MEDICAL MANPOWER MODELS: NEED, DEMAND, AND SUPPLY

PREPARED FOR THE STATE OF CALIFORNIA DEPARTMENT OF HEALTH

JUDITH R. LAVE
LESTER B. LAVE
SAMUEL LEINHARDT

R-1481-CHD
MARCH 1974

Rand
SANTA MONICA, CA 90406
This report, prepared under contract with the State Health Department, examines alternative methods to estimate health manpower requirements and to determine physician shortage. A companion report, Albert J. Lipson, *California Health Manpower: An Overview of Trends and Policy Issues*, R-1572-CHD, discusses supply and distribution of physicians and nurses, medical education, health manpower licensure, and health planning.

These reports are intended primarily to assist governmental officials in examining alternative approaches to deal with emerging problems in the health manpower field. They are produced as part of Rand's continuing health program aimed at analysis and assessment of health policy alternatives.
SUMMARY

Government agencies have taken on the role of planning the delivery of health care and so find themselves confronted with defining "need." How many physicians, hospital beds, ancillary health personnel, and other health facilities are needed in an area? What policies will serve to increase the supply of these health resources when they are needed? An earlier paper (Lave and Lave, 1974) examines this set of questions for hospitals, focusing on the financing of new facilities. This study attempts to answer these questions for physicians.

Section II raises the general issues that must be kept in mind in any examination of health issues, including the absolute and relative efficacy of personal health services and the various ways of producing these services.

Section III reviews models used to forecast the need for physicians. Approaches based on professionally determined requirements and on physician/population ratios are deficient in many ways. There is little indication that the additional services would be used; and if they were used, whether they would be efficacious, or whether additional physicians would choose to settle in the areas of greatest need. Simplicity is perhaps the only criterion that recommends these approaches. Economic models that focus on demand and its determinants have the virtue of forecasting use. We hope that these models will become bolder and include nonmarket as well as market rationing devices. Systems models explicitly include supply and demand factors and thus are conceptually superior to economic demand models. However, existing models have two crucial deficiencies. They do not contain output (health status) indices and have been estimated with only highly aggregate data. Thus, their conceptual contribution is limited. A model that includes indices of health status (Lave, Lave, and Leinhardt, 1974) has been described, but the unavailability of data has so far precluded its estimation.

Section IV discusses models used to forecast the supply of physicians. The focus is on determinants of physician location, since locational mal-distribution is a principal issue and since governmental policies in this area could be effective immediately. Physicians seem to be responding
more to the quest for environmental amenities, including the advantages of urban life, than to narrow income incentives. Within cities, location choice is crucially related to the opportunity to interact with colleagues and to be close to medical care and teaching facilities. For a number of reasons, providing medical services to low density areas that are distant from major metropolitan areas may be the most difficult problem faced in planning health care delivery.

Section V looks at six ways that might be used to determine if an area has a physician shortage. The approaches based on professional standards and physician/population ratios have the disadvantages discussed in connection with forecasting physician requirements. Examining the rate of return to physician education does no more than give an indication of whether there are artificial constraints on supply. Quantifying the extent of nonprice rationing of physician services at least gives some idea as to whether additional services would be used. An alternative to examining nonprice rationing is to survey population satisfaction with the individual care delivery system. A more objective approach to determining the need for additional physicians involves determining the health status of the population and the efficacy of physicians in improving this. However, this approach must be coupled with one surveying nonprice rationing to determine whether the additional services would be used.

Forecasting future shortages of physicians is the subject of Section VI. Investigators have (1) systematically overestimated the population, (2) underestimated the supply of physicians, (3) failed to take account of productivity increases for physicians, and (4) thus, systematically overestimated the magnitude of shortage. Indeed, rather than any future shortage, there may well be a physician glut.

Although the overall number of physicians does not seem to warrant changes in governmental policy, locational problems are important. The research has implications for determining whether an area is underserved and what might be done to increase the supply of physicians. A major unsolved problem is the provision of services to rural areas distant from major cities.
CONTENTS

PREFACE ......................................................... iii

SUMMARY ......................................................... v

Section

I. INTRODUCTION ..................................................... 1

II. SOME ISSUES IN HEALTH CARE DELIVERY ......................... 3

III. MODELS USED TO FORECAST PHYSICIAN REQUIREMENTS .......... 6
    Introduction .................................................. 6
    Normative Approaches to Forecasting Manpower
    Requirements ................................................. 7
    Manpower Planning Using the Ratio Approach .................. 13
    Economic Models ............................................. 17
    Systems Models of Health Care ............................. 22
    Summary and Critique of the Models ......................... 25

IV. MODELS OF THE SUPPLY OF PHYSICIAN SERVICES ................. 27
    Approaches to Studying Physician Supply .................... 27
    The Rural-Urban Choice .................................... 31
    Intra-Urban Location Decisions ............................ 34
    The Role of the Foreign Medical Graduate ................... 38
    New Physicians .............................................. 41
    Summary of the Planning Implications of Current Supply . 43

V. CONCEPT OF A PHYSICIAN SHORTAGE ............................. 45
    Conclusion .................................................. 48

VI. HOW GOOD ARE WE AT FORECASTING SHORTAGES? .................. 49

BIBLIOGRAPHY ...................................................... 53
I. INTRODUCTION

Federal, state, and local governments are inextricably bound up in the medical care system, an involvement that has been increasing over time. Direct provision of care has the longest history, followed by the provision of funds for research and training, the financing of care, and finally the planning of health care delivery. A next logical step is to proceed from the financing of care and the provision of training funds to ensuring that the supply of medical care (location of physicians, facilities, and specialty mix) in a particular locale is adequate. Yet, the determination of what quantity and mix of manpower and facilities are "adequate" is far from trivial, and it is still more difficult to develop reasonable policies that ensure these conditions are met.

Charles Edwards, M.D., Assistant Secretary for Health (HEW), emphasized manpower problems in a speech to the annual meeting of the Association of American Medical Colleges (quoted by the American Council on Education, 1973):

"Less than eight months from today -- on June 30, 1974 -- virtually the entire statutory base for present Federal support of health manpower will come to an end," he said. "On that date, the Comprehensive Health Manpower Training Act, the Nurse Training Act, the Public Health Training Act, and the Allied Health Professions Personnel Training Act will terminate, which means that we now have to decide what forms of legislation -- if any -- should replace these laws."

Total Federal spending for health manpower over the past ten years was nearly $3.5 billion, he said, adding: "Plainly, in my judgment, we now have to ask: What has been the return on this investment?" It seems clear, he said, "that there is little reason to adopt a manpower policy that involves investing large sums of money to further expand our capacity to train health professionals."

If the U.S. merely maintains the current output capacity of health professional institutions, he said, by 1985 there will be 50 percent more physicians, 40 percent more dentists, and 60 percent more registered nurses that there were in 1970. The ratio of physicians in 1985 may reach nearly 220 per 100,000 population, he added, compared to 160 in 1970 and 140 in 1960.
"In my judgement," Edwards said, "even more significant is the possibility we may well be facing a doctor surplus in this country. A number of authorities now see this as a distinct possibility, one that must figure very heavily in both our immediate and long-range planning in the health manpower field."

If one considers the goal of the past decade to have been one of producing more physicians, dentists, nurses, and allied health professionals, he said, "then you would have to conclude that these efforts have been highly successful." But, he said, "if you look at health manpower development efforts as part of a broader national purpose -- namely to make health services available to everyone at a cost that both the individual and the society can afford -- then it is clear that all of us -- not just government, but the whole health community, public and private -- have to reassess our priorities in the manpower field and perhaps adopt some fundamentally different ones. I am not at all sure that a further increase in our capacity to train physicians is in the best national interest at a time when population growth rates are declining and there is ample opportunity to expand the productivity of practicing physicians. There are, after all, some serious disadvantages to having too many doctors...."

Edwards said he had "some serious questions" about the extent to which medical schools and other health teaching institutions "have made an enlightened response to the kind of problems that affect the whole system of health care." He identified the problems as: "rising health care costs that are literally approaching the limit that society is willing to pay; uneven quality in the services that physicians and others provide; and serious imbalances in supply and demand that leave some areas with far too many resources and others literally with none." By and large, he said, health training institutions have "tended to regard these problems as somebody else's."

Doctor Edwards might have added the following figures further detailing the government's role in health care: In 1970, approximately 37.6 percent of all health expenditures and 35.8 percent of personal health expenditures were government financed, whereas in 1950, 27.2 percent of total health expenditures and 22 percent of personal health care expenditures were government financed.

In this report, we treat some of the major problems faced in forecasting medical manpower needs. We begin by looking at expectations of the medical care delivery system as well as looking at what it can provide. We then discuss a range of models used to forecast physician requirements, demand for physicians, and the distribution of physicians and examine their implications for policy. Given the nature of the topic, more questions are raised than answers provided.
II. SOME ISSUES IN HEALTH CARE DELIVERY

We begin by considering some broad issues concerning the organization of health care delivery and the allocation of medical resources. We assume that the goal of a medical care system is to improve the health status of the population served. The most immediate question that derives from this assumption is: What is the effect on the health of the serviced population of having more physicians? In other words, what is the health effect of an increase in the physician/population ratio?

But the question contains a question. How should health be measured? (See Berg, 1973.) Health status encompasses mortality, morbidity, disability, restricted activity, bed-days, hospital days, patient anxiety, and the level of satisfaction with the medical care delivery system. There are conceptual difficulties in formulating what is to be measured and there are methodological difficulties in measuring each of the components so that an acceptable, reliable health status index results. Yet, as we argue in more detail below, there is no reasonable alternative to the use of a quantitative output measure of the medical care delivery system. The major policy issues require such measures as inputs into the solution process.

Supposing here were a health index, the next question that arises is: What is the most effective way of improving the health status of a population? Is it through the provision of personal health services? Is it with increased public health services? Would educational programs aimed at altering personal habits with respect to diet, exercise, and smoking be best? A fundamental paradox in medical care obviously does save specific lives, it does lessen disease, and it does prevent or repair disability; yet, it appears that increases in medical care expenditures have little effect on overall mortality rates (Auster et al., 1969; Stewart, 1971). We infer from these observations that simply increasing medical care expenditures will not necessarily increase the number of those patient-physician interactions where medical care is efficacious. Indeed, more of a good thing can be harmful if iatrogenic disease—that due to drug reactions, mistaken therapies, and unnecessary surgery—becomes important. Although there is little evidence on the effect of increased medical expenditures on morbidity rates, there is evidence that increased expenditures on public health and improved personal habits do reduce morbidity. Obviously, one of the major problems here is distributional. It is difficult to get the
right mix and amount of medical care to all groups, especially those most in need of care: the poor and other disadvantaged groups.

If we knew which forms of care were required, we would next ask, What is the most effective way of producing these services? Traditionally, a visit to a physician has meant the involvement of a physician, possibly a secretary, and possibly an office nurse. However, similar services can be produced from a much more varied set of resources. For example, paraprofessionals working under a physician's direction can produce a great part of the traditional patient visit. In some instances, they can produce the entire visit if the purpose is monitoring a chronic condition or some other aspect of treatment that involves a comparatively low level of skill. (For summaries of some of these issues, see Lave, Lave, and Morton, 1971; and Sadler et al., 1972.)

Returning to the notion of adequacy, one can ask about supply strategies. Is there an optimum spatial distribution of physicians? If physicians were equally distributed, those in urban areas would probably be much busier than those in rural areas, because local time access costs are lower. Those physicians in areas with a heavy concentration of older population would be busier because the aged seek more care than the young. What criteria can be used to evaluate the current gross physician distribution figures and those disaggregated by type of specialty and practice organization?

Assuming we can determine what an appropriate supply is, what are the unique problems of remote areas where there is low population density? The answer to this question is inextricably tied to the answers to the earlier questions. An area with a population density of less than one person per square mile, located far from a metropolitan area, is extremely difficult and costly to serve. Nonacute care can be provided by transporting the patient to the physician or vice versa, but providing emergency care is practically impossible unless physicians are fairly near to the patient. Can one expect "sufficient" numbers of physicians to choose to locate in these areas? The problem of servicing remote and highly dispersed populations is probably the most difficult one in health care delivery, and is not likely to be handled by policies that provide financial incentives for relocation.

We can examine demand and supply independently, but then we must ask, What is the interaction between the supply of and the demand for medical
services? To what extent is the demand for medical care manipulable by physicians; that is, to what extent can physicians increase or decrease the demand for their services? The answer to this question is critical because it will influence our evaluation of data on physician distribution.

One of the most important public policy trends has been to lessen or remove price as a mechanism for rationing the allocation of care. This trend has made the monetary cost of care zero or very small for a significant proportion of the population (because of government programs or insurance). As a result, access costs (travel time, waiting time, transportation cost, etc.) have become one of the most important means of rationing care. We can expect this trend to continue since an increased proportion of the population will soon be facing a negligible price for care. Thus, in the future, access costs will increase in importance. These issues are explored in more detail in the next section.

We raise these general issues to provide a framework within which to consider demand and supply models. At the end of Section III we will provide a more structured framework for handling these issues.
III: MODELS USED TO FORECAST PHYSICIAN REQUIREMENTS

INTRODUCTION

What will be the demand for, the need for, and the supply of physicians at a given time and place? What is the gap (if any) between estimated requirements and estimated supplies? What actions, if any, does a gap suggest?

These are central questions in all efforts to plan and forecast health manpower. But before we ask how to determine demand, need, and supply, we must reexamine the meaning of these terms. In particular, the concepts of need and demand are quite different and are often confused. Distinguishing between them is important because manpower requirements that have been determined from forecasts based on needs have been significantly different from those that have been determined from forecasts based on demand.

A population's need for medical services may be defined as "that quantity of medical services which expert medical opinion believes ought to be consumed over a relevant time period in order for its members to remain or become as 'healthy' as is permitted by existing medical knowledge" (Jeffers et al., 1971). Need is thus a "normative" concept and is identified with the amount of preventive care medical professionals believe the population should have, as well as the amount of care they believe will bring the best medical knowledge to bear on the population's illness problems. These standards are flexible, and as Hiestand (1966) has pointed out: "Such standards will always advance in front of what can exist in fact. In a progressive, increasingly affluent society this is perfectly reasonable."

A related concept is wants. A population's wants for medical services may be defined as "that quantity of medical services which its members feel they ought to consume (at zero price, zero lost wages, zero waiting time, zero access constraints, etc.) over a relevant time period based on their own psychic perceptions of their health needs" (Jeffers et al., 1971). Thus, wants are quite distinct from needs and vary among different groups. Some care that may be deemed "needed" by professionals may not be "wanted" by the population; preventive medical care, especially
preventive dental care, are examples. Other care, such as visits for the common cold, may be "wanted" by the lay population, but not deemed "needed" by physicians.

Demand is yet another related concept. The demand for medical services is defined as the "multivariate functional relationship between the quantities of medical services that its members desire to consume over a relevant time period at given levels of prices of goods and services, financial resources, size and psychological wants of the population as reflected by consumer tastes and preferences for (all) goods and services" (Jeffers et al., 1971). Thus the demand for medical care will depend on the underlying health status, perceptions of the efficacy of medical care, and the cost of medical care. "Demand" will be less than "need" because seeking care involves out-of-pocket cost, travel and waiting time, lost wages, discomfort, and emotional or psychic costs.

There are, therefore, important differences between need and demand: A person may not demand needed medical care because it is not perceived to be effective (not wanted or demanded at a zero price) or because, relative to the desirability of other goods and services (which compete for the individual's scarce money, income, and time), it is not valued highly. In addition, price, waiting time, travel cost, or psychic costs may be so high that the individual demands no care, even though physicians believe it is needed and the individual wants it.

We consider next some of the approaches developed for forecasting manpower requirements. We indicate the strengths and weaknesses of each approach. It should be clear that this analysis is closely related to the analysis of the concept of "physician shortage," which will be discussed in Section V. Note that here we are considering how to forecast future physician requirements; we are not concerned with the precision of the forecasts. This issue will be discussed in Section VI.

NORMATIVE APPROACHES TO FORECASTING MANPOWER REQUIREMENTS

Approaches Based on Professionally Defined Standards—"Needs"

The classic approach to developing manpower estimates based on professional standards or estimates of a population's medical needs was developed by Lee and Jones (1933). The approach consists of four steps:
(1) determining the frequency of occurrence of illness by type in a population; (2) polling experts to determine the amount of services required to diagnose and treat each illness type; (3) estimating the average number of services rendered per hour by a provider; and (4) securing professional opinion on the average number of hours that a provider spends per year in caring for patients. Applying their method, Lee and Jones estimated that the need for individual preventive services and the need for the diagnosis and treatment of diseases and defects—all medical services—required 135 physicians per 100,000 population instead of the 126 per 100,000 existing when the study was performed.\footnote{Lee and Jones were aware of the difference between need and demand. They noted that people were not aware of the need for preventive care. In addition, they pointed to the fact that there was a wide distribution in the physician/population ratios across the states and emphasized that if the supply of physicians were increased the new doctors would probably locate in doctor "surplus" areas. They stressed the need for education and for changes in the way medical care was financed and organized.}

The Lee and Jones study was not replicated until the study by Schonfeld et al. (1972). They interviewed practicing internists and pediatricians and determined what these professionals thought constituted good primary care. They then collected morbidity data from the National Health Survey, and, assuming that pediatricians worked 2227 hours and internists 2198 hours per year, they estimated that 133 physicians per 100,000 were required to give good primary care (not including dental care, mental care, obstetric care, and routine physical examinations for adults). This estimate implies a substantially greater gap between need and supply than that of Lee and Jones since it reflects only a portion of primary care, whereas the Lee and Jones estimate was for total medical care.

Even if we accept the concept of need as the relevant criterion, these estimates suffer from many drawbacks. We detail some of these here.

(1) The standards used represent only mean estimates. Neither Lee and Jones nor Schonfeld et al. indicated the range of appropriate treatment patterns identified by physicians. If there is little agreement among professionals, so that the variation in individual estimates is large, then the need concept is not very meaningful for policy purposes.

Another difficulty derives from the fact that the approach does not
take account of substitution for physician services in the delivery of primary care. Although independent "practice" for paramedics and nurse practitioners is still being debated, few would argue that they cannot perform, under supervision, many of the services currently performed by physicians (Lave, Lave, and Morton, 1971). It would thus appear, as Huebscher (1972) notes, that the number of physicians required to render primary medical care is "fluid."

(2) This approach does not include an evaluation of the health outcomes implied by the standards. How do morbidity and mortality rates compare between groups receiving care at the professional standards level and groups receiving much less care? Is there a noticeable difference and can it be attributed to the level of care? Evidence indicates that the relationship is far from straightforward. For example, the Kaiser Health Plan (which is often cited as an ideal health plan where high quality medical care is believed to be provided) has less than half the number of "needed" primary care physicians (34 internists, 10 generalists, and 16 pediatricians) available per 100,000 subscribers (Somers, 1971). These questions must be considered if the standards approach is to be taken seriously.

(3) The approach is excessively narrow because it regards needs strictly from a professional viewpoint. Do people demand this much medical care? It is not altogether clear that, under the best circumstances, patients behave as physicians would have them. For example, Jacobs (1972), in criticizing this work, noted that in his community of 200,000 there were 122 primary care physicians (less than half the standard). He pointed out that although these doctors provided more inclusive care (for example, routine medical examination) and that the community was very wealthy, the physicians were underemployed. Should the patients be "educated" to demand more care? Surely that depends on the answer given to the question of the preceding paragraph.

(4) The approach fails to consider alternatives. It emphasizes physician based personal medical services, but what priority should the government place on the provision of personal medical services? If resources were available to provide medical care meeting professional standards, then surely the share of national income going to personal
medical services would increase enormously. Should such an allocation of national income be preferred to one of more recreation, books, housing, and the like? If the goal of public policy is to improve health, we must determine or at least consider whether that goal will be better met by providing enough personal medical services to meet professionally determined standards or by the provision of other services such as nutritional programs or education. ¹

These criticisms are quite general and apply equally well to other studies where the estimates of need are based on professional standards. Professional standards and, indeed, all normative approaches should be used for manpower planning purposes only after careful evaluation of alternative approaches.

Approaches Based on Evidence in Comprehensive Prepaid Group Practice

A number of observers have argued that the observed demand for manpower in specific existing prepaid group practice plans is the best guide for general manpower planning. They have argued that this mode of practice is ideal (supporters of prepaid medical practice often ignore the effect of access costs). It is presumed that the plan provides members with all needed medical care and that this care is of high quality. However, we doubt that the staffing ratios observed in these groups could serve to forecast the need for specialist or primary care manpower.

In Table 1 we show the distribution of physicians by specialty averaged across six prepaid group practices. In interpreting the data presented in the table, one should note that the numerator includes only active physicians engaged primarily in clinical care and that the average masks a considerable amount of variation.

The Kaiser System, for example, is composed of numerous different local plans. The number of physicians per 100,000 members varies across

¹Other studies following the Lee and Jones method have been reported. Daitz (1965) proceeds by estimating that 74 million people had a chronic condition in 1960; of those, 10 million people had an incipient or manifest functional impairment associated with chronic disease, injury, or congenital defects that require medical care to prevent further deterioration of functional capacity or to restore functional capacity. He then assumes that each of these 10 million people requires a minimum of 40 hours of professional care services per year, of which five should be physician time. This leads to an estimated need of 25,000 physicians to care for chronically ill patients. Knowles (1969) reports on responses
TABLE 1

AVERAGE NUMBER OF PHYSICIANS PER 100,000 POPULATION SERVED IN SIX MEDICAL GROUPS PROVIDING PREPAID MEDICAL SERVICES, BY SPECIALTY

<table>
<thead>
<tr>
<th>Specialty</th>
<th>Average Number of Physicians</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Median</td>
<td></td>
</tr>
<tr>
<td>Internal medicine</td>
<td>45.2</td>
<td>44.9</td>
<td></td>
</tr>
<tr>
<td>Allergy</td>
<td>1.6</td>
<td>1.4</td>
<td></td>
</tr>
<tr>
<td>Dermatology</td>
<td>2.8</td>
<td>2.5</td>
<td></td>
</tr>
<tr>
<td>Pediatrics</td>
<td>18.0</td>
<td>15.8</td>
<td></td>
</tr>
<tr>
<td>Obstetrics</td>
<td>9.1</td>
<td>8.0</td>
<td></td>
</tr>
<tr>
<td>Orthopedics</td>
<td>3.2</td>
<td>3.0</td>
<td></td>
</tr>
<tr>
<td>Ophthalmology</td>
<td>3.7</td>
<td>3.3</td>
<td></td>
</tr>
<tr>
<td>Otolaryngology</td>
<td>4.6</td>
<td>3.5</td>
<td></td>
</tr>
<tr>
<td>Surgery</td>
<td>6.5</td>
<td>6.7</td>
<td></td>
</tr>
<tr>
<td>Urology</td>
<td>1.9</td>
<td>1.5</td>
<td></td>
</tr>
<tr>
<td>Radiology</td>
<td>4.4</td>
<td>4.0</td>
<td></td>
</tr>
<tr>
<td>Physical medicine&lt;sup&gt;a&lt;/sup&gt;</td>
<td>1.3</td>
<td>1.0</td>
<td></td>
</tr>
<tr>
<td>Anesthesiology&lt;sup&gt;a&lt;/sup&gt;</td>
<td>1.5</td>
<td>1.5</td>
<td></td>
</tr>
<tr>
<td>Pathology&lt;sup&gt;a&lt;/sup&gt;</td>
<td>1.8</td>
<td>1.6</td>
<td></td>
</tr>
<tr>
<td>Neurology&lt;sup&gt;a&lt;/sup&gt;</td>
<td>1.0</td>
<td>1.0</td>
<td></td>
</tr>
<tr>
<td>Psychiatry</td>
<td>2.8</td>
<td>1.5</td>
<td></td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td><strong>109.4</strong></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>


<sup>a</sup>Physical medicine based on three groups; anesthesiology based on two groups; pathology and neurology based on four groups. These services are provided in the remaining groups in other ways.

<sup>b</sup>Exclusive of interns and residents in hospitals.
the different plans. There were 96.9 physicians per 100,000 Kaiser members in 1970. This ranged from 76 in the Colorado Plan to 105 in the Northern California Plan. In plans with over 100,000 members only, the range was from 82 to 105 (Williams, 1971). Before one uses estimates derived from such prepaid group practice programs, several questions must be answered. For example: What factors generated the between-plan manpower ratio differences? Are the populations served different? Is more use made of outside physicians in some plans? Are there more paramedical personnel in some plans? Are the health levels of the groups different? These questions have not been answered, but the variability across plans is striking. This variability is strong evidence that fixed manpower ratios make little sense.

Even if these questions are answered, it is still not clear whether these prepaid comprehensive group ratios should be used in determining manpower requirements for the nation. The data are of doubtful relevance and are likely to underestimate the physicians needed for the following reasons:

(1) The plans in the Kaiser System are tightly administered. The total number of physicians is administratively determined and does not necessarily represent a physician/population ratio that would be observed in fee-for-service practice. The number of patients to be seen per hour and the length of the work week are defined; physicians are closely monitored for such aspects of care as ordering lab tests and hospitalization of patients; patients must often wait substantial periods to have a scheduled visit with their physician.

He received from letters sent to the executive secretaries of the various specialty boards to determine what they thought the manpower needs were in their respective specialty. One executive secretary assumed that there would be one operation per 13 people per year and that the annual caseload of an anesthesiologist could be 800; he estimated that 37,000 anesthesiologists were needed. There are only 7,011 in practice. The estimate of needed anesthesiologists is based on what seems to be a small case-load (about 4 operations per day for a 200 day work year) and it neglects the fact that many anesthetics are given by nurse anesthetists. In addition since Americans have many more operations (per person per year) than western Europeans (or even Americans enrolled in prepaid group practices) one may also want to question the assumed rate of surgical procedures.
(2) The distribution of physicians by specialty type is determined administratively. Supposedly, intensive peer review is practiced, and referral and hospitalization rates are closely monitored resulting in lower rates than in fee-for-service practice. Without the organization of the prepaid comprehensive plan, it is hard to see why their utilization patterns would generalize. Note also that this kind of medical system represents a type of practice that currently serves only 5 percent of the population. The members themselves selected this type of system (all had a fee-for-service alternative) and are not a representative subsample of the national population. They tend to be actively employed, middle class individuals whose level of income and knowledge leads to better health, less need for prolonged hospitalization, and fewer complicated illnesses. In 1970, for example, about 4.5 percent of the members of the Kaiser plans were over 65 and 1.5 percent were on Medicaid. To put these figures in perspective, in 1970, 9.8 percent of the United States population was over 65 and about 10 percent had incomes below the poverty level.

(3) At least in the past, Kaiser patients found incentives quite different from those of other patients; monetary costs were low and access costs could be substantial. Indeed, many patients sought care outside the system at a substantial monetary cost rather than accept the nonmonetary costs within the system. Subtle rationing is associated with centralized clinic ambulatory practice, which leads to greater travel time and cost. A final factor is the emphasis put on educating subscribers to recognize disease that is not helped by medical care and to improve personal habits, such as stopping smoking, getting exercise, and avoiding obesity.

Although the data from the Kaiser System do provide a minimum reference point, they are unlikely to be of direct relevance for planning in the absence of significant structural change in the nation's medical care delivery system.

MANPOWER PLANNING USING THE RATIO APPROACH

There is considerable variation in the ratio of physicians to population among various regions. In Table 2, the 1970 distribution
of M.D.'s by region is presented. The Northeast and West have proportionately more physicians per capita than the South and Northcentral regions; the variation among states is much greater than that among regions. For example, in 1970 for the country as a whole, there were 171 physicians per 100,000 population with a range of 83 per 100,000 in Mississippi to 238 per 100,000 in New York. The District of Columbia had 385 physicians per 100,000. There is some indication that these regional differences are becoming larger. In 1950, Mississippi had 66 physicians per 100,000 and New York had 201; thus, between 1950 and 1970, the gap between the states with the highest and lowest number of physicians per 100,000 rose from 135 to 155. There is also some evidence that these trends favor the coastal states at the expense of those that are more inland.

Within states there are great differences between rural and urban physician/population ratios. In 1970, there were 173 physicians per 100,000 in the urban areas but only 80 physicians per 100,000 in rural areas. Within the state of California, physician/population ratios across the planning areas ranged from 220 per 100,000 in the San Francisco Bay area to 99 in Superior, California. Even within urban areas, the variability in the local supply of physicians is great. Physicians are usually concentrated in only a few census tracts, while most census tracts have none.

We explore spatial distribution of physician manpower more extensively in Section IV. Here we consider an approach to planning for medical manpower that makes use of physician/population ratios. In it, some present physician/population ratio is considered a minimum requirement for the future. This approach has a number of variations involving the selection of the ratios:

A. The minimum required physician/population ratio at some future time is taken as the average ratio now in the United States.

B. The minimum required physician/population ratio at some future time is taken as that now in the states with the highest physician/population ratios (the criterion ratio).

C. In projecting minimum requirements, the current physician/population ratio is projected for various socio-demographic groups of the population in a place where they are "adequately" served. To
TABLE 2

NONFEDERAL PHYSICIANS BY CENSUS REGION, DIVISION, AND ACTIVITY, DECEMBER 31, 1970

<table>
<thead>
<tr>
<th>Census Region Division</th>
<th>Total Physicians(^a)</th>
<th>Patient Care</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Number</td>
<td>Percent</td>
</tr>
<tr>
<td>Total Nonfederal</td>
<td>281,344</td>
<td>100.0</td>
</tr>
<tr>
<td>Northeast</td>
<td>87,641</td>
<td>31.2</td>
</tr>
<tr>
<td>New England</td>
<td>20,391</td>
<td>7.2</td>
</tr>
<tr>
<td>Middle Atlantic</td>
<td>67,250</td>
<td>23.9</td>
</tr>
<tr>
<td>North Central</td>
<td>66,993</td>
<td>23.8</td>
</tr>
<tr>
<td>East North Central</td>
<td>48,162</td>
<td>17.1</td>
</tr>
<tr>
<td>West North Central</td>
<td>18,831</td>
<td>6.7</td>
</tr>
<tr>
<td>South</td>
<td>70,178</td>
<td>24.9</td>
</tr>
<tr>
<td>South Atlantic</td>
<td>37,560</td>
<td>13.4</td>
</tr>
<tr>
<td>East South Central</td>
<td>12,155</td>
<td>4.3</td>
</tr>
<tr>
<td>West South Central</td>
<td>20,463</td>
<td>7.3</td>
</tr>
<tr>
<td>West</td>
<td>54,043</td>
<td>19.2</td>
</tr>
<tr>
<td>Mountain</td>
<td>10,368</td>
<td>3.7</td>
</tr>
<tr>
<td>Pacific</td>
<td>43,675</td>
<td>15.5</td>
</tr>
<tr>
<td>Possessions</td>
<td>2,489</td>
<td>0.9</td>
</tr>
</tbody>
</table>


\(^a\) Excludes 19,621 inactive and 358 not classified.

\(^b\) Percentages may not add due to rounding.
determine the number of physicians required, it is then necessary to forecast the expected population growth and the future demographic characteristics of the population. For example, if the aged need more physicians and if the mean age in the population is rising, then to ensure continued access to physicians similar to that of today, the overall physician/population ratio will have to rise.

Ignoring for the moment the problems that can arise in forecasting, we address the strengths and weaknesses of forecasting requirements on the basis of existing physician/population ratios.

(1) Of the planning techniques, ratio approach A seems to require the least information. Projections that incorporate expected changes in the demographic composition of the population (method C) will provide more reliable insight than projections based simply on total population.

(2) There is an implicit, but untested, assumption that the base ratio is adequate and needed; that is, a reduction in the ratio would be adverse. It is often assumed that the more physicians, the better. However, there is evidence that adding physicians to an area may simply lead to unnecessary medical care or to underemployment of physicians. Thus, the criterion ratio approach (method B) is likely to set future requirements too high.

(3) In the absence of structural changes in the delivery system, the ratio approach will provide reasonable forecasts of demand over the short term. Changes in organization, medical technology, immunology, treatment, financing and payment mechanisms, or mode of practice, for example, are likely to invalidate manpower projections based on current utilization patterns.

(4) Using national ratios hides a great deal—for example, variability in physician distribution by state, county, and type of practice. The wide variation in local physician/population ratios is shown in Table 2. Suppose, for example, that enough physicians are trained to raise the United States physician/population ratio to the level of the highest 10 states (as of 1973). There is little reason to believe that these newly produced physicians will choose to concentrate in states with few physicians, or in areas within states that have the lowest ratios. Indeed,
we would predict that the difference between areas would increase as a result of training these additional physicians, as it did between 1950 and 1970. However, all states had more physicians per capita in 1970 than in 1950; thus, the poorer states were better off in absolute terms, even though they were worse off in relative terms.

(5) If the numerator is to indicate the availability of physician services, it should be derived with respect to full time equivalent active physicians engaged in providing clinical services as opposed to teaching, research, public health, and other activities, and it should take account of physicians who work part time.

(6) The ratio approach ignores all changes in physician productivity and assumes that physicians are used in fixed proportions in delivering medical services. Existing data indicate that the way in which physicians allocate their services to patients is predictable; for example, more tasks are delegated to allied health personnel when the physician/population ratio is low (Riddick et al., 1971). The Volume of services provided by physicians is not fixed and physician productivity has been increasing over time. Klarman (1969) presents a number of estimates of increases in physician productivity; the mean estimated increase is about 3 percent per year. This approach takes no account of changes in medical technology and public health practice or the introduction of new health manpower.

In all fairness, it should be pointed out that these criticisms are acknowledged by researchers who have used this approach. It should also be pointed out that without a good deal of clairvoyance it is impossible to forecast some of these trends. However, these arguments suggest that the system is more fluid than the ratio approach implies and that policies should encourage flexibility, not fixed ratios.

ECONOMIC MODELS

The question in development of an economic model of demand is: What will utilization (use of physicians' services or other aspects of medical care) be at some future time? This question is approached by estimating a demand function, a relationship between the amount of care sought and
its cost, access constraints, and other factors. To predict the future level of services that will be sought, the future value of each factor in the demand function is estimated and the estimated parameters are used to predict future utilization.

Economists have tended to neglect access costs, which are becoming more important than direct payment to the provider. Acton (1973a) and Richardson (1971) have investigated some of the effects of access costs. Costs due to scarcities in rural areas were investigated by Marshall et al. (1969). Their finding that dissatisfaction over travel time to get to a physician in rural Kansas was great and was likely to have adverse effects on patient behavior is corroborated in a study by Weiss and Greenlick (1970), who analyzed the behavior of urban and suburban patients in the Kaiser program. However, Weiss and Greenlick noticed that distance did not have a constant effect on all types of patients. Instead, they detected an interaction effect, which led them to conclude that patients in upper socioeconomic status (SES) groups were little affected by distance, while lower SES groups were led either to delay seeking care until symptoms became quite severe or to substitute emergency room encounters for routine appointments with medical staff. Such findings have been reported elsewhere (Lave and Leinhardt, 1972) and have been incorporated into most current theories of patient illness behavior (Shannon et al., 1969).

The demand for medical care (and physician services) is derived from the demand for health (Grossman, 1972a and 1972b). Medical care is a service that can be bought to improve or maintain health "stock." The demand for medical care is thus dependent upon an individual's underlying health status, perception of the efficacy of medical care, and cost of getting medical care, where cost is a vector consisting of time costs, money costs, and psychological costs. To determine the demand relationship, this theoretical formulation must be made specific and then estimated with empirical data.

Investigators may use age as a surrogate for underlying health status, education as a measure of awareness, reported money income as a measure of both earned and unearned income, sex as another indication of the underlying health status, distance to facilities as a measure of
time cost, and insurance and Medicaid coverage as a measure of the difference between actual price and published price to the individual. Demand functions have been estimated using individuals, groups of individuals, and states as the basic unit of observation.\(^1\) Single equation demand models have been estimated as have general systems (with demand as one part).

Economists have focused their attention on determining the income and price elasticities of the demand for medical care and physician services. Income elasticity is defined as the percentage change in quantity bought due to a 1 percent change in income, and price elasticity is defined as the percentage change in quantity bought due to a 1 percent change in price. The former is a measure of the responsiveness of demand to a change in income, and the latter is a measure of responsiveness of demand to a change in price.

Empirically estimated measures of income elasticity have resulted in varied results, partly because different populations were used. Table 3 presents several estimates of income elasticity.

Economists distinguish between earnings and wealth in estimating income elasticities. Separating earned income from unearned income is particularly important where the time price is high, such as in medical care. As earnings rise, other things equal, the value of time rises, and it is assumed that people are willing to substitute money for time. This may explain why employed men and women are more alike in their demands for care than employed women and homemakers, income level held constant (Newhouse and Phelps, 1972).

Price elasticities have also been measured and been found to be small. The estimates range from \(-.19\) to \(-.36\) (Fuchs and Kramer, 1972). Phelps and Newhouse (forthcoming) argue the price elasticity is approximately \(-.1\). Some recent attention has been directed to nonprice costs and their effects on services demanded. Both waiting time and travel time are shown to have high elasticities (Acton, 1973a and 1973b; Richardson, 1971; Weiss et al., 1971).

These results indicate that as the monetary price (to consumers) of

\(^1\)The ratio model that projects current utilization ratios into the future keeping them constant for each socioeconomic demographic group (an approach explored by Fein, 1967) is a variant of this approach.
medical care falls and as income rises, the quantity of medical services demanded will also increase. As access costs to medical care drop, demand will also rise. However, if the time costs of medical care rise, the number of services demanded will fall, particularly for those with a high opportunity cost for time.

Table 3

SEVERAL MEASURES OF INCOME ELASTICITY OF DEMAND FOR PHYSICIAN EXPENDITURES OR SERVICES

<table>
<thead>
<tr>
<th>Source</th>
<th>Income Elasticity</th>
</tr>
</thead>
<tbody>
<tr>
<td>R. Rein (1967) (visits)</td>
<td>.21</td>
</tr>
<tr>
<td>P. Feldstein (1964)</td>
<td></td>
</tr>
<tr>
<td>a. visits</td>
<td>.62</td>
</tr>
<tr>
<td>b. expenditures</td>
<td>.56</td>
</tr>
<tr>
<td>Gorham (1967) (expenditures)</td>
<td>.33</td>
</tr>
<tr>
<td>Andersen and Benham (1970)</td>
<td></td>
</tr>
<tr>
<td>a. visits (observed income)</td>
<td>n.v.</td>
</tr>
<tr>
<td>visits (permanent income)</td>
<td>.01</td>
</tr>
<tr>
<td>b. expenditures (observed income)</td>
<td>.41</td>
</tr>
<tr>
<td>expenditures (permanent income)</td>
<td>.63</td>
</tr>
<tr>
<td>Fuchs and Kramer (1972) (visits)</td>
<td>.04 - .57</td>
</tr>
</tbody>
</table>

a This table is adapted from Klarman (1969).

Some investigators have argued (Ginzberg, 1971; Fuchs and Kramer, 1972; and Evans, 1973) that the demand curve may be affected by the supply of physicians (Stevens and Brown, 1971). In this conceptualization the physicians themselves are thought to influence the number of visits people make by suggesting patterns of care (which include revisits) through follow-up recommendations, and by lobbying for the construction of medical institutions. The crux of the physician induced demand argument is that the pattern of recommended care may vary with the physician
load. That is, in areas with many physicians, physicians can maintain their income by recommending more visits, which may be of marginal medical efficacy.

Some major work on the demand for and supply of physicians has been done by Fuchs and Kramer (1972). They estimated a five equation model using state data that contained a demand function (number of visits per capita), a supply function (number of doctors per 100,000), an output per physician function (number of visits per thousand), an insurance benefits function, and two identities: the quantity demanded of physician services equals the quantity supplied, and one defining net price. Since it is a simultaneous system, two stage least squares was used to estimate the parameters.

Fuchs and Kramer (1972) hypothesized that one of the factors affecting the number of visits demanded per capita was the number of M.D.s per capita. This variable was included because it was argued that physicians were able to generate a demand for their services without lowering price. Fuchs and Kramer (1972) suggest that when physicians are abundant, they may order care that is not medically indicated (unnecessary surgery) or of marginal important (numerous post-operative visits, follow-up visits, or overzealous well-baby care); when physicians are scarce, patients may lower their expectations and handle minor complaints themselves. They also suggested that when physicians are plentiful, the nonmonetary costs of care (waiting time, time to get an appointment, and travel costs) are likely to be lower. In the regression results, the M.D. variable (physicians per capita) was the one with the highest elasticity and with the highest level of significance, presumably indicating that the supply of physicians had the most influence on the demand for care.

Although the number of M.D.s per capita may be a surrogate for nonmonetary costs, Fuchs and Kramer believe that the importance of the variable stems from physicians' ability to control demand. They quote Ganzberg (1969):

The supply of medical resources has thus far effectively generated its own demand. Much unnecessary surgery continues to be performed...There is substantial over-doctoring for a host of diseases, including, in particular,
infections of the upper respiratory tract...[Physicians] usually have wide margins of discretion about whether to recommend that a patient return to the office for one or more follow-up visits.

It could be argued that Fuchs and Kramer have uncovered a simultaneous equation problem rather than physician-induced demand. If physicians moved their practice to places where most care was needed, the data would show a close association between supply and demand. Although many of the characteristics expected to lead to increased need are included in the analysis, the possibility remains that the result reflects the altruistic nature of physicians rather than their artificially induced demand.

Results similar to those obtained by Fuchs and Kramer were found in a study in Canada. Examining physicians in British Columbia, Evans (1973) argued that a strong case could be made for supplier-induced demand. These results are open to question (and have been questioned). Nonetheless, they do have very important implications for manpower planning.

SYSTEMS MODELS OF HEALTH CARE

A number of theoretical systems models of health care delivery have been proposed, but only two have been fully specified and estimated. The first is actually a series of models developed by Martin Feldstein (1970, 1971a, 1971b). These are econometric models with about half a dozen equations (and endogenous variables). For example, in his analysis of the effects of Medicare, the endogenous variables are the proportion of enrollees with supplementary insurance, the hospital admissions rate, the rate of extended care admissions per hospital admission, and the hospital insurance benefits per hospital episode. These variables are modeled as being determined by demographic variables, such as the proportions of white enrollees, male enrollees, those living in large cities, the ratio of enrollees over 75 to total enrollees, per-capita income, and other variables including the proportion of the population under 65 with health insurance, the proportion of enrollees for which the state government declines to pay premiums (and who are indigent), short term beds per capita, private physicians per capita, per capita state expenditure for the aged on health, the number of months the state had participated in Medicaid,
extended care beds per capita, and hospital cost per patient day. The equations were estimated by an instrumental variables technique (similar to two stage least squares) on observations for each state. Since the model was designed to explain the variation among states in the endogenous variables, it has little to do with the sort of behavior one would expect to observe in individual patients, physicians, and hospitals. In this sense, it is a "macro" model and ignores the details of an individual's behavior, assuming it to be unaffected by the policy variables except as summarized in the macro variables. The other conceptual difficulty is the lack of an output measure, but we return to this difficulty after we discuss the second model.

Yett et al. (1970, 1971, 1972) have developed an elaborate model of the medical care sector that uses more than 100 equations (and endogenous variables) to characterize almost every aspect of the system. There are a number of submodels, concerned with manpower, hospitals, and so on, that are joined by some interaction equations. The rationale given for each of the behavioral equations is that of the decisions to be made by an individual and the factors affecting them. Thus, the Yett et al. model is different in orientation, as well as in complexity, from the Feldstein models, but data are not available to estimate the relations. The best that Yett et al. can do is to use state observations to get most of the variables. Even so, it is impossible to obtain the relevant measures of insurance coverage and other factors. Given the size of the model and the quality of the data, the authors choose to estimate the equations individually, using ordinary least squares. This means that the parameter estimates are inconsistent compared with estimates generated by, for example, two stage least squares, although, given the quality of the data, their choice of an estimating technique is appropriate. Unfortunately, the good intentions of constructing a model that is argued to reflect individual behavior are not realized because of the lack of data for estimating the model.

Yett et al. have attempted to spell out the implications of their model using simulation techniques. Using the parameter estimates gained from ordinary least squares regression of state data, the authors vary one or another parameter and use simulation to explore the resulting
change in the equilibrium of the model. For a model of this size, there is no alternative way of investigating the system influences of a change in a set of parameter values. For example, their manpower planning sub-system uses the forecast demand for services, the production relations for these services, and the equations determining the supply of manpower. The output is then a forecast of the number of physicians that would be required to meet the demand, as well as an estimate of the number of hours per week they would work and their income.

The Feldstein and Yett et al., models share the problem of not having a measure of the output of the medical care sector. They determine the conventional measures of utilization, available manpower, and cost but have no explicit measure of output. One might continue to use utilization and input measures (number of patient visits, inpatient days, and so on) as surrogates for output, but the implications in a model of this sort would be that more and more resources should be put into medical care delivery. Although these models can be helpful in exploring the structure of the medical care system, they are inherently limited by their lack of an output measure. Policy implications must be drawn cautiously from a model when it cannot be known whether increasing the number of patient visits will have a positive or negative effect on health (much less on social welfare). The models do tell us the implications of parameter changes, such as those that might result from a particular type of national health insurance. If one knew whether an increase in one aspect of utilization was good or bad, such models might be used to determine what parameters ought to be altered to achieve a desirable result.

We have constructed a model with these properties (Lave, Lave, and Leinhardt, forthcoming) and are currently attempting to estimate its parameters. Summary inputs and outcomes in the model consist of physician hours, hospital days, physicians visits, total cost, and several health status measures. The underlying health status of the population and an individual's health status at any point is the primary factor determining whether care will be sought or prescribed treatment compliance will occur. Since health status is treated explicitly in the model, there is no alternative to specifying spontaneous rates of improvement or deterioration
in health, and the efficacy of some contact with the health care system in improving health status. We regard the model as giving many insights, but we must stress the obvious difficulties in getting health status measures and in estimating how these change spontaneously or with medical care.

SUMMARY AND CRITIQUE OF THE MODELS

The ratio model has the advantage of being simple and will predict utilization accurately insofar as the underlying conditions determining demand and supply do not change (no major changes in medical knowledge, organization, or financing of medical care). However, since the ratios reflect the current system, their use guarantees a future system where utilization is similar to the current one. In contrast, the Lee and Jones (1933) approach is more of an attempt to get at ideal requirements. Its principal problem is that the actual level of medical care demanded is far less than that predicted by the model (since not everyone who needs care seeks it, nor do people adhere to their prescribed regimen). Thus, manpower policies based on the ratio approach will tend to preserve the current system, while policies based on the Lee and Jones approach would lead to an oversupply of physicians. Both approaches ignore the substitution possibilities in producing care and increasing productivity.

The simple economic models go far in allowing one to gain insight into the factors that affect physician utilization. They allow one to predict what effect changing socioeconomic or demographic characteristics of the population will have on the demand for medical care and what the effects will be of increasing population or the amount of insurance. However, many of the crucial variables, such as travel time and waiting time, are necessarily the result of interactions between supply and demand. Since these models look only at the demand side of the picture, they cannot determine the level of these access variables and so are conceptually incomplete (although one might guess at future values of these parameters).

Two models of medical care delivery that look at the simultaneous interaction of supply and demand have been described briefly. However, they are unsatisfactory in two important ways. First, they were estimated
using aggregate (state level) observations. Second, they had no output measures associated with them. Thus, one knows only that by manipulating the system one can change the number of patient visits or hospital visits, but one cannot know whether these changes increase or decrease the public welfare.

To construct a model capable of predicting demand after a structural change (such as that represented by national health insurance or reorganization into health maintenance organizations), to take account of the simultaneity of supply and demand, and to estimate the social welfare implications of changes, one must have a model of the health care system like the two presented but having the additional property that output is included directly. This is the crucial issue, but, in addition, the system must be estimated using disaggregated data so that it reflects individual consumer and physician preferences while acknowledging that there are other ways of providing care.
IV. MODELS OF THE SUPPLY OF PHYSICIAN SERVICES

Physician services must be defined in terms of the number of hours of service by physician's specialty and location. Thus, one must know the current stock of physicians, changes in the stock due to deaths, retirements, and new graduates, factors affecting choice of specialty, factors affecting the number of hours a physician devotes to medical care, and factors affecting geographical location (by region, by rural versus urban area, and by location within an urban area). While there have been insightful discussion of these issues (see Fein, 1967), there has been little formal modeling and empirical investigation.

In this section we examine research findings that bear on physician supply. These studies have focused predominantly on the maldistribution of physician services. Determining such maldistributions, measuring their extent, and studying the factors associated with them are central issues since many government agencies have reacted to perceived local scarcities by elaborate, and often expensive, intervention in the local delivery system. Unfortunately, this research area is filled with contradictory results; thus, one must look carefully at the underlying data:base and analytical methods to resolve the contradictions. We briefly review the major analytic approaches and then present a summary of findings concerning factors that appear to influence the location choices of physicians. Given the contradictory nature of the findings, we have no alternative to an in-depth review of the data and methods.

APPROACHES TO STUDYING PHYSICIAN SUPPLY

Three techniques have been most commonly used to study maldistribution and the factors causing it. By far the most common involves a simple tabular comparison of physician/population ratios. Variability across the units is often presumed to be sufficient evidence of a maldistribution of physicians. The extent of this maldistribution is identified with the range of the ratios. Thus, DeVise (1973) argues that coastal regions have a comparative advantage over inland regions, and Fahs et al. (1968) argue that the central states lose medical graduates to the western states. Comparisons of this sort inevitably show that the more heavily urbanized states have more physicians than the rural states.
and, within the states, urban counties have more physicians than rural counties. However, these findings should be viewed as suggestive but not definitive. The studies themselves suffer from an absence of control for other factors known to be important determinants of physician location. For example, the coastal states with the greatest number of physicians are also highly urbanized states; is the coastal location or the urbanization the true cause of the movement? What aspects of urban or coastal environments attract physicians? What qualities of rural communities repel them? For purposes of policy formulation, such simple comparisons represent only the beginnings of an analysis.

The second most common approach employs multivariate statistical analysis and represents an advance over tabular comparisons. Studies using these methods attempt (at levels of aggregation such as regional, state, county, or census tract) to relate the physician population of a locality to qualities of the locality that are presumed to act as positive or negative inducements to physicians. These studies are often predicated on a well-formulated model of physician spatial behavior. Typically some form of regression is used to estimate the relation. In their classic study, Rimlinger and Steele (1963) used least squares techniques to estimate a linear model relating physician/population ratios (for 200 county groups that constituted the United States) to characteristics of these areas in 1959. Their findings indicated that income, leisure, and mobility were significant explanatory variables; and they argued that such factors are positively associated with urbanization. The power of the multivariate approach is that it has the potential of allowing the investigator to examine the influence of a host of variables simultaneously and to estimate the individual effect of each. Since some of these variables are amenable to policy manipulation, the technique can provide a basis for choosing among alternative policies.

Although this approach is superior to cross tabulations, it has many pitfalls. Ordinary least squares (OLS) may not be an appropriate estimation technique because of multicollinearity or simultaneous influences among the variables. Often the relevant variables are so closely associated in the observed data that it is impossible to distinguish among them. (See, for example, the discussion in Fuchs and Kramer, 1972.) Other problems in the applicability of OLS lie in the assumptions that must be made
concerning the distribution of errors. Thus, estimation techniques such as two stage least squares, ridge, Tobit, and Poisson regression may be more appropriate.

The third approach, and one that is becoming more common, involves surveying physician opinions. Here researchers attempt to ascertain from physicians themselves the factors they regard as most important in the choice of a practice location. This approach has the potential of uncovering factors that may be neglected by analysts who have little personal experience with practice location. Of course, the outcome of such studies may be input to more elaborate econometric analyses of aggregate data. For example, Cooper et al. (1972) report the results of a survey conducted by the AMA in an attempt to ascertain the conditions under which particular physicians decide to locate in rural or urban areas. Such studies focus on differences among physicians, whereas spatial models focus on differences among localities. Both are useful for policy formulation. On one hand, surveys help determine which background qualities of physicians make them better choices for rural locations. On the other hand, spatial models may reveal whether professional amenities lead more physicians to choose a particular locale.

These three approaches are not exhaustive of those used to investigate physician distribution, but they do typify the reported research. In general, the questions that have been phrased are these: Where are the physicians? What are the characteristics of the areas they favor and of those they shun? What are the characteristics of physicians in areas where there are scarcities? The underlying assumption of the research is that by understanding the attractive factors or the personal predispositions of types of physicians, policies can be instituted to modify or equalize the availability of medical care. We proceed to review this literature first in terms of location choice between urban and rural settings and then within the city; we consider the role of the foreign medical graduate; and last, we examine what has been learned regarding the initial locational decisions of recent medical graduates.

In Table 4 we have categorized factors that have been the subject of research and indicated the direction of their apparent influence on physicians. Some effects, such as that of loan forgiveness programs, are debatable. But in the absence of further research the displayed effects seem reasonable. The results emphasize the attractive nature
<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>Relationship</th>
<th>Independent Variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of physicians</td>
<td>+</td>
<td>Per capita income in state</td>
</tr>
<tr>
<td></td>
<td>+</td>
<td>State educational expenditures</td>
</tr>
<tr>
<td></td>
<td>+</td>
<td>Per capita income in county</td>
</tr>
<tr>
<td></td>
<td>+</td>
<td>Construction of hospital in community</td>
</tr>
<tr>
<td></td>
<td>+</td>
<td>Median income in community</td>
</tr>
<tr>
<td></td>
<td>+</td>
<td>Population of area</td>
</tr>
<tr>
<td></td>
<td>+</td>
<td>Physicians' price practices, based on per capita income of area</td>
</tr>
<tr>
<td></td>
<td>+</td>
<td>Failure rate of licensing examination</td>
</tr>
<tr>
<td></td>
<td>+</td>
<td>Physician income in state</td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>Lack of recreational facilities</td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>Construction of hospital in rural county</td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>Cyclic variations in income levels in area</td>
</tr>
<tr>
<td></td>
<td>+</td>
<td>Presence of commercial activity</td>
</tr>
<tr>
<td>Practice in urban areas</td>
<td>+</td>
<td>Graduation from certain medical schools</td>
</tr>
<tr>
<td></td>
<td>+</td>
<td>Graduation from urban medical school</td>
</tr>
<tr>
<td>Practice in rural areas</td>
<td>+</td>
<td>Rural background</td>
</tr>
<tr>
<td></td>
<td>+</td>
<td>Participation in loan forgiveness program</td>
</tr>
<tr>
<td>Practice in same state</td>
<td>+</td>
<td>Internship and residency training in state</td>
</tr>
<tr>
<td>Ability to attract physicians</td>
<td>+</td>
<td>Mobility of community residents</td>
</tr>
<tr>
<td></td>
<td>+</td>
<td>Educational level of population</td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>Percent population in agriculture</td>
</tr>
<tr>
<td>Number of primary care physicians</td>
<td>+</td>
<td>Percent population white</td>
</tr>
<tr>
<td></td>
<td>+</td>
<td>Percent population 0-5 years old and 65+ years old</td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>Inadequate cultural and recreational resources in community</td>
</tr>
<tr>
<td>Number of specialists</td>
<td>+</td>
<td>Educational level of population</td>
</tr>
<tr>
<td></td>
<td>+</td>
<td>Number of supportive institutions</td>
</tr>
<tr>
<td></td>
<td>+</td>
<td>Number of general hospital beds per 1000 population</td>
</tr>
<tr>
<td></td>
<td>+</td>
<td>Medical school in community</td>
</tr>
<tr>
<td></td>
<td>+</td>
<td>Presence of high concentrations of commercial activity</td>
</tr>
<tr>
<td></td>
<td>+</td>
<td>Presence of university medical facility</td>
</tr>
<tr>
<td>Presence of physician</td>
<td>+</td>
<td>Economic growth rate of town</td>
</tr>
</tbody>
</table>

*Adapted from Cooper et al. (1972)*
of urban areas and the physician's personal acquaintance with a particular area or type of area. Although the table summarizes the most relevant research, we briefly consider some of the studies from which it was constructed.

THE RURAL-URBAN CHOICE

The discussion of rural-urban differences in physician manpower is commonplace in state medical journals. Baker et al. (1967), MacQueen (1968), Martin et al. (1968), Matthews (1971), Reas (1972), Royce (1972), and Stine (1968, 1970) represent a small sample of studies in which either surveys of physicians' opinions or tabular comparisons of physicians by county are presented to document urban-rural disparities. This genre of studies can be considered together since they are quite similar in design and findings. Generally, the physician surveys indicate that those who chose to locate in small communities report they did so simply because they and their spouses liked these types of locales. Yett and Sloan (1972) report that physicians tend to set up practice in areas near their place of rearing or medical training. These findings and the associated finding that most small town practitioners were reared in small towns suggest that, of those physicians practicing in small towns, personal life style preferences rather than professional issues predominate. (See also Parker et al., 1969.) For physicians who report that professional considerations predominate, urban locations tend to be most attractive. Such physicians indicate the importance of proximity to colleagues and supporting medical facilities, the ability to have a varied practice, and the possibility of establishing group practice. Heavy workloads are reported by nonurban physicians, while higher pay scales and more readily scheduled activities are reported by urban physicians, as are access to institutional medical facilities and group practice organizations. The surveys also indicate that young physicians are more anxious than older physicians to join group practices and that they are tending more often to choose financially secure and scheduled positions on academic or medical staffs and in other institutional settings. These findings imply that extant urban-rural differences are likely to increase with increases in the supply of physicians in the absence of fundamental changes in medical education.
Multivariate regression studies have addressed the issue of lifestyle variables and urban versus rural location at all levels of data aggregation. In an analysis of state level data, Fuchs and Kramer (1972) used two stage least squares techniques to estimate a model of physician distribution for 33 states in 1966. The most important variables they identified were per capita income, the presence of medical schools, the price of care, and the number of hospital beds. However, the relevant variables were so closely associated that they could not separate the various hypotheses. They concluded by suggesting that, in general, professional convenience and urban "life style" factors were attractive to most physicians. In a follow-up study to their 1963 report, Steele and Rimlinger (1965) reported the results of county level analyses for 1949 and 1959. Although some specific results seemed at odds with their earlier findings, the general results from both studies were similar. Family income was less important as an explanatory variable in 1959 than in 1949, while the amenities the authors associated with an "urban environment" became more important over time.

Marden's (1966) study of physician location in 369 metropolitan counties was particularly important because he grouped physicians as either specialists or general practitioners. The following characteristics of the metropolitan areas were included: educational attainment, age composition, racial composition, and number of hospital beds in the counties. Race and age composition were found to be the most important explanatory variables, with hospital beds important only for the smallest cities. However, Marden's analysis was replicated by Joroff and Navarro (1971) with contradictory results. Examining 299 cities and using 10 independent variables, they did not find race to be important once the presence of a medical school and the population's educational attainment were taken into account. Additional important factors were hospital beds and the population's age distribution.

Further support for the role played by medical facilities or professional amenities is provided in two Ph.D. dissertations using state data. Weiss (1968) reported that physicians were attracted to states in proportion to the number of medical centers they contained; and Sloan (1968) reported that high physician/population ratios were associated with high income, low cyclical economic variability, high capital stock, and high numbers of medical students (a likely surrogate for the medical center
effect detected by Weiss). Finally, Benham et al. (1968) performed analyses on 1930, 1940, 1950, and 1960 levels of physicians and dentists in the states and found that, over time, increasing importance was being placed on high per capita income, population, and medical training facilities.

These results for the effects of population and medical facilities on physician distribution have received widespread support. Income seems not to be the most important factor any longer. For example, Marshall et al. (1971) performed an elaborate analysis using factor analytic techniques to ascertain the common characteristics of Kansas counties with fairly high physician/population ratios. Their method prevented them from quantifying the influence of their explanatory variables. However, they concluded that affluence alone was not sufficient to explain physician location. Population size, they found, was the single most important factor.

Although this finding was inconsistent with an earlier report by Terris and Monk (1956) dealing with conditions in upstate New York, it represents an increasingly popular belief that physician salaries are currently so high that the lure of still higher income influences only a few. Instead, many researchers argue—and this is most likely the gist of the rural-urban disparity—that physicians are now more concerned with cultural, environmental, and professional amenities. These include leisure time (attained by a not overly demanding practice and the effectiveness of nonprice rationing procedures [Sloan et al., 1973a]), medical resources, interesting case variety, income security, colleagues, and other qualities associated with large populations or urban areas. This was, in fact, the conclusion of a study by Parker et al. (1969) who observed that most physicians choosing to practice in small towns were quite distinguishable from the bulk of physicians in not sharing their views about urban and professional qualities. Instead, they were often originally from small towns, had a preference for this lifestyle, and claimed to be unconcerned over urban-rural physician income differentials. Nonetheless, Steinhall and Sloan (1973) argue that income is still important since a profit-maximizing model of physician fee-setting dominates alternative models. But this conclusion does not conflict with individual location choice based on nonincome considerations.
These findings indicate that as the bulk of physicians are socialized through the medical education system, they acquire a set of values that predisposes them to choose an urban location for their practice. Although some physicians, because of earlier, more deeply set personal concerns, may be able to overcome this inclination and choose rural locations, they are unlikely to offset the growing disparity in urban and rural physician supply. Both the increasingly urban nature of the population and the increasing importance placed on hospital and specialty medicine in the medical profession support this conclusion.

INTRA-URBAN LOCATION DECISIONS

The state, metropolitan, and county level studies provide some information on the factors affecting the spatial distribution of physicians in the United States. They also suggest variables that may be relevant to intra-urban spatial research. The research we have reviewed indicates that physicians tend to gravitate toward areas that satisfy their own personal objectives in terms of lifestyle and professional behavior. Income alone does not appear to be an overwhelming consideration. Any simple association between income and physician/population ratio may be due to other factors, such as urban amenities, that are also associated with the high income area. Surprisingly, the racial composition of the population living in an area where the office is located does not appear to be very important. Urbanization, medical specialization (leading to greater dependence among physicians on cooperative activities), dependence on hospital-based technological care, and the rise of staff positions in institutional medicine seem to be the primary processes controlling contemporary spatial distribution of physicians across the nation.

The processes affecting regional distribution seem readily understood, although work is needed to improve precision and determine important interactions. On the local level, however, the operative factors are not as clearcut. Yet understanding spatial distribution at this level has greater immediacy. This issue is an urgent one because of the likelihood that the urban population groups adversely affected by local distributional scarcities are the poor, the aged, or the nonwhite segments of the population. When such groups are faced with adverse conditions, they are likely to develop patterns of behavior accentuating the severity of an illness episode and
frustrating the efficient delivery of therapeutic medical services (Shannon et al., 1969).

The literature bearing on local distribution is large but lacking in analytical sophistication. We will return to this issue below and here detail typical approaches and findings.

Chicago, New York, and Boston have been the sites of extensive investigations. Some other cities have also been studied (Pittsburgh by Kaplan and Leinhardt, 1973; Baltimore by McMillan et al., 1970, and by Fine, 1971; and Buffalo, Syracuse, and Rochester by Terris and Monk, 1956). In what follows, we concentrate on recent studies of the three major sites, detailing others only when relevant.

Chicago has been a traditional site for studies of socioeconomic factors and their role in urban ecology. Lepper et al. (1967) performed their study of physician/population ratios in Chicago under the auspices of a grant to the Chicago Board of Health to examine patterns of medical care in the city's OEO-defined poverty areas. They found physicians to be fairly scarce in these locales. Similar conclusions have been reached by Rees (1967a, 1967b, by DeVise and Dewey (1972) and by Dewey (1973). DeVise and others on the staff of the Chicago Regional Hospital Program have completed numerous reports that echo these findings. In his two studies, Rees (1967a, 1967b) attempted to document what he claimed was a continuing exodus of physicians from the inner city to the surrounding suburbs. But the spatial unit he chose was extremely gross (the city was either broken into three concentric zones or a few neighborhoods) and, consequently, only vague conclusions could be drawn. Thus, although he determined that the numbers of physicians setting up practice in the Chicago suburbs outpaced those starting up in the city, it was impossible to determine exactly where in the city new physicians were going. DeVise and Dewey (1972) report a summary of Dewey's (1973) monograph. The study described in the two reports examined decennial data on physical/population ratios for 1950, 1960, and 1970 and contrasted these with data on four explanatory variables: race, population, retail buying power, and hospitals. The conclusions, drawn after comparing "mean centers" located in the concentric zones and sectors of the city, were that physicians were fleeing the inner city in favor of the more affluent white suburbs; that physicians seek office locations near expensive residential tracts to minimize their own travel time to the office; that socioeconomic levels were the most
influential qualities of a locale; that suburban shopping centers and office buildings have high attractive potential for physicians; and that these processes were aggravating the disparity between high and low income areas over time. But the coarse nature of the zone boundaries in this analysis confuses the influence of many variables, some of which may be important and unrelated to locale, affluence, or racial characteristics.

One of the most extensive analyses of intraurban physician office location was performed by Elesh and Schollaert (1972) using data describing conditions in Chicago around 1960. In this study, classical least squares estimation techniques were used to estimate several log-linear regression models expressing the log of physician frequencies (general practitioners, specialists, and all physicians) as functions of population, commercial activity, hospitals, age distribution, education, and racial composition of Chicago census tracts. The coefficients of these independent variables were then examined for sign, magnitude, and significance to determine whether each variable either attracted or repelled physicians. The results indicated that, in Chicago, physicians avoid predominantly black areas but are attracted to tracts that are highly populated, have more commercial activity, more hospitals, are located in the central business district (CBD), have older populations, are more educated, and have high median incomes. However, the explanatory power of the model was low ($R^2$ never exceeded .4), indicating that important variables influencing physician distribution had not been included in the models.

In a similar though less ambitious study submitted as a B.A. honors thesis at Harvard University, May (1970) performed a series of regression analyses on the distribution of physicians in Brooklyn, New York in 1960. However, in contrast with the Elesh and Schollaert study, May found race was not a significant variable. Using both ratios and counts of physicians at the census tract level, he observed that median tract education was the only important socioeconomic variable and that population size and the number of hospitals were other crucial variables. His findings also contradicted the observation of Piore and Sokal (1968) that in New York City, physicians avoided low income areas because of their evident poverty-stricken character.
This last conclusion was precisely the one drawn by Roemer (1966) in his description of conditions in Los Angeles and, regardless of the publication of seemingly contradictory evidence, it is the viewpoint most often stated in describing the "doctor shortage" or "physician maldistribution" (see the review in Lave and Leinhardt, 1972).

The true complexity of the situation, though, is best illustrated by the work reported by Dorsey (1969) and Robertson (1970) on conditions in Boston, and by Hambleton's (1972) analysis of 15 SMSAs (see also his dissertation, 1971). Dorsey examined data for census tracts in Boston and Brookline for the period 1940-1961. He observed the growth of an increasingly specialized population of physicians (at what he construed to be the expense of general practice). Using income, occupation, and education as a composite indicator of SES, he determined that the changing physician distribution over time had left low SES tracts with few local primary care physicians. However, Robertson drew quite different conclusions in a study of Boston census tracts covering the same period, 1940-1960. He observed that because of a tendency for physicians to cluster together, simply longitudinal analyses are likely to be erroneous. This would be because of the lagged effect caused by clusters of physicians attracting more physicians who seek out such physician clusters. To avoid this problem he used a differential equation model of change originally described by Coleman (1968). The change coefficients in this model were estimated using multiple regression techniques after factor analysis had reduced census data to four factors, chief of which was a factor identified with SES. The results were complex, but they presented a consistent and interpretable pattern. SES was found not to be important for the location of general practitioners, but it was related to the distribution of internists and pediatricians. All three types of physicians tended to concentrate in areas characteristically high in the second factor derived from the census data. This factor was identified with low owner-occupied housing and low median income. Robertson concluded that, in Boston between 1940 and 1960, medical practice had become more specialized and physicians had clustered into a few locales that were characterized by office buildings and proximity to hospitals. This clustering had proceeded with little regard to changes in other local conditions. The result was that although some few areas were well served by locally available primary physicians,
it was generally accidental. In general, because of clustering, physicians were located near only a small portion of the population.

A study by Hambleton (1971) further disputes common assumptions about urban income and race-related physician maldistribution. He used an analysis of variance technique to investigate the relations among race, income, and population of postal zones in 15 SMSAs in 1960. His findings were that the clusterings of physicians in the central business districts of these cities and in other inner zones actually left predominantly poor or nonwhite zones with advantages over other areas. This was especially true for specialists who, he argued, tend to cluster in the central business district more than do general practitioners because they have a city-wide market orientation. Poor and black residents near business districts were consistently found to be near to medical offices, which, it appeared, had been located without regard to the race or income of their immediate neighborhoods. However, Hambleton raised the cautionary note that proximity in no way guarantees utilization. The problem to access, he observed, is not simply one of geographic availability. Hambleton's conclusions were similar to those drawn by Kaplan and Leinhardt (1973) in a multivariate study of physician location in Pittsburgh. Observations were at the census tract level, and physician location was explained in terms of commercial activity, hospital beds, and socioeconomic and demographic composition. Pittsburgh physicians were found to be more concerned with professional amenities than with local socioeconomic or racial conditions.

The importance of professional factors is consistent with evidence from other cities. Physicians were not uniformly distributed across the city but instead are concentrated in places offering professional amenities. Although there is no overt advantage to the middle class in that physicians are not located any more closely to them than to the poor, there are subtle advantages. The middle class can more easily travel the distances, and dealing with the system of specialists and appointments is easier.

THE ROLE OF THE FOREIGN MEDICAL GRADUATE

A large portion of the current physician supply in the United States and the one most amenable to short term policy-induced variation is the group that has received training outside of the United States. Some of these foreign medical graduates (FMGs) are U.S. citizens who sought
training in foreign countries. Most, however, are foreign citizens who have come to the United States for a period of years to improve their training or who wish to immigrate. These FMGs represented approximately 20 percent of the active physicians and about 33 percent of the hospital interns and residents in 1970 (Dublin, 1972). Between 1960 and 1970 the influx of FMGs increased at a faster rate than the domestic production of new physicians. In 1968-70, FMGs were 29 percent of newly licensed physicians. These physicians were trained primarily in underdeveloped countries. Of all physicians admitted as immigrants in 1970, 70 percent (approximately 6,300 physicians) came from Central and South America, Asia, and Africa. Between 1962 and 1971, nearly 29,000 FMGs immigrated to the United States. During the same period, 46,812 FMGs came to the United States under the exchange visitor physician program, and most of them have stayed well beyond the original two year term of the program.

This flow of skilled manpower from underdeveloped areas to the United States poses many important questions for international relations and represents a form of reverse foreign aid that may be quite detrimental to the source countries (see the report of the Macy Conference on the Migration of Medical Manpower, Bowers and Rosenheim, 1970). However, this flow of foreign physicians has come to be seen by United States policymakers as a significant aid in alleviating the presumed shortage in the domestic supply of U.S. trained physicians. Indeed, such thoughts led to 1970 legislation making it fairly simple for an FMG to convert a visitor visa to permanent-resident status.

The question most often raised by medical professionals about FMGs is the adequacy of their training. In 1970 only two out of five FMGs taking an equivalency test passed it on the first try, and only two-thirds of those attempting ever obtain an equivalency certificate. But such tests do not separate language difficulties from training deficiencies. Regardless of the adequacy of this testing procedure in assuring comparability of medical knowledge, there is no systematic procedure assuring that FMGs will adapt to the complex organization of health care delivery or that the effectiveness of care they deliver will not be compromised by cultural or ethnic barriers. Perhaps language and organizational integration programs would ensure more effective utilization of this manpower resource.
A related issue regards the growing dependence of the United States on a physician resource over which U.S. policymakers do not have complete control. Although, by evidence of their increasing numbers of applications, there is currently a large and anxious stock of FMGs seeking entrance to the United States, there is no way of assuring continuity in this supply. The health needs of other nations are certainly as great as those of the United States, and it is unclear how long these nations will permit skilled manpower to be drained off by the incentives of the U.S. market for physicians.

Still a third issue concerns the effect of this increasing supply of FMGs on the general distribution and availability of physicians within the United States. If one assumes that a nonuniform distribution of physicians is detrimental, then, clearly, the initial locational choice decisions of FMGs and their migratory propensities within the United States can work either to alleviate or to aggravate this maldistribution. Although important research on this topic is limited, Margules and Bloch (1969) indicate that, at the state level, FMGs tend to concentrate in locales that are more urbanized and therefore tend to have more physicians than others. By far the most extensive investigation of the effect of FMGs on the spatial character of the U.S. physician supply has been performed by Butter and Schaffner (1971). Addressing the distributional effects of the FMGs, they used 1968 data from the AMA giving physician/population ratios at the state and at the SMSA-rural levels. Their analysis was simple: a comparison of deviations from a uniform distribution of physicians that first excludes FMGs and then includes FMGs. On the basis of these comparisons, they argued that the interstate and urban-rural differences in the physician/population ratios were exacerbated by the addition of FMGs.

FMGs appear to distribute themselves selectively and, although they add to the aggregate physician supply, they increase the disparity between states and within states between urban and rural areas. Several factors may explain this finding. FMGs may exaggerate the locational decision propensities of U.S. trained physicians because of their dependence on institutional support, their desire for cosmopolitan environments, their likely urban backgrounds, and their possible desire to serve their own ethnic groups. Unless urban FMGs are viewed as replacements for U.S. trained physicians who are thus freed to serve rural communities (an
unlikely assumption when demand in urban areas is expanding), they should not be considered as a force arresting the growing rural-urban disparity in physicians.

In a second study Schaffner and Butter (1972) examined the interstate geographic mobility of FMGs. Their objective was to determine whether long term trends in the regional mobility of FMGs would act to equalize the apparent selectivity of their initial locational decisions. Using data from the 1966 and 1968 physician census taken by the AMA, the authors computed FMG in- and out-migration rates for the two year period for 36 states (95.4 percent of all FMGs). As surrogate measures for determining the relative physician shortage in a given state, they used the physician/population ratio, the five year rate of change of physician income, and the state level of physician income (this assumes that physician income is responsive to demand and that relative increases in income indicate increases in demand without concomitant increases in supply -- that is, shortage). They determined that only 8 percent of all FMGs moved between states during the two years and that the mobile FMGs tended to move to only those states that ranked low in terms of their measures of medical manpower shortage. Thus they concluded that the mobility of FMGs added to the disparity between the states. However, these conclusions must be viewed with caution. Only 36 states were included in the study, a period of only two years was analyzed, and the AMA records on FMGs are likely to be incomplete. An adequate study of initial locational propensities of FMGs, their migratory patterns, and the factors influencing their local spatial distribution would require a longer time span, a lower level of aggregation, controls for factors likely to attract or repel FMGs, and data detailing residency location, specialty, and subjective characteristics. Informed planning policy must take into consideration the large and increasing FMG physician population. To the extent that incentives for these physicians can be expected to differ from those acting on American trained physicians, separate analyses are required.

NEW PHYSICIANS

Understanding the locational decisions of physicians can be decomposed into two related issues: (1) the decision to locate and initiate practice
and (2) the decision to move a practice. Presumably, different factors influence these decisions. Physicians making their initial locational decisions will be younger and less experienced, and their families will be at a different stage of development than physicians considering moving a practice from one locale to another. This latter decision is often tied to the issue of changes in type of practice or specialty type (Crawford and McCormack, 1971). Although an important distinction, few analyses of physician location or migration have attempted to distinguish between physicians who have relocated and those who are locating practices for the first time. Clearly, if different concerns motivate these two groups of physicians, specific policies to take advantage of them might be formulated. The work that has distinguished between these two groups has focused on the behavior of new physicians. These studies are important because they reflect the effectiveness of programs that have been instituted during residency to alter the specialty and geographic distribution of physicians. We consider this material next.

Although there are several surveys of medical school graduates (Martin et al., 1968; and Weiskotten et al., 1960), these are usually descriptive studies of the current character of various past graduating classes. They typically fail to distinguish initial from subsequent decisions or to look closely at the behavior of recent graduates.

Sloan and Yett have performed several studies investigating the behavior of recent medical school graduates. Sloan has examined choice of specialty (1970) and practice mode (1973) and Yett and Sloan have modeled spatial distribution (1972). Their findings suggest that earnings differentials do not account for the strong trends away from general practice in favor of careers in specialty medicine. Elasticities based on lifetime earnings coefficients for most specialties were near zero, and, although there was a significant positive effect in some regression results, the effect of the income variable was always small. The number of FMGs in a specialty was significant in several estimates, but it always had a negative coefficient. In other words, medical students are attracted to those specialties that have high lifetime incomes and few FMGs, but neither effect is very great. Indeed, the negative FMG coefficient may indicate that FMGs are allocated into residual categories—that is, they fill up positions domestic medical graduates shun.
On choice of practice, Sloan's (1973) report is merely descriptive, presenting the results of a survey carried out by the Hospital Physician. Partnerships and groups were clearly preferred by the residents sampled in the study, and even academic medicine was preferred to solo practice. The survey indicated that although practice mode decisions are made fairly late in the medical education process, the decision to engage in academic medicine seems to be made while the physician is still a medical student. Many residents report that financial considerations are important in their choice of practice mode, although it is likely that the net income of non-solo practice is lower than the net income of solo practice.

In an extensive investigation of a 1966 survey carried out by Medical Economics, Incorporated, Yett and Sloan studied the effect of several variables on the state-level spatial distribution of new physicians. Variables describing the following general factors were included in their analysis: previous attachment to the state, income, population growth, barriers to entry (using licensure failure rate as a proxy), opportunities for professional development, general environmental conditions, and level of effort required to establish a satisfactory practice. Their results indicated that previous attachment to a state (through birth or attendance at a medical school, internship, or residency program) was an attractive factor. Income levels and environmental conditions were significant, too. But the only action a state could take to increase its physician supply (besides attachment) that is predicted to have a significant effect would be to lower the failure rate on state licensure examinations. The environmental and income factors, when adverse, might be offset, they suggest, by financial incentives; but other research reports suggest this is unlikely to be successful.

SUMMARY OF THE PLANNING IMPLICATIONS OF CURRENT SUPPLY

The supply of physician manpower seems best explained by behavioral theories that emphasize decisionmaking by individual physicians in which the role of pecuniary incentives has diminished over time. Although fee setting by physicians may involve profit-maximizing behavior, locational and specialty choices do not. Instead, professional and personal amenities
and conveniences seem most important, with income security rather than income maximization gaining in importance. These findings suggest that physicians will be more, rather than less, ready to accept staff positions in institutional facilities and they will become more, rather than less, ready to accept paraprofessional substitution. Increasing concentration in group practice also indicates that there is an increasing readiness for the individual physician to relinquish control over decisionmaking.

Policies that aim at reducing distributional inequities must build on these results if they are to succeed. Other policy effects that can be manipulated are the propensity of physicians to delay decisions until the time of residency and to emphasize life style qualities in the locale they choose. Possible alternative include programs to reduce the difficulty of entering practice in locales that are attempting to gain physicians, that promote the social and organizational integration of FMGs, that attempt to motivate residents to choose certain areas, and that entice students from areas with low physician populations to choose medicine as a career and to take training in their own state.
V. CONCEPT OF A PHYSICIAN SHORTAGE

Section III examined models of the demand for physicians, and Section IV examined models of physician supply. Here we consider the following questions: Under what conditions does the interaction of supply and demand factors lead to a condition requiring government intervention? How is physician shortage to be detected? Answering these questions will require a recapitulation of some of the arguments in Section III. The six most important criteria for a physician shortage are set out below.

1. Professional standards: A shortage is said to exist if the number of physicians available at a given place and at a given time is inadequate to meet some professionally defined standard of medical care. In Section III, we noted a number of methods that had been used to develop professional standards and argued that such measures were unlikely to help policymakers determine whether action should be initiated to change a situation. We also noted that in almost all areas, the amount of manpower available will be much less than that necessary to meet professionally defined standards and that, by these definitions, shortages are perpetual.

2. Comparative ratios: A shortage is said to exist in all those states (or counties) with a physician/population ratio lower than the mean ratio across states (or counties) or with a physician/population ratio lower than that of the "best" areas—defined, for example, by the areas with the highest ratios. Since it is extremely unlikely that physicians will be uniformly distributed across regions, such a definition will always imply a shortage in some areas. As noted in Section IV, these ratios make no sense unless the numerator is full-time equivalent physicians providing patient care, and the denominator is adjusted for age, sex, and race. Consider, for example, a situation in which one county has a physician/population ratio of 75/100,000 and another has a physician/population ratio of 150/100,000. Assume that half the physicians in the latter county spend their time in teaching, research, and administration, and in staffing a hospital that provides specialty care to citizens of the entire state; and some of the remaining physicians are retired or treat patients living in other counties. If, in the former county, all physicians engaged in
full-time patient care, strict manpower ratios would grossly overestimate the differential availability of physician services.

These ratios suffer the same problems as the professional standards approach. There is little reason to believe that additional physician services would be used, no assurance that additional physician services would improve the health of the populace, and no reason to believe that additional physicians would choose to settle in "underserved" areas.

(3) Demand/supply differential: A shortage is said to exist if, at current prices, the demand for medical care exceeds the supply of medical care. This is a strict economic definition of a shortage. Consider, for example, the market for rental apartments. An economist would argue that a well-functioning market would equate the quantity supplied to the quantity demanded (and determine an equilibrium price). If the demand for apartments suddenly expanded or some event curtailed the supply, the rental price might rise a great deal, but this would not be considered a shortage. According to this interpretation, a shortage is possible only if the market is not functioning. For instance, in the apartment rental example, if an apartment owner cannot or will not raise prices but customers want more apartments than the owner can supply, an economic shortage would be said to exist. However, from an economic viewpoint, this shortage is artificial and created by the constraint on price. As with rent control in New York City, the pernicious effects of keeping the market from clearing include curtailing future supply, increased litigation, and immobility. When such constraints occur in the market for medical care, they give rise to economically defined shortages. In such situations, the rationing role played by price is replaced by nonmonetary rationing devices.

Constraints on the supply of physicians in a well-functioning market for medical care lead to a high price per patient visit and high incomes for physicians. If physicians cannot or will not raise prices then the market would tend to equilibrate (that is, ration services among those demanding them) through the use of nonmarket rationing. In the delivery of physician services, the most common rationing devices are: (1) service unavailability (physicians refuse to see a new patient); (2) long waits for service (a delay of several weeks for an appointment); (3) deterioration in the product or service itself (a long wait to see the physician once one arrives, a small amount of time with the physician, a less than thorough examination, and perhaps little effort by the physician to be
reassuring or friendly); and (4) other increased difficulties in gaining access (such as a greater distance to travel, less convenient office hours, and a general way of putting more burden on the patient and having the service take more of the patient's time and effort).

Much current evidence indicates that such shortages exist. The market for medical care is not well-functioning since numerous factors prevent price from playing a rationing role. Some institutions commit themselves to deliver care at zero price to the patient, but they do not hire the manpower necessary to deliver that care. Physicians traditionally take an oath not to deny care to those who cannot pay for it; Blue Shield and other review mechanisms often set an effective upper bound on the price that physicians can charge (to specified groups); the welfare associated with medical care tends to induce strong expectations in the physician and patient that the price of care be related to the ability of each individual patient to afford it. Thus, high prices are not considered a socially acceptable way of rationing medical care. Nonetheless, price may have an important role to play as an incentive.

(4) Rate of return: A related way of detecting a physician shortage is to determine whether there is a high rate of return to physicians at a given time or place. A high rate of return across regions could indicate that physicians were able to create an artificially high demand for their services or that supply constraints existed. The latter is a market signal to attract more people into medical professions. Thus one could determine the rate of return to physician education at different locations (or equivalently, look at physician incomes and education costs at different places). This technique is likely to yield evidence that is contrary to the assumption of a general shortage. For example, urban areas, which have the highest physician/population ratios and, therefore, should have low prices and low physician incomes, have the highest prices and highest (hourly) incomes.

(5) Health levels: A more difficult (but more objective) way of determining the existence of a shortage involves surveying the health of a population. Correcting for age, race, sex, and income, such a survey would determine the rates of mortality, acute disease, chronic disease, disability, and bed days. If one assumes that medical care is a principal factor influencing health, this approach could be used to identify the need for more physicians.
An alternative to a health survey is based on the assumption that a shortage of physicians leads individuals to seek care only when they are very ill. If so, looking at the mortality rate and at the severity of patient presentations to physicians in an area would provide an estimate of the population's health status.

(6) Community satisfaction: A final measure of shortage involves surveying a population to determine whether there is general satisfaction with local medical services. The level of satisfaction need have little correspondence to the physician/population ratio. If the health status indexes indicate a shortage of physicians, but people are satisfied with the level of service and there is no indication of nonmarket rationing, it would make little sense to provide additional service since it would go unused. Alternative policies are required in this instance to motivate the use of extant supply. Note that if nonmarket rationing is important, the provision of extra services would be reasonable and the services would probably be used (since nonmarket rationing is an indication that current demand exceeds current supply at current price).

CONCLUSION

We have described six methods of determining whether a physician shortage exists. Professional standards and ratios have little to recommend them other than their simplicity; there is no guarantee that additional services would be used or, if they were, that they would be efficacious. Looking at the rate of return to physicians is really nothing more than one way of determining whether there are artificial constraints holding down physician supply. Examining the existence of nonmarket rationing of physician services is relevant because of the many constraints that prevent price from equilibrating supply and demand. This approach has the virtue of indicating that at current prices people desire more medical services than are being supplied. A similar approach, although a more costly one, is to survey the population regarding their satisfaction with the medical care system. Clearly, the best approach is to survey the health status of a population and determine whether additional physicians would be efficacious. This approach, together with one indicating unsatisfied demand, would indicate not only that additional physicians would be used, but that the additional services would be efficacious.
VI. HOW GOOD ARE WE AT FORECASTING SHORTAGES?

The demand and supply models described in Sections III and IV imply that there is a physician shortage. Even more important, they indicate that this shortage is likely to become worse over time. Analysts have forecast population, number of physicians, and the requirement for physicians. On the basis of forecasts indicative that we are in the midst of a worsening shortage, it has been argued that governmental policies influencing the supply of physician services must be rethought and that new policies, such as expanding the number and size of medical schools, must be established. However, as indicated in the earlier sections, we do not accept the conclusions that derive from the models that have been used to estimate physician requirements or supply. Table 5 presents selected forecasts that have been made in the past for population, physician supply, and physician requirements. The notes to the table spell out the assumptions underlying each forecast. The table shows the shortages that were expected and the actual population and physician levels.

In general, these results indicate that: (1) the population estimates covered a wide range, and the smallest estimate exceeded the actual population of the United States (and its outlying territories) in 1970; and (2) the estimates of physician supply covered a wide range with the highest estimate lower than the actual number of physicians in 1970. Clearly, forecasting is not a science and is subject to a considerable error. The table indicates that forecasters should make their assumptions clear and instead of generating a single estimate they should present a range of possibilities.

In Table 6, we present some estimates that have been made for 1975 (in the notes to the table, the assumptions underlying the projections are given). The projections, made in 1966, indicate that, once again, a large doctor shortage is expected.

The forecasts of both physician supplies and population are poor. By 1970, we had more physicians than were predicted by the Bane Committee Report for 1975. By 1971 (according to the AMA), there were 344,823

---

1 Table 5 is adapted from Butter (1967).
2 Table 6 is taken from Hansen (1970).
<table>
<thead>
<tr>
<th>Sources</th>
<th>Date of Projection</th>
<th>Pop. in Thousands</th>
<th>Supply MD and Osteo. (active and inactive)</th>
<th>Supply MD Only</th>
<th>Requirement MD and Osteo.</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1958</td>
<td>209,380</td>
<td>273,474</td>
<td>274,469 (b)</td>
<td>276,458 (b)(c)</td>
<td>2,984</td>
</tr>
<tr>
<td>B</td>
<td>1959</td>
<td>213,810</td>
<td>294,900 (a)</td>
<td>296,500 (f)</td>
<td>286,938 (d)</td>
<td>12,469</td>
</tr>
<tr>
<td>C</td>
<td>1959</td>
<td>213,810</td>
<td></td>
<td></td>
<td>299,000 (e)</td>
<td>4,100</td>
</tr>
<tr>
<td>D</td>
<td>1960</td>
<td>213,810</td>
<td></td>
<td></td>
<td>279,000 (g)</td>
<td>2,500</td>
</tr>
<tr>
<td>E</td>
<td>1964</td>
<td>214,570</td>
<td>327,900 (j)</td>
<td>324,900 (k)</td>
<td>319,900 (m)</td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>1966</td>
<td>212,683</td>
<td>335,000 (n)</td>
<td>340,000 (p)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>G</td>
<td>1966</td>
<td></td>
<td>306,954 (q)</td>
<td>326,915 (r)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>H</td>
<td>1966</td>
<td>208,576</td>
<td>332,700</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Actual</td>
<td>1970</td>
<td>207,976</td>
<td>348,300</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(a) Present production rate. (b) Increase graduates of U.S. schools. (c) Maintaining 1955 physician/population ratio. (d) Increase graduates sufficiently to maintain 1955 ratio of graduates to population 20-24. (e) To maintain 1959 ratio. (f) Recent growth rate. (g) Graduates at levels currently predicted. (h) Increase graduates to maintain 1957 ratio. (j) At current planned growth, increase graduates (1,600 foreign graduates licensed annually). (k) 1,000 foreign graduates annually. (m) No foreign graduates licensed after 1965. (n) Low estimates of U.S. graduates and new foreign unlicensed, stable new foreign licenciates. (p) High estimates of graduates and foreign unlicensed, stable new foreign licenciates. (q) Based on HMP growth in 1950-60, 4-1/2 percent per year. (r) HMP growth rate, 5-1/2 percent per year.
### TABLE 6

**SUMMARY OF PHYSICIAN PROJECTIONS FOR 1975**

<table>
<thead>
<tr>
<th>Projection Study</th>
<th>Requirements (I)</th>
<th>Supplies (II)</th>
<th>(-) Deficit (+) Surplus (III) (Col. II minus Col. I.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Bame Committee Report (U.S. Surgeon General, 1959)</td>
<td>330,000 (minimum)</td>
<td>(I) 312,800</td>
<td>-17,200</td>
</tr>
<tr>
<td></td>
<td>(II) 318,400</td>
<td></td>
<td>-11,600</td>
</tr>
<tr>
<td>2. Fein (1967)</td>
<td>(I) 340,000 to 350,000</td>
<td>361,700</td>
<td>+21,700 to +11,</td>
</tr>
<tr>
<td></td>
<td>(II) 372,000 to 385,000</td>
<td></td>
<td>-10,300 to -23,</td>
</tr>
<tr>
<td>3. U.S. National Advisory Commission on Health Manpower (1967)</td>
<td>(I) 346,000 (minimum)</td>
<td>360,000</td>
<td>+14,000</td>
</tr>
<tr>
<td>5. U.S. Public Health Service (1967)</td>
<td>(I) 400,000</td>
<td>360,000</td>
<td>-40,000</td>
</tr>
<tr>
<td></td>
<td>(II) 425,000</td>
<td></td>
<td>-65,000</td>
</tr>
</tbody>
</table>

*Physicians include both MDs and DOs, except for Line 2, which excludes DOs.*

**SOURCES:**

- **Line 1.** U.S. Surgeon General’s Consultant Group on Medical Education (1959). Column (I) Table 2, p. 3; Column (II) Table 2, p. 3; Column (III) Calculated. Requirements based on assumption that 1959 represents minimum rates to maintain health of population. Supply: Continuation of physician growth rate.
- **Line 2.** Fein (1967). Column (I)(I) Based on 12-15 percent increase due to population growth above. (II) Based on 22-26 percent increase due to all factors. See pp. 134-135. Column (II) Table III-9, p. 87. Column (III) Calculated. Requirements based on the demand for physician services at 1965 prices given expected changes in population composition by 1975. (I) shows estimated effect only accounting for population change; (II) shows the effect of a whole range of factors. Supply takes into account expected increase in medical school graduates as well as the immigration of foreign trained physicians.
- **Line 4.** U.S. Bureau of Labor Statistics (1967). Column (I) Page 18. Column (II) No figure is given. We assume National Advisory Commission on Health Manpower supply figure of 360,000 is appropriate to use. Column (III) Calculated. Requirements taken into consideration: Population changes, increased demand for services across all age groups, and need for an expansion of physicians engaged in research.
- **Line 5.** U.S. Public Health Service (1967). Column (I)(I) and (II), Table 8, p. 15, and accompanying text. Column (II) Same as Column II, Line 4. Column (III) Calculated. Requirements: (I) Based on the application of professional standards—namely, the utilization rate for members of prepaid group practice plans to the entire 1975 population; (II) applies the highest physician utilization rate among the four major regions of the United States to the entire 1975 population.
physicians. Little has changed in forecasting techniques, and it is very likely that again all forecasts of physician supplies will be underestimates.

It should also be noted that these forecasts do not take increases into account. If one assumed that physician productivity were to increase at 4 percent per year, the available "effective" supply in 1975 would be about 390,000 and the projected deficit would disappear.

Based on the criteria in the literature discussed above, there is no overall shortage of physicians. If anything, there may be a doctor glut over the next decade. There are significant problems with maldistribution of physicians, especially in rural areas. No simple policies to equalize the distribution of physicians are likely to be successful. However, it is not legitimate to assume that an area has a physician shortage just because some other area has a higher physician/population ratio. Instead, one must gather evidence of nonprice rationing or of unsatisfactory health indices (mortality ratio, morbidity ratio, or disability days).

In view of past attempts to solve medical care delivery problems by good intentional intervention, we would caution that good intentions are not enough. If funds are not to be wasted or to have a pernicious effect, careful data collection and analysis are necessary.
BIBLIOGRAPHY


Hansen, W.L., "'Shortages' and Investment in Health Manpower," in The Economics of Health and Medical Care, Bureau of Public Health Economics and Department of Economics, University of Michigan, Ann Arbor, 1964.


Lave, J.R., L.B. Lave, and T.E. Morton, "The Physician's Assistant -
Exploration of the Concept," Hospitals, Journal of the American Hospital
Association 45 (11), 42-51, July 1, 1971.

Lave, J.R. and S. Leinhardt, "The Delivery of Ambulatory Care to the Poor:
A Literature Review," Management Science 19 (4, part 2), 78-99, December
1972.

Lee, R.I. and L.W. Jones, The Fundamentals of Good Medical Care, The
University of Chicago Press, Chicago 1933.

Lepper, M.H., J.C. Lashof, M. Lerner, J. German, and S.L. Andelman,
"Approaches to Meeting Health Needs of Large Poverty Populations,"

Lubin, J.W., I.M. Reed, G.L. Worstell, and D.L. Drosness, "How Distance
Affects Physician Activity," The Modern Hospital 107 (1), 80-82, 156,
July 1966.

MacQueen, J.C., "A Study of Iowa Medical Physicians," Journal of the Iowa
Medical Society 58 (11), 1129-1135, November 1968.

Marden, P.G., "A Demographic and Ecological Analysis of the Distribution
of Physicians in Metropolitan America, 1960," American Journal of Sociology
72, 290-300, 1966.

Marguiles, H. and L.S. Bloch, Foreign Medical Graduates in the United

Marshall, C.L., K.M. Hassanein, R.S. Hassanein, and C.L. Paul, "Principal
Components Analysis of the Distribution of Physicians, Dentists, and
Osteopaths in a Midwestern State," American Journal of Public Health 61
(8), 1556-1564, August 1971.

Marshall, C.L., K.M. Hassanein, R.S. Hassanein, and C.L. Paul, "Time & Dis-

tance-Rural Practice. Dissatisfaction with Travel Distance to the Physi-
cian in a Rural Area," Journal of the Kansas Medical Society 70, 93-96,
March 1969.

Martin, E.D., R.E. Moffat, R.T. Falter, and J.D. Walker, "Where Graduates

Mason, H.R., "Manpower Needs by Specialty," Journal of the American Medi-


May, L.A., "The Spatial Distribution of Physicians--The Special Case of

the Balance of Physician Manpower in a Metropolitan Area," Public Health
Reports 85 (11), 1001-1011, November 1970.

Navarro, V., "Planning for the Distribution of Personal Health Services,"

Newhouse, J.P. and C.E. Phelps, Price and Income Elasticities for Medical
Care Services, The Rand Corporation, R-1197, Santa Monica, 1972.


Weiss, J.E. and M.R. Greenlick, "Determinants of Medical Care Utilization: The Effect of Social Class and Distance on Contacts with the Medical Care System," Medical Care 8 (6), 456-462, November-December 1970.


