuting flows were obtained from the U.S. Census Bureau, and RN estimates were adjusted based upon the ratio of workers living in the county to workers working in the county. This methodology assumed that the commuting patterns of RNs did not differ substantially from the commuting patterns of the civilian workforce overall.

Example: Albany County, New York

In 2000, 117,668 residents of Albany County worked in Albany County, and another 24,174 residents of Albany County worked outside of Albany County (a total of 141,842 residents of Albany County worked, with 17% commuting out). An additional 101,045 workers commuted into Albany County, resulting in a total workforce of 218,713 in Albany County [117,668 + 101,045], with 46% commuting into the county from other counties. The ratio of workers (both residents and non-residents) working in the county to residents of the county who worked (both within and outside the county) was 1.5419 [218,713/141,842]. This adjustment factor was applied to the number of RNs living in Albany County [3,205 x 1.5419] to estimate that 4,942 RNs actually worked in Albany County.

E. Estimating RN Shortages

The estimation of RN shortages was based upon the difference between estimated demand for RNs and the number of RNs in the county (adjusted for commuting patterns). Raw shortage numbers were then standardized as a percent of demand. This methodology did not assess shortages at the national level because it theoretically redistributed the current number of RNs into counties according to patterns of health care utilization. While a small national shortage occurred using our procedures, this may have been an artifact of using data from different years for different types of care (hospital ratios used 2004 nurse data and 2003 hospital data, while ratios for other types of care used 2000 nurse and hospital data).

At the state level, however, some interesting patterns emerge (Table 42, below). Half of the states were not seen to have shortages, and those with the largest relative supplies of RNs were Vermont, New Hampshire, and Alaska. On the other hand, the District of Columbia had a 49% shortage, while Louisiana had a 25% shortage, and Oklahoma had a 20% shortage.

Table 42. Estimated Percentage Shortages of RNs in the U.S.

FIPS State Code	Estimated RNs	Unadjusted Demand	Percent Shortage
District of Columbia	4,267.6	8,672.8	49%
Louisiana	11,210.7	44,913.0	25%
Oklahoma	5,765.5	29,281.2	20%
Nevada	2,732.2	14,182.6	19%
Mississippi	4,955.5	27,235.9	18%
New York	29,697.8	187,629.5	16%
Texas	25,686.1	163,456.0	16%
West Virginia	2,654.7	17,625.0	15%
Arkansas	3,280.4	23,831.6	14%
Hawaii	1,190.0	9,650.9	12%
California	28,761.9	233,938.4	12%
Rhode Island	1,232.7	10,761.8	11%
Virginia	5,797.4	57,588.5	10%
Georgia	6,027.4	63,405.5	10%
Florida	10,510.1	134,832.0	8%
Idaho	394.2	8,434.0	5%
New Jersey	2,570.5	70,834.1	4%
Kentucky	962.3	35,434.4	3%
Tennessee	1,292.1	49,246.3	3%
Alabama	878.9	37,830.2	2%
Arizona	366.6	34,685.2	1%
New Mexico	111.6	12,177.3	1%
Utah	92.8	13,787.5	1%
Missouri	77.7	50,013.8	0%
South Carolina	-127.4	32,188.9	0%
Montana	-76.6	7,054.3	-1%
North Carolina	-1,503.8	67,261.0	-2%
Pennsylvania	-2,632.2	116,156.5	-2%
North Dakota	-249.1	6,312.6	-4%
Colorado	-1,398.4	28,716.7	-5%
Maryland	-2,480.0	41,098.5	-6%
Indiana	-2,936.6	48,152.2	-6%
Wyoming	-212.2	3,370.4	-6%
Michigan	-5,070.8	73,520.9	-7%
Massachusetts	-5,061.8	63,465.1	-8%
Iowa	-2,506.4	26,343.7	-10%
Kansas	-2,293.4	22,984.5	-10%
South Dakota	-793.3	6,863.4	-12%
Nebraska	-1,920.6	14,639.6	-13%
Ohio	-12,283.5	90,622.4	-14%
Connecticut	-4,402.6	28,395.4	-16%
Oregon	-3,846.4	21,216.4	-18%
Maine	-2,077.4	9,736.7	-21%
Delaware	-1,400.2	6,488.7	-22%
Wisconsin	-8,450.0	38,179.5	-22%
Washington	-8,082.2	35,861.5	-23%
Illinois	-22,402.1	99,354.8	-23%
Minnesota	-9,245.4	38,000.9	-24%
Alaska	-1,156.6	3,805.3	-30%
New Hampshire	-3,224.8	8,929.8	-36%
Vermont	-1,653.6	4,052.7	-41%
Total	**43,029.4**	**2,282,219**	**2%**

Eighteen counties in the U.S. had a 100% shortage (all of these counties had no RNs), but a handful more counties had shortages of more than 90%.

When counties were aggregated into metropolitan and micropolitan areas (shown below), the MSA with the greatest shortage was the Boone, Iowa micropolitan area (80%). Relatively few major metropolitan areas had serious shortages -- the notable exceptions were Las Vegas (with a 25% shortage), New Orleans (22%)[7], and New York (also 22%). Oklahoma City, Los Angeles, Topeka, and Honolulu also had shortages (16%, 14%, 13%, and 12%, respectively). Despite the serious shortage estimated for the District of Columbia proper, the Washington-Arlington-Alexandria MSA (which included counties in Maryland, Virginia, and West Virginia, as well as D.C.) had a shortage of only 2%.

[7] This was using data from before Hurricane Katrina in 2005. New Orleans may currently have a much greater shortage.

V. Additional Analyses and Explorations

At the final meeting of the advisory panels, suggestions were offered for several additional analyses prior to finalizing the report. The results of these analyses are summarized below.

A. Adjustments for Patient Acuity

Perhaps the greatest shortcoming of the Preferred Method is that it does not account for patient acuity. This may systematically underestimate RN demand and need in counties with large medical centers with trauma units, which might be expected to have higher levels of patient acuity on average than small community hospitals.

Related to this, larger hospitals may have more patients coming in for complex surgeries, and may require larger surgical staffs (including OR RNs) than their smaller counterparts. While accounting for ICU and surgical patients is a rough estimate of patient acuity, it may be better than no adjustment at all.

Although inpatient days are not broken into type of stay, ARF does provide data on different types of beds in short-term general hospitals, including medical and surgical intensive care beds, cardiac intensive care beds, neonatal intensive care beds, neonatal intermediate care beds, pediatric intensive care beds, burn care beds, other special care beds, and other intensive care beds. This can be used to assign inpatient days proportionately to ICU and regular care.

One difficulty in accounting for ICU patients, however, was that not all counties in ARF had their short-term general hospitals beds correctly allocated into different types of beds. This was evident when the totals for various types of beds did not sum to the total number of inpatient beds. This occurred in 14% of counties. In some of these counties bed counts were sufficiently similar using total number of beds and the sum of specific bed types that the percent estimate of ICU beds did not differ greatly. Overall, 90% of counties had data that permitted a reliable estimate of percent of inpatient beds that were ICU beds. For the other 10%, data on specific bed types either was not provided or was sufficiently different from the total number of beds that estimates of ICU beds could not be reliably made[8]. For these counties, demand estimates using the first method were maintained with no acuity adjustment. An alternative strategy would be to use a regression-based imputation method for assigning values to the missing cases.

Another limitation was that bed types were not available for short-term non-general hospitals, which may also have ICUs and operating rooms. RNs, on the other hand, cannot be separated by general versus non-general short-term hospitals, so RNs in ICUs in both types of hospitals were factored into the staffing ratio for ICU, but inpatient days in short-term non-general hospitals were not adjusted down by parsing out the ICU days. The net effect of this was to overestimate how many RNs were being used per patient day in ICUs and underestimate how many were being used per patient day in other inpatient units. This bias favored counties with more ICU days. The adjustments for non-psychiatric hospitals are shown in Table 43.

[8] This was considered to be the case when percent ICU beds calculated using the sum of specific bed types was different by more than two percentage points from when it was calculated using total number of beds.

Table 43. Estimated RN Utilization Adjustment for ICU and Surgeries for Non-Psychiatric Hospitals

RNs working in Med/Surg, Step-Down Transitional, Recovery, or Labor and Delivery Unit in ST General Hospital or InPt Units (excluding OR) in "Other" Hospitals (2004)	568,955	Short-Term Inpatient Days (excluding estimated ICU) and STNG inpatient days (2003)	155,847,193	3.65 RNs per 1,000 InPt Days
RNs working in ICU Units (2004)	203,504	Short-Term Estimated ICU inpatient days (2003)	19,314,122	10.54 RNs per 1,000 InPt Days
RNs working in ORs in ST and LT Hospitals (2004)	101,379	Total number of hospital-based surgeries in Short-Term general and non-general/LT hospitals (2003)	28,009,403	3.62 RNs per 1,000 Surgeries
RNs working in Inpatient Units in LT Hospitals (2004)	71,937	Long-Term Inpatient Days	7,261,248	9.91 RNs/1,000 inpatient days

The net effect of this method was to reduce the estimated nursing shortage for many counties, but to increase it for a few. This was partly because this adjustment was expected to level the playing field for more urban counties with higher rates of acuity in their hospitals, but the counties that were most likely to lack reliable ICU data tended to be the largest urban counties. Of the 20 largest U.S. counties in terms of population, 10 lacked data to support the acuity adjustment. Of those that had adequate data, however, the shortage estimate using the adjusted measure still tended to be lower than that using the unadjusted measure. Because this adjustment had promise, it could be considered as the theoretical standard, even though currently available data did not permit its use in practice.

B. Commuting Patterns for RNs

The original version of the Preferred Method assumed that RN commuting patterns were similar to those of the overall workforce. This was generally true in the aggregate -- RNs were no more or less likely than other workers to work outside the county where they live. At the individual county level, however, RN commuting patterns sometimes varied dramatically from the patterns for all workers. It appeared that RN commuting patterns depended more on county characteristics than on characteristics of RNs (e.g., gender, income level, etc.).

1. Models to Predict Commuting Patterns

This finding led to efforts to formulate a model to predict RN commuting patterns based on county characteristics. 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 have reflected many of the primary drivers of commuting behavior (e.g., relative wages, cost of living, etc.). Another key independent variable was assumed to be the opportunities for RN employment available 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 1% increase in overall 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, such that 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.

Models were estimated separately for three types 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). The results presented in Table 44 show some potential for fine-tuning the results for different types of counties, but differences in the model coefficients were not dramatic. Differences in model fit were substantial, however. The best-fitting model was for counties not adjacent to a metropolitan area, with an adjusted R^2 of 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 44. Ordinary Least Squares Regression Coefficients Predicting RN Incommuting, By Type of County

	All Counties		Metro Counties		Counties Adjacent to Metro County		Counties Not Adjacent to Metro County	
	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 in-commuting	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	

* $p \le 0.05$ ** $p \le 0.01$ *** $p \le 0.001$

The series of tables below gives descriptive statistics for the difference between the various commuting estimates and actual RN commuting patterns for all 244 counties combined (Table 45) and broken out by relationship to a metropolitan area (Tables 46 through 48). The mean value allowed us to judge whether estimates were biased in a particular direction, but gave a limited sense of overall accuracy (e.g., estimates that were extremely inaccurate but equally likely to be overestimates or underestimates could produce a mean difference of 0 from actual patterns of commuting).

Table 45. Differences Between Selected Commuting Estimates and Actual Commuting Patterns, All County Categories Combined

	N	Minimum	Maximum	Mean	Std Dev
Difference between actual RN commuting and estimate based on all-counties regression	244	-0.76	1.48	0.00	0.207
Difference between actual RN commuting and estimate based on RUCC-specific regression	244	-0.60	1.25	0.00	0.193
Difference between actual RN commuting and commuting of all workers	244	-2.12	1.62	-0.098	0.288

Table 46. Differences Between Selected Commuting Estimates and Actual Commuting Patterns, Metropolitan Counties

	N	Minimum	Maximum	Mean	Std Dev
Difference between actual RN commuting and estimate based on all-counties regression	93	-0.76	0.53	-0.025	0.189
Difference between actual RN commuting and estimate based on RUCC-specific regression	93	-0.60	0.53	-0.00	0.179
Difference between actual RN commuting and commuting of all workers	93	-2.12	0.86	-0.067	0.305

Table 47. Differences Between Selected Commuting Estimates and Actual Commuting Patterns, Counties Adjacent to Metropolitan Areas

	N	Minimum	Maximum	Mean	Std Dev
Difference between actual RN commuting and estimate based on all-counties regression	86	-0.30	1.48	0.025	0.252
Difference between actual RN commuting and estimate based on RUCC-specific regression	86	-0.42	1.25	0.00	0.241
Difference between actual RN commuting and commuting of all workers	86	-0.61	1.62	-0.087	0.302

Table 48. Differences Between Selected Commuting Estimates and Actual Commuting Patterns, Counties Not Adjacent to Metropolitan Areas

	N	Minimum	Maximum	Mean	Std Dev
Difference between actual RN commuting and estimate based on all-counties regression	65	-0.34	0.63	0.004	0.160
Difference between actual RN commuting and estimate based on RUCC-specific regression	65	-0.26	0.39	0.00	0.131
Difference between actual RN commuting and commuting of all workers	65	-0.66	0.67	-0.154	0.234

The next pair of tables provides descriptive statistics for the absolute difference between the various commuting estimates and actual RN commuting patterns for all 244 counties combined (Table 49) and for metropolitan counties (Table 50 through 52). In contrast to the tables above, the mean absolute values below allow one to judge overall accuracy, but not whether the estimates were biased in a particular direction.

Table 49. Absolute Differences Between Selected Commuting Estimates and Actual Commuting Patterns, All County Categories Combined

	N	Minimum	Maximum	Mean	Std Dev
Absolute difference between actual RN commuting & estimate based on all-counties	244	0	1.48	0.144	0.149
Absolute difference between actual RN commuting & estimate based on RUCC analysis	244	0	1.25	0.143	0.129
Absolute difference between actual RN commuting and commuting of all workers	244	0	2.12	0.208	0.221

Table 50. Absolute Differences Between Selected Commuting Estimates and Actual Commuting Patterns, Metropolitan Counties

	N	Minimum	Maximum	Mean	Std Dev
Absolute difference between actual RN commuting and estimate based on all-counties	93	0	0.76	0.143	0.125
Absolute difference between actual RN commuting & estimate based on RUCC analysis	93	0	0.60	0.138	0.114
Absolute difference between actual RN commuting and commuting of all workers	93	0	2.12	0.181	0.253

Table 51. Absolute Differences Between Selected Commuting Estimates and Actual Commuting Patterns, Counties Adjacent to Metropolitan Counties

	N	Minimum	Maximum	Mean	Std Dev
Absolute difference between actual RN commuting and estimate based on all-counties	86	0	1.48	0.161	0.194
Absolute difference between actual RN commuting & estimate based on RUCC analysis	86	0	1.25	0.178	0.161
Absolute difference between actual RN commuting and commuting of all workers	86	0	1.62	0.229	0.214

Table 52. Absolute Differences Between Selected Commuting Estimates and Actual Commuting Patterns, Counties Not Adjacent to Metropolitan Counties

	N	Minimum	Maximum	Mean	Std Dev
Absolute difference between actual RN commuting and estimate based on all-counties	65	0	0.63	0.122	0.103
Absolute difference between actual RN commuting & estimate based on RUCC analysis	65	0	0.39	0.102	0.081
Absolute difference between actual RN commuting and commuting of all workers	65	0	0.67	0.218	0.176

Interestingly, the most accurate method for estimating commuting varied by county type, as shown in Table 53. In metro counties, the commuting flow of all workers was the most accurate estimate 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.

Table 53. Percentage of Observations in Which Each Estimate is Closer Than Others to the Actual Value

	Metro County		Adjacent to Metro County		Not Adjacent to Metro County	
	Count	%	Count	%	Count	%
All-counties regression	28	30.1%	40	46.5%	15	23.1%
RUCC-specific regression	29	31.2%	28	32.6%	33	50.8%
Commuting of all workers	36	38.7%	18	20.9%	17	26.2%

The "best" estimate, however, might be 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, we found that 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 (Table 54). It was never preferable to use the overall commuting pattern.

Table 54. Percentage of Cases in Which Estimated Commuting Differed From Actual by More Than 10%

Group (N)	Statistic	All-Counties Regression Estimate Off by > 10%	RUCC-Specific Regression Estimate Off by > 10%	Estimate Based 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 Estimates to Predict Actual RN Supply

The evaluation of the estimates using the counties from which they were derived was somewhat tautological, however. We also needed to assess whether these corrections brought our estimates of RN employment by county closer to actual employment data in other states (for which real commuting patterns were not available). These states were Iowa (some counties), Texas, Pennsylvania, South Dakota, and Tennessee.

When compared to actual counts of RNs working in particular counties, the revised commuting adjustments did very little to improve the estimated counts. The estimated supply was actually closer to the actual supply on average when overall commuting was used as the adjustment factor. This was true both in terms of the average difference (Table 55) and the average absolute difference (Table 56).

Table 55. Average Commuting Adjustment for RN Supply Across All County Categories (N = 812)

Group (N)	Minimum	Maximum	Mean	Std Dev
Difference between actual RN commuting and supply estimate based on all-counties commuting regression	-10,000	4,499	-116.7	754.7
Difference between actual RN commuting and supply estimate based on RUCC-specific commuting regression	-10,389	5,204	-128.3	840.0
Difference between actual RN commuting and supply estimate based on commuting of all workers	-3,635	2,153	-42.6	341.0

Table 56. Average Absolute Commuting Adjustment for RN Supply Across All County Categories (N = 812)

Group	Minimum	Maximum	Mean	Std Dev
Absolute difference between actual RN supply and supply estimate based on all-counties commuting regression	0.04	10,000	198.5	737.4
Absolute difference between actual RN supply and supply estimate based on RUCC-specific commuting regression	0.01	10,389	211.6	822.9
Absolute difference between actual RN supply and supply estimate based on commuting of all workers	0.01	3,635	122.8	321.0

Table 57 shows that there was some variation by 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.

Table 57. Differences Between Actual RN Supply and Predicted RN Supply Based on Various RN Commuting Estimates, by Metropolitan Status of County

Group (N)	Statistic	Difference between actual RN commuting and estimate based on all-counties regression	Difference between actual RN commuting and estimate based on RUCC-specific regression	Difference between actual RN commuting and commuting of all workers
Metro County (261)	Mean	-348.8	-376.4	-105.1
	Std Dev	1,298.5	1,448.0	588.2
Adjacent to Metro County (305)	Mean	-18.2	-20.6	-23.2
	Std Dev	81.6	91.3	80.9
Not Adjacent to Metro County (246)	Mean	7.5	1.6	-0.42
	Std Dev	54.7	57.0	58.0
Total (812)	Mean	-116.7	-128.3	-42.6
	Std Dev	754.7	840.0	341.0

Table 58. Absolute Differences Between Actual RN Supply and Predicted RN Supply Based on Various RN Commuting Estimates, by Metropolitan Status of County

Group (N)	Statistic	Absolute difference between actual RN commuting and estimate based on all-counties regression	Absolute difference between actual RN commuting and estimate based on RUCC-specific regression	Absolute difference between actual RN commuting and commuting of all workers
Metro County (261)	Mean	540.6	574.7	296.5
	Std Dev	1,230.8	1381.1	518.4
Adjacent to Metro County (305)	Mean	45.9	50.8	49.9
	Std Dev	69.8	78.6	67.8
Not Adjacent to Metro County (246)	Mean	24.8	25.7	29.0
	Std Dev	49.3	50.9	50.2
Total (812)	Mean	198.5	211.6	122.8
	Std Dev	737.4	822.9	321.0

Table 59 shows that generally the estimates based on overall commuting were the closest to actual counts 39% of the time, those based on the all-county commuting model were closest 33% of the time, and those based on the RUCC-specific models were closest 28% of the time.

Table 59. Percentage of Observations in Which Each Supply Estimate was Closer Than Others to the Actual Supply Value, All Metropolitan Groups Combined

Group	Count	%
All County Model	267	33%
RUCC-Specific Model	226	28%
Overall Commuting Model	315	39%

Once again this varied by RUCC category, as shown in Table 60 below.

Table 60. Percentage of Observations in Which Each Supply Estimate was Closer Than Others to the Actual Supply Value, by Metropolitan Groups

Group	Metropolitan Counties		Counties Adjacent to Metropolitan Area		Counties Not Adjacent to Metropolitan Area	
	Count	%	Count	%	Count	%
Supply Estimate Based On All County Commuting Model	64	24.5%	124	40.8%	79	32.5%
Supply Estimate Based on RUCC-Specific Commuting Model	63	24.1%	86	28.3%	77	31.7%
Supply Estimate Based on Overall Worker Commuting	134	51.3%	94	30.9%	87	35.8%

As noted before, however, the “best” estimate might be better than the “next best” estimate by only a point or two (Table 61). 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, there was little difference in the likelihood that an estimate was off by this threshold. Table 62 shows that this more or less holds true for every RUCC category, although a very slight advantage might be gained by using the all-county model to estimate commuting in non-metro counties.

Table 61. Percentage of Cases in Which Supply Estimate Differed From Actual Supply by More Than 10% (N = 812)

Group	Minimum	Maximum	Mean	Std Dev
Supply Estimate Based On All County Commuting Model Is Off By > 10%	0	1	0.784	0.411
Supply Estimate Based on RUCC-Specific Commuting Model Is Off By > 10%	0	1	0.799	0.401
Supply Estimate Based on Overall Worker Commuting Is Off By > 10%	0	1	0.800	0.400

Table 62. Percentage of Cases in Which Estimated Commuting Differed From Actual by More Than 10%

Group (N)	Statistic	All-Counties Regression Estimate Off by > 10%	RUCC-Specific Regression Estimate Off by > 10%	Estimate Based On Commuting of All Workers Off by > 10%
Metro County (261)	Mean	83.1%	84.7%	75.9%
	Std Dev	37.5%	36.1%	42.9%
Adjacent to Metro County (305)	Mean	77.0%	79.3%	82.3%
	Std Dev	42.1%	40.6%	38.2%
Not Adjacent to Metro County (246)	Mean	75.2%	75.6%	81.7%
	Std Dev	43.3%	43.0%	38.7%
Total (812)	**Mean**	**78.4**%	**80.0**%	**80.0**%
	Std Dev	**41.1**%	**40.1**%	**40.0**%

It is important to remember that the accuracy of the commuting estimates was only one source of inaccuracy in estimated supply of RNs working in a county. Another source of inaccuracy was related to the estimated supply of RNs living in a county. It may be possible for future work to estimate confidence intervals around commuting estimates and estimates of supply of resident RNs, and for designation to be based on the lowest estimate of RNs in these confidence intervals. This would remove some of the disadvantage potentially faced by rural areas due to shortcomings of sampling, although this would increase the likelihood of designating some counties as shortage areas that should not, in fact, qualify.

C. Other Analyses

The meeting of the advisory panels concluded with a description of plans for the remainder of the study. The results of these analyses are summarized below, along with the conclusions drawn.

1. Shortage Counties and Persistent Poverty Counties

The tables below show the average values of the variables used in determining nursing shortage status. Table 63 shows that on average those persistent poverty counties that did not receive a shortage designation had smaller populations, but more RNs than persistent poverty counties that were defined as having shortages. Table 64 demonstrates that the persistent poverty counties defined as shortage counties had higher ratios of both RNs to population and RNs to age-adjusted population than counties not defined as shortage counties. Finally, Table 65 shows that, despite having higher numbers of RNs per capita, persistent poverty counties not defined as shortage counties had lower average utilization of hospital-based services than persistent poverty counties that were shortage counties. It also shows that hospital-based services were the primary driver of RN demand (3,021 general inpatient days per 10,000 population versus 5,546 days/10,000 population; 13,681 outpatient visits per 10,000 population versus 20,188/10,000; and 3,785 emergency department visits per 10,000 population versus 5,808/10,000).

Table 63. Population and RNs in Selected Classes of Counties in the U.S.

County Characteristics	Total Population	Number of RNs Living in County	RN Supply Adjusted for Commuting
Both shortage and persistent poverty	31,848	131	139
Shortage but not persistent poverty	47,037	279	282
Persistent poverty but not shortage	28,261	181	176
Neither shortage nor persistent poverty	103,357	840	842
Total	89,596	711	713

Table 64. Measures of RN Supply for Selected Classes of Counties in the U.S.

County Characteristics	RNs / 10,000 Population	RN / 10,000 Adjusted Pop
Both shortage and persistent poverty	35.2	30.7
Shortage but not persistent poverty	46.3	42.1
Persistent poverty but not shortage	59.2	51.5
Neither shortage nor persistent poverty	77.5	69.1
Total	**71.9**	**64.1**

Table 65. Health Care Utilization Rates per 10,000 Population

	Nursing Home Pop	Home Health Pts	Long-Term Hospital Inpt Days	Psych Hospital Inpt Days	Outpt Hospital Visits	ER Visits	General Hospital Inpt Days
Both shortage and persistent poverty	77.4	51.9	403.4	1,101.9	20,188.5	5,807.9	5,546.4
Shortage but not persistent poverty	98.5	58.2	2,155.6	4,672.6	27,374.2	5,927.5	9,883.4
Persistent poverty but not shortage	80.9	51.8	4.9	237.4	13,681.4	3,784.6	3,020.8
Neither shortage nor persistent poverty	87.3	55.6	45.3	563.1	16,900.5	3,507.0	3,669.2
Total	87.3	55.3	226.8	888.9	17,571.6	3,803.8	4,180.7

Although the intent of this exercise was to simply compare the characteristics of counties with and without shortage designations, the results indicated that persistent poverty counties did not necessarily suffer from RN shortages. Combining the relatively strong supply of RNs per population with the relatively low rates of utilization per population, as well as the persistent poverty, counties failed to make the cut because their number of RNs relative to use of services was relatively high.

2. Supply Validation

The next validation exercise compares the commuting-adjusted Census figures to the non-adjusted Census figures and the numbers available in the NSSRN.

Adjusted versus Unadjusted Census figures. Overall, we found that the commuting-adjusted estimates from the Census were better than the unadjusted Census estimates. Fully half (49.6%) of the time, the adjusted Census figures were the best estimate (compared to both unadjusted Census and NSSRN figures). Another 26.7% of the time, unadjusted figures would have been more accurate, and 23.7% of the time NSSRN figures would have been more accurate. It was worth noting, however, that when the unadjusted figures provided a closer estimate, the difference between the unadjusted and adjusted figures was generally small (an average of 17.7 RNs in all validation counties, with the adjusted figures being an underestimate).

There were some differences by county type, however, in terms of whether the adjusted or unadjusted figures were more accurate. In counties with a RUCC code of four (non-metro urban counties of 20,000 or more, adjacent to a metro area), the unadjusted figures were the most accurate 46.5% of the time, compared to the adjusted figures being more accurate 37.2% of the time. Also, in counties with a RUCC code of 9 (non-metro completely rural counties of less than 2,500 urban population, not adjacent to a metro area), the unadjusted figures were the most accurate 40.7% of the time, compared to 30.5% of the time for adjusted figures.

Again, however, the raw numbers tended to be small. In counties with a RUCC code of 9, when unadjusted figures were superior they were nonetheless very similar to the adjusted figures --

different by only 0.24 RNs on average. This did not seem a convincing rationale for using the unadjusted figures in these counties, particularly because in counties where the adjusted figures were superior, the differential between adjusted and unadjusted was much greater -- 10.4 RNs on average. A similar pattern was obtained for counties with a RUCC code of 4. When the unadjusted numbers were better, they were only different from the adjusted numbers by an average of 13.6 RNs. When the adjusted numbers were better, they were different from the unadjusted numbers by an average of 103.1 RNs. Clearly, the use of adjusted numbers, even when unadjusted numbers were closer, was less likely to produce a highly skewed result than using unadjusted numbers when adjusted numbers were closer.

These analyses provided little evidence to indicate that the county-level estimates of the supply of RNs based on the NSSRN would ever be a better choice than estimates based on Census data.

Accuracy of Census Estimates. In sum, the percent of the time estimates were substantially off (here arbitrarily defined as more than 50% in either direction), was 28% using adjusted Census numbers, 40% using unadjusted Census numbers, and 58% using NSSRN numbers. This substantially deteriorated for counties in RUCC categories 8 (50% substantially off) and 9 (41% substantially off), which was a significant error given that in very small counties a differential of 3 or 4 RNs could exceed 50% of the actual supply. The fact that the differentials in counties with a RUCC code of 8 or 9 tended to be overestimates could potentially prevent those rural counties from being designated. In fact, of the 49 counties in the five states with a RUCC code of 8 or 9 and estimates that were substantially off, 63% had actual RN supplies that fell below RN demand but had estimated RN supplies that exceeded RN demand. Still, relatively few of those counties would become shortage counties based on this change. About one in eight (12.4%) of the counties in the five states actually stood to lose their shortage designation based on the inaccuracies of the estimates, while 6.2% stood to gain a shortage designation inappropriately as a result of these inaccuracies.

Who loses? The effect of inaccuracies on where the counties actually fell in terms of shortage status was worst for counties with a RUCC code of 8. About one in five (20.4%) of these counties would have qualified for shortage counties if actual data were available rather than Census estimates. And 18.2% of counties with a RUCC code of 3 would have been similarly incorrectly designated.

This was also a problem for persistent poverty counties, 21.3% of which were misclassified in the five states as not qualifying for a shortage designation, while 9.8% were misclassified as qualifying for a designation inappropriately. Overall, the estimated data resulted in 36% of persistent poverty counties being classified as shortage counties, while the actual data resulted in 48% of the persistent poverty counties being classified as shortage counties.

Error introduced by sampling versus the commuting adjustment. It was possible to separate the error introduced by sampling from the error introduced by the commuting adjustment in states in which actual RNs by county data were available both for place of residence and place of employment. This level of data was available for both New York and North Carolina. In some cases, the residence figures from the Census were virtually identical to the residence figures from the state, but the employment estimate was substantially off from the actual employment figures, indicating that the commuting adjustment was the source of most error. In other cases, however, calculating the commuting adjustment actual residence data would have resulted in employment estimates almost identical to actual employment figures, indicating that the difference between estimated and actual resident RNs was the source of most error. When these patterns were quantified and measured, it proved to be the case that sampling error was implicated no more or less often than commuting error. The only clear pattern was that more rural counties had more of both types of error than did urban counties. Furthermore, it was usually the case for individual counties that much more of the error originated with one source than the other.

3. Conclusions

Nearly the only way to correct effectively for sampling errors was to use data taken from a larger sample. The counties suffering the greatest effects of sampling error were more rural counties, and 75-80% of the time, the sampling error overestimated RNs living in these counties. It might be possible to adjust the resident RN estimates down for the rural counties (with the understanding that this would worsen the estimates for the 20-25% of counties in which the Census data were an underestimate), but actual residence data for RNs were needed from more than two states to accurately judge how much of an adjustment would maximize accuracy.

It may be possible to correct for commuting adjustment error in a relatively systematic way. It was expected that differences between RN commuting patterns and the commuting patterns of the workforce overall might be largely explained by the health infrastructure existing in a given county and its neighboring counties. Again, however, data on actual RN commuting patterns were needed from more than two states in order to fully study the divergence between RN and civilian labor force commuting patterns.

VI. Conclusions and 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. These recommendations are presented below.

1) Of the methods examined in this study, the Preferred Method outlined in this report was the best choice for assessing the severity of nursing shortages in counties in the U.S. It met 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 would be 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, community, county, or region of the country? If so, how should the bias be addressed and 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 due primarily 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 or 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 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. A promising avenue for research will open up when the revised Nursing Supply Model becomes available sometime in 2007.

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. Hecker D. (2001). *Occupational employment projections to 2010. Monthly Labor Review*. U.S. Bureau of Labor Statistics, November 2001.

6. 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.

7. 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.

8. U.S. Government Accounting Office. (2001). *Nursing Workforce: Emerging Nurse Shortages Due to Multiple Factors*. http://www.gao.gov/new.items/d01944.pdf

9. 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/

10. 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.

11. 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.

12. Aiken L., Clarke S., Sloane D., Sochalski J., & Silber J. (2002). Hospital nurse staffing and patient mortality, nurse burnout, and job dissatisfaction. *Journal of the American Medical Association, 288*:1987-1993.

13. American Nurses Association. (2002). Government Affairs, State Government Relations: *2002 Legislation: Nurse Staffing Plans and Ratios*. http://nursingworld.org/gova/state/2002/staffing.htm