The study examined 7 demographic and ecological variables that influence the distribution of physicians in North Carolina: population size; percent change in population size; per capita income (population composition); "medical environment" - hospital facilities (number of hospitals and beds); physicians' proximity to a hospital center; and proximity to a medical school. The use of stepwise regression analysis and multivariate analysis of variance permitted an assessment of the relative importance of each of these variables for counties and hospital service regions. Findings indicated that substantial amounts of the variations in 3 of the 4 categories of physicians (physician/population ratios, percent of physicians who are general practitioners, and percent of specialists who are board certified) could be explained by one or a combination of 3 variables -- hospital facilities, population size, and income. (Author/KM)
A DEMOGRAPHIC AND ECOLOGICAL ANALYSIS
OF THE
RURAL-URBAN DISTRIBUTION OF PHYSICIANS IN NORTH CAROLINA

by

M.R. Grigg
U of Maryland
College Park, Maryland

Presented at the Annual Meeting of the Rural Sociological Society,
Montreal, Quebec, Canada. August, 1974.
ABSTRACT


This study examined seven demographic and ecological variables that influence the distribution of physicians in North Carolina: population size; percent change in population size; per capita income (population composition); and "medical environment" - hospital facilities (number of hospitals and number of beds); physician's proximity to a hospital center; and proximity to a medical school. The use of stepwise regression analysis and multivariate analysis-of-variance permitted an assessment of the relative importance of each of these variables with regard to county data and data grouped into hospital service regions. The findings indicated that substantial amounts of the variation in three of the four categories of physicians (physician/population ratios, percent of physicians who are general practitioners, percent of specialists who are board certified) could be explained by one or a combination of three variables - hospital facilities, population size, and income.
A Demographic and Ecological Analysis of the Rural-Urban Distribution of Physicians in North Carolina.

I. Introduction

A. Relevant Background Literature.

In recent years much time and energy has been expended pursuing information on the nature and extent of the maldistribution of physicians between rural and urban areas in the U.S. The literature pertaining to physician distribution is extensive for the nation as a whole and for individual states. This literature cites studies, most of them descriptive in nature which deal with the distribution of physicians in terms of a rural-urban continuum. These studies consistently show that the distribution of physicians in rural and urban areas is unequal and becoming increasingly unequal, i.e., the rate of physicians to population is low in rural areas and declining, and high in urban areas and rising.

One study dealing with the rural-urban distribution of physicians is Health manpower in the Upper Midwest (1966). The Health Manpower Commission postulated that physicians have followed the national trend of population shifts from farms and rural areas to urban centers for many years, but that they may be shifting even more rapidly than the population as a whole. Their data (active physicians/100,000 population ratios by degree of urbanization of county, 1940, 1950, 1960, for Minnesota, North
Dakota, South Dakota, and Montana) supported this contention. In general, they found that the ratio of physicians to population was declining in the most rural counties and increasing in those that were most urban.

In their investigation of "no-physician towns" (defined as a town without any physician and also fifteen or more miles from another place with at least one active physician), the Health Manpower Commission found that a population of less than 500 persons was characteristic of these towns. Also, most of the no-physician towns had a declining population or were growing at a slower rate than the national average.

These findings are repeated again and again in all areas of the nation. The literature (American Medical Assn., 1966; Fein, 1954; Lyons, 1967; Wather and Yaramoto, 1958; Rinlincr, 1963; Stewart and Pennell, 1960) documenting this disparity of distribution of physicians between rural and urban areas will therefore not be reviewed because of its consistency.

With regard to the characteristics of rural and urban physicians, the literature reported that the median age of rural physicians was higher than that of urban physicians, and appeared to be increasing. The majority of the physicians in rural areas were general practitioners, whereas the majority of physicians in urban areas were specialists. High income areas (usually urban areas also) attract younger physicians who are specialists, while low income areas (if they have a physician at all) tend to attract
physicians in general practice, most of whom are older. Thus the literature showed that urban, high income areas had more physicians, more specialists, fewer general practitioners, and more young physicians in proportion to their population than do rural, low income areas.

Another aspect of the rural-urban distribution of physicians phenomenon is the physician's choice of a practice location, and the factors influencing his/her decision. In general, studies investigating these factors have shown that physicians tend to choose a practice location for professional reasons - standard of medical care, availability of hospital facilities, proximity to a medical center, work load, opening for specialty, prospects for building a good practice, etc. Social, cultural, and educational factors and family ties and influences also had an important effect on the physician's decision. Physicians choosing to locate in urban areas seemed to be influenced more often by professional factors in their choice than were their counterparts in rural areas. The latter appeared to be more influenced by their families and by social, cultural, educational, and recreational factors.

Two studies of the influence of demographic and ecological factors on the social structure are noteworthy. The first, by Hawley (1941), examined the relationship between certain demographic variables and urban service institutions ("agencies established for the service of the needs of the general
The observation that the number and variety of these institutions depends directly upon certain demographic and environmental factors was evidenced by Hawley's findings, which indicated that "population differences are more or less closely related to differences in the institutional structure of cities." (Hawley, 1941:633.) The variations within groups of cities of similar size were found to be greater than those between size groupings. Many of these differences, however, could be attributed to variables of population composition, among which Hawley found income to exert the most important influence.

A second study, by Marden (1966) examined the influence of five demographic and ecological variables on the distribution of physicians within metropolitan areas in the United States. Marden patterned his study after the aforementioned study by Hawley, and set out to test the following hypothesis:

Differences in the distribution of institutionalized medical services provided by physicians in the metropolitan areas of the United States in 1960 are associated with differences in the population that supports them and in the environment in which they practice. (Marden, 1966:293.)

The demographic and ecological variables employed by Marden were: population size; population composition - age, race, education; and "medical environment" - the presence of hospital facilities.

Marden's findings - in accord with Hawley's - indicate that these five variables significantly influence the distribution of physicians' services and that there are marked differences between their influence on general practitioners and on specialists. All variables studied combined to explain substantial amounts of the
variation in the distribution of the three categories of physicians - total physicians, general practitioners, and specialists - under consideration. (Harden, 1966:300.)

b. The Problem.

This investigation was similar in content to that of Harden's study, to the extent that his major hypothesis may be paraphrased to serve as the major hypothesis for this analysis: Differences in the distribution of institutionalized medical services provided by physicians in North Carolina in 1966 were associated with differences in the population that supported them and in the environment in which they practiced.

The two studies differ in the demographic area units utilized and in choice of variables. Harden utilized metropolitan counties as his units of analysis. There are acknowledged problems with the use of county data or with any other data applicable to political areal units. The problem of defining a population as being serviced by a particular set of physicians simply because the population and the physicians are all located in the same county unit is obvious. This study attempted to cope with problems inherent in county data by defining hospital service areas consisting of groups of counties related to a "core county" containing a hospital service center. These groups of counties were shown to be related to the core county in terms of utilization of medical services (see discussion of unit of analysis below).
Harden used population size, population composition - race, education; and "medical environment" - hospital facilities, as his independent variables. In this study, population size, percent change in population size, population composition - income, and "medical environment", i.e., presence of a general hospital(s) and distance from the nearest large hospital center and to the nearest medical school were the independent variables.

Harden treated his dependent variable - distribution of physicians - in terms of total physicians and in terms of general practitioners and specialists. In this study physicians were treated according to this same classification.

This analysis extended Harden's analysis of the influence of demographic and ecological factors on the distribution of physicians. It has been expanded from a consideration of distribution within a metropolitan area to a consideration of distribution within the entire state of North Carolina, both metropolitan and rural areas. The studies by Fawley and Harden and the extensive work in the whole field of physician distribution all buttress the major hypothesis of this study stated above. The literature also lends itself to the development of several research hypotheses:

1) Total physicians per population ratios are directly related to: population size; percent change in population size; income level of the population; number of hospitals and number of hospital beds; and inversely
related to distance from a major hospital center and from a medical school.

2) Percent of total physicians who are general practitioners is inversely related to: population size; percent change in population size; income; number of hospitals and number of hospital beds; and directly related to distance from a hospital center and from a medical school.

3) Percent of specialists who are board certified is directly related to: population size; percent change in population size; income; number of hospitals and number of hospital beds; and inversely related to distance from a hospital center and from a medical school.

4) Median age of physicians is inversely related to: population size; percent change in population size; income; number of hospitals and number of hospital beds; and directly related to distance from a hospital center and from a medical school.

C. Theoretical Framework.

Generally, the theoretical framework for this analysis was a human ecological one, involving the relationship between environment and social structure, where social structure was the variable to be explained. Hawley's definition of environment was adopted here. Environment is:

composed of all external conditions or factors that affect human behavior. (Hawley, 1951:529)
...environment refers to the medium in which an organism exists. Environment comprises the raw materials of life and the conditions, favorable and unfavorable, that affect the use of those materials. (Hawley, 1950:12-13)

Organization and its product, social structure, are the means by which human beings adapt to their environment. Society exists by virtue of the organization of a population of organisms, each of this is individually equipped to survive in isolation. Thus we have a human ecological study - "a study of the morphology of collective life in both its static and dynamic aspects." (Hawley, 1950:67)

This study therefore focuses on the manner in which various human activities are ordered by the population that supports them and the environment in which they exist. An observation made by Duncan describes the general problem of this research effort:

institutions or institutionalized services differ because:

each territorially delimited aggregate confronts a special set of environmental circumstances and differs from other such aggregates in size and composition. (Duncan, 1959:6n3)

In this study, demographic variables are within the scope of ecological research as independent variables, as "determinants or limiting conditions of ecological organization." (Duncan, 1959:6n3) Hawley stated (1957:361), "Demographic variables, together with population size, are important conditions affecting the formation and change of social structure."

Institutionalized medical services offered by physicians were selected as the object of this analysis - as that aspect of the social structure whose variation is to be explained - to study the
differences in the distribution of physicians serving the state of North Carolina as influenced by their environment. The demographic factors considered were population size, percent change in population size, population composition - income, and "medical environment", including number of hospitals and number of hospital beds in area, proximity of physicians to a large hospital center, and proximity to a medical school.

II. Methods and Procedures.

A. Measures and Sources of Data.

The data needed to measure each of the variables in this study were readily available in various publications and from the North Carolina Regional Medical Program. The detailed sources are cited elsewhere (Grier, 1959).

B. Unit of Analysis.

The entire state of North Carolina was included in this study. Data were analyzed by hospital service areas and by county. For the purposes of this study North Carolina was divided into seventeen hospital service areas, each consisting of one or more counties located around a core county containing a general hospital with 200 or more beds. The seventeen core counties were selected from 20 counties having a hospital of that size. The three counties which were not selected as centers were in all cases adjacent to counties chosen as cores. The counties included in each service area were chosen by means of several procedures.
The first of these procedures - a good indicator of routine hospital use - involved determining the number of births to mothers residing in a particular county and then computing the percentage of hospital births occurring outside the mother's county of residence. Utilizing the interrelationships between counties established by this first procedure, final placement of the counties into hospital service regions was made by using procedures involving three different divisions of the state of North Carolina - divisions utilized by the North Carolina Regional Medical Program (1960). These divisions were based on general hospitalization patterns, on geography, and on rural-urban characteristics.

Using these procedures and divisions, the interrelationships between counties were delineated, and each county was placed in a hospital service region. The result of these methods is the seventeen hospital service regions shown in Map 1. The core counties of each region are the shaded counties.

The general problem with these units of analysis and their definition is the problem of defining areal units which are sociologically meaningful. It was in the attempt to avoid sociologically meaningless areas that county units and geographical areas were eliminated as the sole units of analysis, since people frequently cross county lines to obtain medical services. Nevertheless, the use of the political boundary makes it possible to examine the relations between county and other
socioeconomic data since the Bureau of the Census and similar agencies rather their statistics on a county rather than a service trading area basis. Also, planning, particularly where tax funds are involved (as in the Hill-Burton program) is frequently best carried out on the county level. To cut across county lines would be most difficult. Both of these considerations underlay the choice of county data for this study, as supplemented by data for counties gathered into hospital service regions.

C. Analysis of Data.

The first phase of the analysis of the data in this study involved the use of a stepwise regression program (Efron et al., 1960). All of the independent variables in the study were run against each of the dependent variables. In each of these analyses population size was controlled, by forcing it into the regression equation first. An analysis was also made of all of the independent variables except population size against the dependent variable physicians per 10,000 population. The purpose of the runs was to discover which independent variables - after controlling on population size - were the best predictors of each of the dependent variables.

A multivariate analysis of variance program was used in the second phase of this analysis. (Clyde, et al., 1966) It was used to compare the seventeen hospital service regions in terms of the dependent variables of this study. The program was designed to perform univariate and multivariate analysis of variance, of
covariance and of regression. It provides an exact solution in either the orthogonal or the nonorthogonal case.

The first analysis with this program controlled population size by using it as a covariate. The purpose of the analysis was to discover whether there were large differences between the values of the seventeen regions for each of the dependent variables. When such differences were found, other analyses were made to determine which independent variables best explained the differences. The results of the step-wise regression analysis were used here. The independent variables found to be the most important predictors of each of the dependent variables by the regression analysis were run in the multivariate analysis as covariates to discover if they were also important with regard to the hospital service regions.

III. Findings.

A. Results of Stepwise Regression Analysis of County Data.

Population size was the only independent variable controlled throughout the stepwise regression analysis. Thus all of the independent variables - with population size controlled - were run against each of the dependent variables.

The results of the analysis showed that the two dependent variables which deal directly with number of total physicians (total physicians and physician/population ratios) may both be predicted best by the independent variable number of hospital
beds. The three remaining dependent variables deal primarily with characteristics of physicians and the independent variable income was important in predicting two of them. In the case of median are of physicians, income, of all of the independent variables, was the best predictor, but even it was not significant. Income alone was the best predictor of percent board certified, whereas with percent general practitioners, income in the company of population size and number of hospitals comprised the best set of predictors.

Table 1 displays the relationship between the dependent variables and the independent variables with which each is most highly correlated, as shown for individual variables by the correlation matrix and as shown for the joint distribution of the variables by the regression analysis. For the correlation matrix all variables correlating >+.60 with the dependent variable are listed in order of highest correlation to lowest. For the regression analysis, only those independent variables which were found to be significant predictors of the dependent variable are listed. The analyses predict quite similar relationships.

3. Results of Multivariate Analysis of Variance of Hospital Service Regions.

The second phase of the data analysis utilized a multivariate analysis-of-variance program and procedures as described above. This analysis dealt with data at the hospital service region level but still used county data as the sampling unit. The purpose of
TABLE 1

RELATIONSHIPS OF VARIABLES

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>Related Independent Variables According To:</th>
<th>Correlation Matrix</th>
<th>Regression Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Physicians</td>
<td></td>
<td>Income</td>
<td>Hospital Beds</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Population Size</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Hospital Beds</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Number of Hospitals</td>
<td></td>
</tr>
<tr>
<td>Median Age</td>
<td></td>
<td>Income*</td>
<td>Income**</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Distance - Medical School</td>
<td></td>
</tr>
<tr>
<td>Percent GPs</td>
<td></td>
<td>Income</td>
<td>Income</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Hospital Beds</td>
<td>Population Size</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Population Size</td>
<td>Number of Hospitals</td>
</tr>
<tr>
<td>% Board Certified</td>
<td></td>
<td>Income*</td>
<td>Income</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Population Size*</td>
<td></td>
</tr>
<tr>
<td>Physicians/Population</td>
<td></td>
<td>Hospital Beds*</td>
<td>Hospital Beds</td>
</tr>
<tr>
<td>Ratios</td>
<td></td>
<td>Income*</td>
<td></td>
</tr>
</tbody>
</table>

*This variable does not correlate over .60, but it is one of the two highest correlates of this dependent variable.

**This variable was not significant, but was the best predictor of median age.
this analysis was to discover whether any of the seventeen hospital service regions differed significantly from each other in terms of the dependent variables. As mentioned previously, population size was the only independent variable controlled in the first part of the multivariate analysis.

Table 2 shows the results of regressing population size on as many as four dependent variables repeatedly and on each of them alone. As shown, the P values are less than .001 in all instances except in the case of median age. This means that when these variables are regressed on population size, all but median age are highly significantly related to population, and therefore population must be controlled when analyzing them, hence justifying the use of population size as a covariate in the analysis of these variables.

(Table 2 about here)

Table 3 shows the results of the between cells analysis. The P values here show the differences between all of the hospital regions for all of the dependent variables to be significant at .041. With regard to the individual variables, only for percent general practitioners and physician/population ratios are the regions significantly different from each other. Their P values are .015 and .011 respectively. This implies that the means of each of these variables are significantly different in at least two regions. It should be noted here that these two variables are highly correlated with each other.
### TABLE 2

**Multivariate Test of Within Cells Regression**

Tests of Significance Using Wilks Lambda Criterion:

<table>
<thead>
<tr>
<th>Test of Roots - F</th>
<th>DFHP</th>
<th>DFERR</th>
<th>P Less Than</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 through 1</td>
<td>30.661</td>
<td>4.000</td>
<td>79.000</td>
</tr>
</tbody>
</table>

**Univariate F Tests:**

<table>
<thead>
<tr>
<th>Variable</th>
<th>F(df=1.82)</th>
<th>P Less Than</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>0.433</td>
<td>0.512</td>
</tr>
<tr>
<td>% General Practitioners</td>
<td>83.323</td>
<td>0.001</td>
</tr>
<tr>
<td>% Board Certified Physician</td>
<td>47.551</td>
<td>0.001</td>
</tr>
<tr>
<td>Physician Ratios</td>
<td>46.745</td>
<td>0.001</td>
</tr>
</tbody>
</table>

### TABLE 3

**Multivariate Test of Hospital Regions**

Tests of Significance Using Wilks Lambda Criterion:

<table>
<thead>
<tr>
<th>Test of Roots - F</th>
<th>DFHP</th>
<th>DFERR</th>
<th>P Less Than</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 through 1</td>
<td>1.376</td>
<td>64.000</td>
<td>311.547</td>
</tr>
</tbody>
</table>

**Univariate F Tests:**

<table>
<thead>
<tr>
<th>Variable</th>
<th>F(df=16.82)</th>
<th>P Less Than</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>1.133</td>
<td>0.340</td>
</tr>
<tr>
<td>% General Practitioners</td>
<td>2.119</td>
<td>0.015</td>
</tr>
<tr>
<td>% Board Certified Physician</td>
<td>0.609</td>
<td>0.868</td>
</tr>
<tr>
<td>Physician Ratios</td>
<td>2.197</td>
<td>0.011</td>
</tr>
</tbody>
</table>
On the basis of this first part of the multivariate analysis, the data for percent general practitioners and physician/population ratios were reanalyzed with additional covariates. The results of the stepwise regression analysis were employed here, and the best predictors of each of these variables were entered as covariates and thereby controlled. The covariates for percent general practitioners were income, population size, and number of hospitals, and for physician/population ratios, they were population size and number of hospital beds. The F and P values for this regression show that the regression of the three covariates with percent general practitioners was highly significant, a fact which confirmed the regression analysis. The analysis of hospital regions (between cells) showed highly significant F and P values. Thus there were highly significant differences between regions. The three independent variables used as covariates did not explain the differences.

In the case of physician/population ratios, the reanalysis showed that this dependent variable’s two covariates did explain the differences between regions. The regression of the covariates with the dependent variable was shown to be highly significant by the F and P values. However, the F and P values for comparison of hospital regions were not significant, indicating that the two covariates population size and hospital beds explained the differences between hospital regions for the dependent variable.
physician/population ratios.

The only dependent variable to emerge from the multivariate analysis-of-variance still showing significant differences between hospital regions was percent general practitioners. It was felt that further analysis of this variable on the hospital region level was needed, so a multiple comparison test was administered using the means of the regions. The procedure chosen was Tukey's procedure for comparing individual means. (Edwards, 1960:330-332) Tukey's test for a significant gap was used to discover which regions differed significantly from each other. The results showed a slight patterning of differences between hospital regions with a metropolitan county as core and regions with a semi-rural county as core.

IV. Discussion of Findings.

The regression analysis showed population size to be an important predictor only of percent general practitioners, and then it is a good predictor only when accompanied by income and number of hospitals (See Table 1). However, the multivariate analysis discredited the influence of population size by showing that there were significant differences between hospital regions even after all three variables were controlled. Thus population size did not appear to be an important factor in explaining the variation in any of these measures of physician distribution.

Percent change in population size also related as predicted.
However, it did not correlate highly with any of the dependent variables (See Table 1), and did not appear in the regression analysis as an important predictor of any of them.

Income made a better showing. It correlated in the same (predicted) directions with the same variables as did both population size and percent change in population size, but it correlated highly with all of the variables, and was a significant predictor of two of them - percent general practitioners and percent board certified (See Table 1). However, the multivariate analysis of percent general practitioners showed that income could not explain away the significant differences between hospital regions.

The independent variable number of hospitals related in the expected direction with each of the dependent variables. It correlated highly with total physicians only, and just under -.60 with percent general practitioners, but was an important predictor of the latter only (See Table 1). Number of hospitals suffered the same fate as population size and income in the multivariate analysis of general practitioners.

Number of hospital beds followed the same pattern of relationship with the dependent variables as did the independent variable just described, and it related highly with three of them - total physicians, percent general practitioners, and physician/population ratios (See Table 1). The multivariate analysis of physician/population ratios showed number of hospital
beds to be capable of accounting for the significant differences between the hospital regions on that variable.

The independent variables of distance to the region's hospital center and distance to the nearest medical school each related as predicted. However, only distance to the nearest medical school was highly related to any dependent variable (median are). Neither of these variables was significant as a predictor of any of the dependent variables.

Thus, the independent variables which appeared to be most important for the prediction of and explanation of the variation in the dependent variables were number of hospital beds, income, population size, and number of hospitals. These variables did not explain all (and in some cases they explained very little) of the variation in the dependent variables. Other factors aside from the ones tested in this study need to be investigated for their power to explain the distribution of physicians. Such other factors might include an urban-rural index other than the one utilized in this study - population size, percent change in population size, and per capita income. Other indicators of urban-ruralness might include measures of: number and variety of certain non-medical services provided in an area; industrialization; etc. Personal factors influencing the physician's choice of a practice location are also worthy of investigation for their power to explain distribution.

While certain questions remain unanswered, and several
problems have been identified for future study, the results of this study confirmed the general hypothesis and the research hypotheses which guided the analysis. In theoretical terms, this study showed that differences in the social structure - distribution of institutionalized medical services provided by physicians in North Carolina - were associated with differences in the population that supported them and in the environment in which they practiced.
BIBLIOGRAPHY

Association for the North Carolina Regional Medical Program
Flalock, Hubert W.

Clyde, Dean J., et al.

Coral Gables, Florida: Biometric Laboratory, University of Miami.

Department of Tax Research, State of North Carolina.

American Medical Association

Duncan, Otis D.

Duncan, Otis D. and Leo F. Schnore.
1957 "Cultural, Behavioral, and Ecological Perspectives in the Study of Social Organization," American Journal of
Sociology 65 (September): 132-149.

Edwards, Allen L.

Efroymson, H.A.

Fein, Rashi.

Gins, .R.

Hamilton, C. Horace.

Havley, Amos II.

Havley, Amos II.

Hawley, Anos N.


Health Manpower Study Commission.


House, Floyd.


Lazarsfeld, Paul F.


Lyons, Richard D.


Marden, Parker C.


Mather, William C. and Kinzo Yamanoto.

1953  Pennsylvania Rural Physicians - Their Satisfaction With

Mechanic, David

Parker, Ralph C., Jr., and Thomas R. Tuxill.

Rimlinger, Gaston V., and Henry H. Steele.

Stewart, W.M., and J.Y. Pennell