THE ZERO PRICING OF HEALTH CARE: A REGIONAL ANALYSIS OF THE BRITISH NATIONAL HEALTH SERVICE

By

KURT HARLAN JACKSON

Bachelor of Science
Oklahoma State University
Stillwater, Oklahoma
1969

Master of Science
Oklahoma State University
Stillwater, Oklahoma
1977

Submitted to the Faculty of the Graduate College
of the Oklahoma State University
in partial fulfillment of the requirements
for the Degree of
DOCTOR OF PHILOSOPHY
December, 1984

THE ZERO PRICING OF HEALTH CARE: A REGIONAL

ANALYSIS OF THE BRITISH NATIONAL

HEALTH SERVICE

Thesis Approved:

Ansel M. Sharp
Thesis Adviser

Joseph M. Jadlow

Keith Olson

Patrick B. Dorr

Norman N. Durham
Dean of the Graduate College

PREFACE

This dissertation's origins are in England. On a sabbatical to the London School of Economics, my major adviser and dear friend, Dr. Ansel M. Sharp, met Dr. Anthony Culyer, a noted English health economist. Inspired by the experience and his study at the London School of Economics, Ansel carried the seed back to the new country. Admittedly overwhelmed by the scope of the project, but interested at the same time, I adopted the project as my own.

This project has been a part of my life and the lives of those close to me for a long time. Without the love and support of these people, I could not have completed it. In fact, one of the most important things I have learned from this dissertation is how important family and friends are. To begin, I wish to thank my major adviser and dear friend, Dr. Ansel M. Sharp. Ansel's encouragement and patience were always there when I needed it. In addition, I would like to thank my committee members for their support and constructive advice. I benefited greatly from Dr. Joseph M. Jadlow's experience with the methodology I used. In addition, Dr. Kent W. Olson and Dr. Patrick B. Dorr provided many helpful comments that greatly improved the manuscript. Other members of the faculty, especially Dr. Michael J. Applegate and Dr. Ronald L. Moomaw, also offered valuable assistance in the earlier stages of the dissertation.

During this project I enjoyed continued and strong support from both my parents, Ed and Virginia Jackson. They provided emotional and

financial support throughout the project, making our lives much easier than they would have been otherwise. In addition, I would like to thank my dear wife's parents, C. C. and Mildred King, for their support and the many days and nights of lodging while I completed the dissertation.

Last, but certainly not least, I wish to thank my loving family. My dear wife, Pamela, was always there to help me out and pick me up when I needed it most. My daughter, Brooke, was so patient and forgiving for the many nights Daddy had to go back to the office. My son, Weston, was there with his laugh and smile to boost me in the final stages of the project. And lest I forget, I wish to thank God for making all of these wonderful people and allowing me to be a part of their lives.

TABLE OF CONTENTS

LIST OF TABLES

Table | Page

LIST OF FIGURES

CHAPTER I

INTRODUCTION

Between 1974 and 1982 health care expenditures in the United States increased from 8.1 percent of GNP to 10.5 percent.[1] At the same time, medical care prices increased over 118 percent, while the consumer price index increased 65 percent.[2] Faced with this alarming and continuing trend, critics of the performance of the United States health care industry have offered national health insurance as an alternative to current market arrangements. In pursuing this path, the market critics seek to achieve three broad social objectives concerning health care. These objectives are: (1) to insure all persons access to adequate quality health care; (2) to eliminate the financial risks posed by serious illness; and (3) to control or limit health care costs.[3]

The primary technique offered to achieve the first and second objectives involves reducing the price paid by the health care consumer at the point of demand. Coinsurance rates vary from the currently common range of 20 percent offered by private insurance carriers to zero. Unfortunately, insuring access by lowering the price of health care at the point of demand creates excess demand. Excess demand leads to price increases and these price increases provide the economic stimulus for growth of the health care sector. Consequently, the three objectives cannot be met simultaneously by demand side policies alone. Controls on the supply side must limit the resource flows into the industry. If the supply side controls are not successful, the excess demands must be eliminated by some form of non-price rationing.

The approach outlined above is essentially the same approach used by England when it formed the National Health Service (NHS). While national health insurance proposals do not advocate nationalization of the health care sector, there are two basic similarities with the British approach. First, the price faced by the consumer is less than cost. Second, control over the supply of health care is required to offset the incentives designed to insure access to health care. The NHS uses both of these approaches to meet similar social objectives. Consequently, a study of the NHS's attempt to achieve the three objectives listed above should provide valuable information on both the strengths and weaknesses of this alternative to the current structure.

Purpose of Dissertation

The idea that medical care is a right is not new. Able-Smith reports that Nye Bevan, the Minister who introduced the National Health Service, said that medical care should be made available to everyone on the basis of medical need and no other criteria.[4] Klarman reports, in 1951, that a similar view of the individual's right to health care was well established in American thinking.[5] Should the citizens of the United States choose to make this statement valid for all persons, they should be aware of the problems that may be encountered along the way.

This dissertation, in attempting to provide some insight into the above problems, has two major purposes. The first is to consider the contributions to economic theory related to the public provision of health care, the zero-money price rationing of health care, and the separation of the concept of health care need from health care demand. The second major purpose of the dissertation is to empirically investigate three areas of NHS operation.

Areas of Investigation

The first area of investigation concerns the NHS's performance in allocating health care resources among geographic regions based on those regions' medical needs. The analysis will concentrate on the regions' hospital sectors. As part of this area of investigation, each geographic region's hospital sector's response to need within the region will be evaluated.

In the second area of investigation, we seek to evaluate the relative efficiency of the regions' hospital sectors. To do so involves estimating a hospital sector production function using linear programming techniques. After estimating the production function, we use this information to construct indices of efficiency for each region, and then compare these efficiency indices with measures of regional medical need and other variables. The objective here is to determine which regional characteristics support efficiency and which do not.

The third area of analysis is related to the second. It explores the behavior of NHS hospital doctors in allocating resources. As part of this analysis, this section evaluates the hypotheses that doctors' utility maximizing behavior will not necessarily support the NHS's stated objective of allocating health care resources on the basis of medical need.

Organization of Dissertation

Chapter II discusses the economic literature related to market failure in the medical care market. The literature on market failure is divided into two parts. The first part discusses market failure in the absence of externalities. The second part discusses market failure due to the presence of externalities.

Chapter III reviews the literature concerning the separate concepts of health care demand and health care need. In addition, the chapter discusses time prices as a non-money price form of rationing health care when the objective is to allocate health care equitably on the basis of need as determined by a third party and not by individual buyers. It also compares the performance of time prices and money prices in achieving equity in the allocation of health care resources.

Chapter IV presents the findings of the analysis concerning the regional distribution of NHS resources on the basis of three alternative measures of regional need. The simplist measure of need used to evaluate the NHS's performance in achieving its equity objective is equal per capita allocations and outputs. Two alternative indices, the age-sex adjusted population and the age-sex standard mortality ratio adjusted population are also used to evaluate the NHS's performance in this regard.

Chapter V presents the estimation procedure and findings of the hospital sector production function analysis. In addition, it reports the results of the investigation concerning which regional characteristics are associated with relative efficiency and relative inefficiency. Finally, it presents the findings concerning hospital doctor behavior in supporting the NHS's objectives at the regional level.

Chapter VI summarizes the conclusions of the research. Additionally, it presents suggestions for improving NHS performance in light of the findings. Lastly, the chapter suggests directions for future research.

ENDNOTES

[1]U. S. Department of Commerce, Bureau of the Census, Statistical Abstract of the United States, 1984 (Washington, D.C., 1984), p. 102.

[2]Ibid., p. 106.

[3]Karen Davis, National Health Insurance: Benefits, Costs, and Consequences (Washington, D.C., 1975), pp. 2-5.

[4]Brian Abel-Smith, National Health Service: The First 30 Years (London, 1978), p. 1.

[5]Herbert E. Klarman, "Requirements for Physicians," American Economic Review (May, 1951), p. 633.

CHAPTER II

MEDICAL CARE AND MARKET FAILURE

Private markets may fail to allocate medical care resources efficiently. The economics literature concerning these market failures can be divided into two groups. The first group concerns market failures that can occur in the absence of any externalities associated with medical care. The second group concerns market failures arising from externalities associated with medical care. This chapter discusses each group in turn.

Market Failure Without Externalities

Private markets may fail to allocate medical care resources efficiently even in the absence of externalities. These market failures may arise from several sources. First, a significant group of consumers may behave irrationally in their purchase and consumption of health care. If they do so, evaluating their behavior using the efficiency criteria of welfare economics would be inappropriate. Second, uncertainty in both the diagnosis and treatment of illnesses may prevent patients from equating marginal costs to marginal benefits. Third, and related to the second, competitive behavior of insurance companies combined with their inability to evaluate medical care consumers accurately by risk class leads to overinsurance which in turn motivates society to consume medical care past the point where marginal social benefits equal marginal social costs.

Market Failure Due to Consumer Irrationality

One would not expect private markets to allocate medical care resources efficiently if most consumers behave irrationally with respect to medical care. Culyer considers whether the existence of certain classes of persons provide sufficient evidence to support this proposition.[1] Three classes of persons cited as evidence of wide spread irrationality are: (1) mentally ill persons; (2) unconscious persons; and (3) persons who are sick but do not demand medical care.[2] If these groups represent most consumers, market critics argue that the principle of consumer sovereignty would no longer apply. Free choice should be replaced by a social welfare function prescribing the appropriate behavior for medical care consumers.

Culyer dismisses the first two classes as insignificant for two reasons. First, no one expects consumer rationality to apply to these groups. Society already makes decisions for these people. Second, the number of persons in these groups is relatively small. In addition, Culyer goes on to argue that the existence of a large group of these persons would not justify the violation of the consumer sovereignty of the sane and conscious.

Culyer examines the last class of persons as an obviously more serious case. These "sane and conscious" persons would be considered sick by medical experts yet they do not demand care. Can their behavior be called "irrational?" Culyer discusses why their behavior may not be irrational. He begins by dividing these persons into two groups:

(1) those who make a conscious decision not to consume medical care, and (2) those who do not consume medical care because they do not know they are sick.

Culyer starts his argument by noting that neither group is necessarily irrational and therefore they do not provide evidence of wide-spread irrationality.[3] Sick persons who consciously choose not to demand care may do so because they perceive their marginal costs of treatment to exceed their marginal benefits. If they correctly perceive the situation, they are rational within the context of welfare economics and there is no justification for violating their consumer sovereignty. If they perceive the situation incorrectly, then does this necessarily prove that they are irrational? Again, Culyer argues that it does not because their behavior could be due to a lack of information concerning their marginal benefits and marginal costs. Since information is an economic good, medical care consumers will acquire it just like any other economic good to the point where marginal benefits equal marginal costs. For optimality, some ignorance will exist. For these consumers to be considered irrational in the context of our model, they must know that their marginal benefits of treatment exceed their marginal costs and then choose not to demand care.

Culyer concludes that even if most consumers fall generally into this third group, this is not a sufficient condition to justify an allocation mechanism which discounts consumer sovereignty. The existence of sick persons who do not demand care does not prove irrationality.[4] It may be an indicator of the level of ignorance existing within the current system. Whether this level of ignorance is optimal or not is an empirical question. The group's existence does not provide us with the necessary information concerning the number of persons who actually satisfy our defintion of irrationality.[5]

Market Failure Due to Uncertainty

A second reason markets may fail to allocate medical care resources efficiently is associated with uncertainty, both in the diagnosis and treatment of disease. Culyer notes that patients often can not calculate the cost of treatment in advance or evaluate the quality of treatment received. In addition, actuarily fair insurance may not be available due to loading charges and moral hazard, which cause actual insurance premiums to exceed actuarily fair premiums.[6]

If a patient can not determine the quality of care in advance or the cost of the care, how can the patient equate marginal benefits to marginal costs? Two institutional arrangements have evolved to deal with this problem. The first of these is the doctor-patient relationship while the second is health insurance.

The Doctor-Patient Relationship

Feldstein notes that the agency relationship between doctor and patient develops because it is easier for the patient to tell the doctor the relevant parameters of his demand function, such as his financial position, insurance coverage, fears, needs, than it is for the doctor to give the patient all the necessary information to make the decision himself.[7] Feldstein adds that if the agency relationship is complete, observed demand would be identical to the demand of an informed consumer.[8] But if the relationship is not complete, the two demands will not be identical. Titmuss, a critic of market provision, argues that technological advances and increased specialization of doctors' skills have eroded the agency relationship that once existed between the patient and the family doctor.[9] In the private market, where the doctor is both a demander and supplier of services, the doctor has a financial incentive to prescribe higher quantities and qualities of care than an informed consumer would demand. Titmuss holds that "the conflict between professional ethics and economic man should be reduced as far as is humanly possible."[10] His solution is public provision where doctors would not be compensated on a fee-for-service basis, as in the British NHS.[11]

While fee-for-service payment schemes provide incentives to increase the number of services provided, and may increase the "conflict between professional ethics and economic man" as Titmuss warns, complete nationalization of medical care is not the only alternative. Prepayment schemes, such as HMO's, avoid the possible distortions of fee-for-service while keeping medical care resources in private hands. Additionally, other aspects of the patient-doctor agency relationship may counter some of the adverse effects of the conflict of interest in the agency relationship.

Friedman argues that both patients and doctors can expect to have "lower aggregate costs and better-informed decisions" if they form a long-term association.[12] Potential benefits from long-term association for the patient include shorter waiting time for appointments and free services by telephone. The doctor benefits from shorter consultations and increased knowledge about his patient.[13] Citing U.S. data for 1973-75, he notes that the average days wait for an appointment for an established patient is approximately 50 percent shorter than for a new patient.[14]

Long-term associations are available to patients in the United States and in England. Increasing specialization and technological change may indeed place a patient in "strange hands" but this characterization applies to both the United States and England. And, in England, hospital doctors (specialists) represent a different group of persons than the family doctor (GP). In the United States, family doctors practice in hospitals along with specialists. One might argue that this would facilitate communication between specialist and GP, thus enhancing the agency relationship. From the foregoing, Titmuss' argument of effects of technical change and specialization would not give the British NHS any obvious advantage in terms of an improved agency relationship. So if the agency relationship in the United States were found to diverge more from the optimum than in Britian, the obvious culprit would be fee-for-service compensation. And, as already noted, nationalization is only one alternative for removing "the conflict between professional ethics and economic man."

Market Failure and Health Insurance

Health insurance represents a second institutional arrangement which attempts to resolve the problems of medical care consumption created by uncertainty. But while conflicts of interest may distort the agency relationship between doctor and patient and lead to non-optimal results, non-optimal outcomes can also result from both the absence of health insurance and the presence of too much health insurance. The works of Arrow and Pauly cover these problems.

The Absence of Health Insurance. A risk-averse medical care consumer would purchase health insurance as long as his utility level after purchasing the insurance is higher than it would be if the loss occurred.[15] The same consumer would self-insure if the purchase of insurance leaves him at a comparatively lower utility level. If the market fails to provide insurance for certain medical care risks, a welfare loss results when consumers who would purchase insurance do not have that opportunity. Under these circumstances, Arrow contends a case for government provision of insurance exists.[16]

Arrow argues that uncertainty associated with the diagnosis and treatment of illness prevents the market from providing insurance against certain risks. This non-marketability of risks violates one of the principle assumptions of the competitive model and results in a welfare loss to society.[17,18] Some risk-averse persons will be unable to obtain insurance at an actuarily fair premium (which equals the probability of a loss multiplied by the amount of the loss). As a result, they will self-insure.

A second reason actual premiums may exceed fair premiums is because fair premiums are "loaded" with administrative costs and profits.[19] If the insurance industry is competitive, these loading charges should reflect the alternative costs of the resources providing the insurance.[20] On the other hand, if barriers to entry or other forces limit the level of competition, loading charges will exceed the alternative costs of the resources involved. Either way, loading charges raise the actual premium above the fair value, and may motivate some risk-averse persons to self-insure.

"Moral hazard" is another reason actual premiums may exceed actuarily fair premiums. Pauly cites the insurance industry's definition of moral hazard as "the intangible loss-producing propensities of the individual assured."[21] In other words, the insured person will demand more at a zero price than at the positive price. While the insurance literature associates this behavior with morality, Pauly notes that "moral hazard" represents rational economic behavior.[22] If the demand for units of medical care is not perfectly inelastic and having insurance lowers the private marginal cost of a unit to zero, the rational consumer will consume medical care to the point where private marginal benefits are zero. Since this level of consumption exceeds the level where marginal benefits equal marginal costs for the uninsured person, the size of the loss resulting from a given illness is increased. The insurer will recognize this and adjust premiums upward to compensate. Consequently, actual premiums again may exceed actuarily fair premiums, resulting in some self-insurance among the risk-averse.

Because actuarily fair premiums are not available, some risk-averse persons will self-insure and some welfare loss will result. But Pauly warns that this does not provide a sufficient case for the public provision of health insurance. He contends that only if the costs incurred from an illness are completely random would Arrow's argument be valid. The existence of moral hazard makes some medical expenses "uninsurable" in the strict sense envisioned by Arrow.[23] While the probability of becoming sick may be a completely random variable, only a perfectly inelastic demand curve for medical care would insure that the actual costs of the illness be completely random. Since actual premiums would incorporate the effects of moral hazard, the insurance market for certain forms of price sensitive medical care would not be observed because most risk-averse persons would self-insure as opposed to paying the necessary premium.

The consumer, as a purchaser of insurance, equates marginal benefits to marginal costs. But if the insurance purchased reduces the marginal cost of medical care to zero, the consumer, as a puchaser of medical care, equates marginal benefits to zero. Pauly concludes that commercial insurance will be provided for those events:

a. For which the quantity demanded at a zero price does not greatly exceed that demanded at a positive price;

b. For which the extent of randomness is greater, so that risk spreading reduces the risk significantly, and;

c. For which individuals have a greater risk-aversion.[24]

A case for the public provision of health insurance cannot rest solely on the presence of risk-averse uninsured persons. Mandatory public insurance involves risk spreading over the country's total population. Just as with private group insurance, some persons would gain at the expense of others. Public provision of insurance must be justified some other way. For example, if the increased risk spreading over the entire population provides a means of reducing the risk beyond that obtainable by the private market, premiums could be lower and net welfare gains could result. But if Pauly's three conditions adequately describe the insurance that is likely to be provided by the private market, all potential gains from risk spreading may have been already exploited.[25]

Alternatively, public provision of insurance may realize economies of scale in administration. If these economies do exist, loading charges would be lower. Actual premiums would exceed fair premiums by a smaller amount. As a consequence, fewer risk-averse persons would self insure.

The Presence of Too Much Health Insurance

Pauly presents another argument for the public provision of insurance. Competition among insurance companies may prevent them from charging premiums based on the expected losses of the insured.[26] Based on the insured assessment of his (her) needs, the probability of a loss varies directly with the amount of insurance bought. The insured's expected losses vary directly with the amount of insurance purchased. If the insurer knew this, they could identify the persons with larger expected losses and charge appropriately higher premiums. If single sellers try to increase the price for additional units they sell, the customer can simply go to another firm. Pauly argues that the competition for business will motivate sellers to conceal this information from other sellers. This competitive behavior allows the insured to purchase more insurance than if insurers knew how much insurance all persons had purchased. Increased insurance translates into more moral hazard.[27] Pauly concludes that an optimal result would be produced by a law that required individuals to buy exactly the number of units of public insurance they would have purchased had premiums varied in accordance with their expected losses and no more.[28]

While Pauly states that the public provision of insurance provides a solution to the problems created by too much insurance, he goes on to recommend that the prefered role of government would be to provide insurance companies with information concerning the total amount of insurance purchased by an individual.[29] Given this information, firms could charge premiums based on expected losses and individuals would not over-insure. Consequently, the level of moral hazard would be reduced to the point where the benefits from risk reduction just equaled the expected loss resulting from the purchase of insurance.[30]

Feldstein looks at overinsurance created by private markets in a different way.[31] As he sees it, medical care consumers are caught in a dilemma. For consumers as a group, increased insurance coverage causes the price of medical care to increase. The price increase induces an increase in the quality of inputs used to produce medical care, which reinforces the price increase. This process increases expected losses from illness and motivates consumers to increase their coverage.[32] This sets off another round of price and quality increases. This places the consumer in a vicious cycle that magnifies the welfare loss created by moral hazard. Feldstein concludes that substantial net gains to society would result by constraining medical care consumers to paying higher coinsurance rates. In other words, if medical care consumers are forced to pay a larger proportion of their medical costs at the margin, the welfare loss from increased risk bearing will be more than offset by the welfare gain from reduced price distortion.[33]

Again, a type of market failure has been identified with uncertainty. Without government intervention, it seems unlikely that insurers will raise coinsurance rates. Competition among themselves prohibits this. If the government could establish a law setting the coinsurance rate at some level, price distortion could be reduced. But Pauly warns against standardization. He concludes:

> If persons differ (a) in the strength of their risk aversion and (b) in the extent to which insurance of various types alter the quantity of medical care they demand, an optimal state will be one in which various types of policies are purchased by various groups of people.[34]

Feldstein's article points to another potential problem created by insurance coverage for medical care expenses. Increased insurance coverage may lead to too much of a country's resources being devoted to medical care. Figure 1 begins this argument.[35] Without insurance, the market price of medical care is P_1 and quantity M_1 is provided. When insurance is made available with a coinsurance rate (C), the net price consumers must pay falls to CP_1. At this price, M_3 units of medical care will be demanded. Since suppliers will produce only M_1 units at price P_1, a shortage results which causes price to rise. When the price of medical care reaches P_2, an equilibrium quantity of M_2 units results.

If Feldstein's dilemma argument is correct, the attained equilibrium will not last. Higher prices induce suppliers to increase quality. If consumers perceive the quality to be higher, the demand curve will shift to the right.[36] Assuming consumers do perceive the quality change, demand increases to D_1D_1. Consumers facing a net price of CP_2 will now demand M_4 units of medical care. A shortage of M_2M_4 units results which drives the price up to P_3. The net price faced by consumers is now CP_3, which for convenience only, equals the original equilibrium price of medical care without insurance. But now quantity M_3 is associated with the same outlay by the consumer due to the quality induced increase in demand. Resource allocation decisions concerning medical care are now based on a price of P_3 dollars per unit. Without insurance, a lower quality of medical care was provided and resource allocation decisions were based on a price of P_1 dollars per unit.

Figure 2 illustrates the effects of the price distortion created by medical care insurance on the distribution of the economy's resources between medical care and other goods and services.[37] Society's

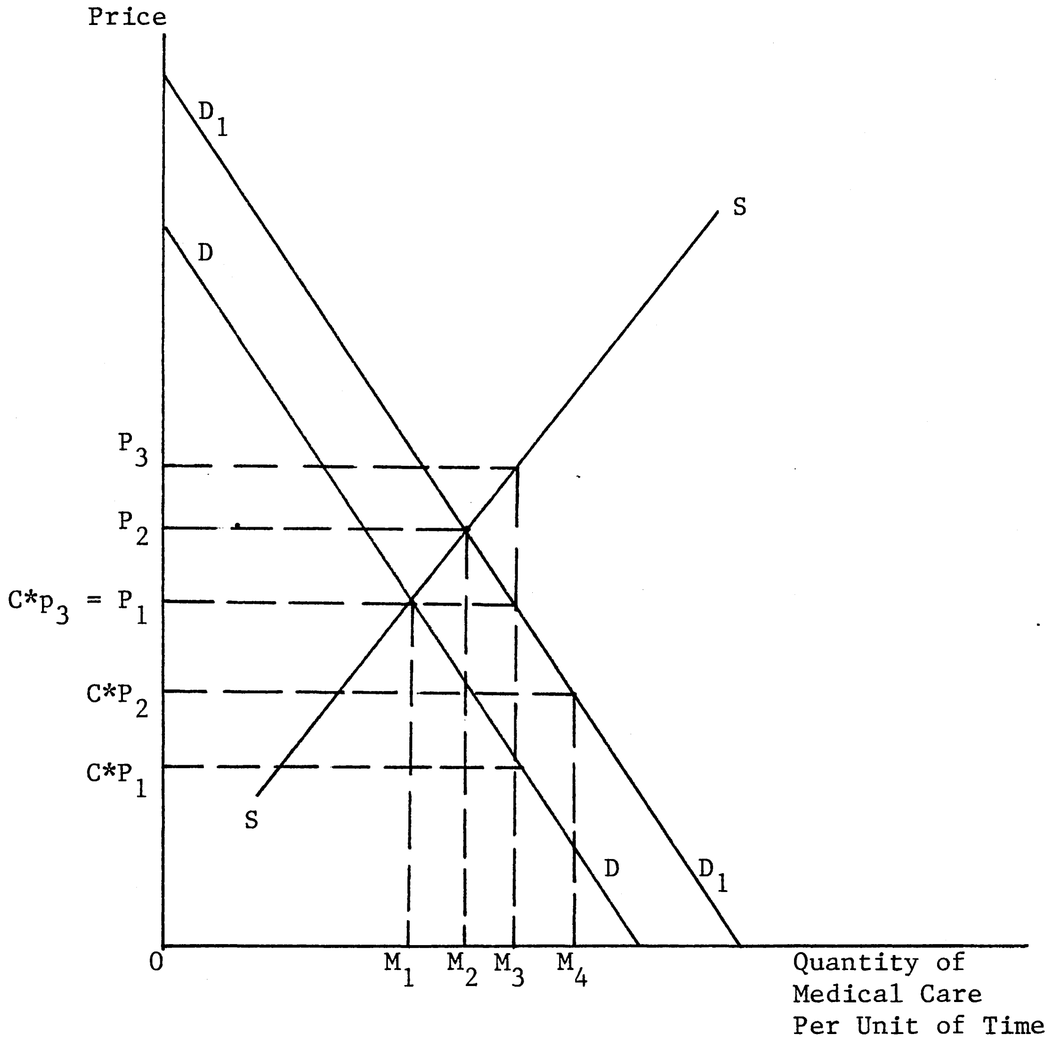

Figure 1. Insurance and Medical Care Price Inflation

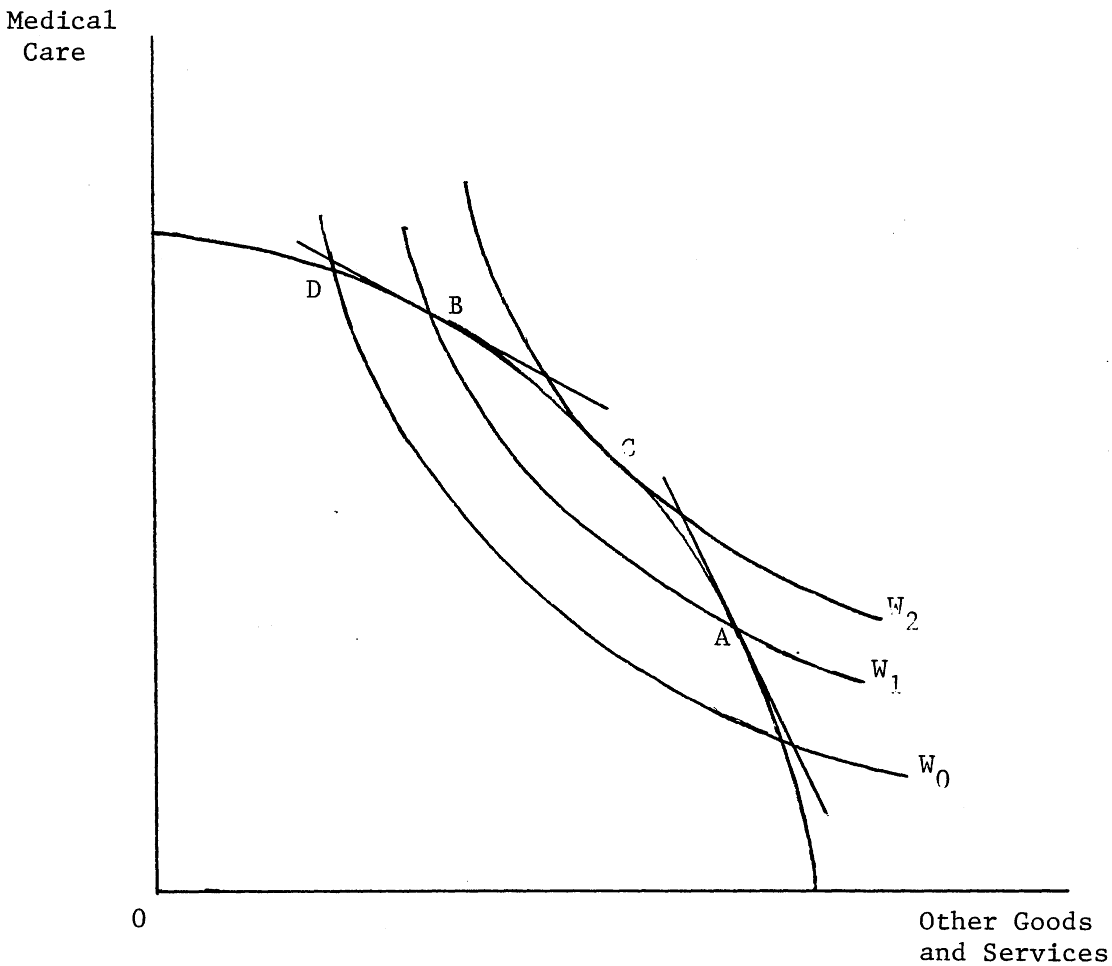

Figure 2. Insurance Induced Price Distortion and Resource Allocation

initial position without insurance for medical care is point A on indifference curve W_1. Assuming perfect competition in all markets for clarity of exposition, the slope of the production possibilities curve (PPC) at point A equals the negative of the ratio of the marginal cost of other goods to the marginal cost of health care. In perfect competition, price and marginal cost are equated so the slope of the PPC is P_{og}/P_1.

Now point A is clearly not a welfare maximum because risk-averse persons are forced to self-insure. Introducing medical care insurance with a coinsurance provision would cause the price of medical care to rise. If this price increase raises medical care prices to P_3, the economy may move past point C to point B where the slope of the PPC is $-P_{og}/P_3$. Given the social welfare function drawn, the welfare losses created by the price distortion have now completely offset the welfare gains resulting from insuring risk-averse persons for medical care costs. Unless someone resolves the dilemma medical care consumers face, further quality increases could actually lower welfare to W_0. Based on the arguments presented by Pauly and Feldstein, it is unlikely that competitive forces in the market place will be able to resolve the dilemma without some form of government assistance. As noted earlier, one possibility would be for the government to set minimum coinsurance limits. Raising the coinsurance rate above C in Figure 1 would increase the net price to consumers and create surpluses. These surpluses should place downward pressure on prices which would provide incentives for some resources to move out of the medical care sector. This corresponds to movement from B toward point C in Figure 2.

The preceding analysis uses the assumption of perfect competition. Monopoly forces, such as the AMA, certainly exist. But many of the resources in the medical care sector are accounted for by hospitals. In the United States, nonprofit hospitals account for the majority of hospital assets. If these hospitals have and use monopoly power in a conventional way, they will equate marginal revenue to marginal cost. The result would be a restriction of output and a restriction of the flow of resources into the medical care sector. This would tend to offset the expansion of the sector induced by insurance. But this does not occur because non-profit hospitals use the excess revenues generates by insurance induced price increases to increase the quality of care provided. Quality increases stimulate further increases in demand and pull more resources into the medical care sector.

Market Failure with Externalities

Critics of the market provision of medical care also attribute to externalities part of the failure of private markets to allocate medical care resources efficiently. If indivisible external benefits accure to others in society from the production and consumption of medical care, private production and consumption decisions will be non-optimal from society's viewpoint.[38] Culyer discusses three cases involving external benefits cited in the economic literature. They are:

a. Public health activities and immunization from communicable diseases,

b. The provision of sufficient capacity to satisfy fluctuations in demand; and

c. The desire by some rational members in society for other rational members in society to consume more medical care.[39]

Public Health Activities

Economists generally agree that significant external benefits are associated with public health activities and immunization programs. The benefits from these activities generally accrue to all members of the community. Since it is difficult or impossible to exclude people who do not pay from enjoying these benefits anyway, people will be motivated to understate their preferences for these activities. This free rider problem will prevent private markets from producing an optimum quantity of these services and justifies government intervention in one of two ways. The government can subsidize private production to provide these services at a very low cost, usually zero, to members of the community. Financing these subsidies out of general tax revenues reduces the number of persons taking a "free ride." To the extent that general tax revenues are not regressive, this solution is generally superior to a law requiring all persons to purchase immunizations against communicable diseases at private market prices.[40] The alternative course of action would involve government provision of these services.[41]

Optimum Hospital Capacity

Weisbrod points out a second type of externality that relates to health care.[42] The services of a hospital are needed infrequently by most persons and the actual moment of need is uncertain. Furthermore, the cost of expanding capacity is very high. Given these characteristics, there is a demand for what normally would be considered excess capacity. Since all persons would generally benefit from this excess capacity, voluntary provision of excess capacity will put the burden for payment on current users in a fee-for-service system. Under these circumstances, one would expect a suboptimum amount of capacity to be provided by the private markets and potential users to benefit without paying, thus creating a free-rider problem.

The British NHS (government provision) represents one possible solution to the problem. Financing hospital capacity out of general tax revenues will prevent "free rides" of this type. But Culyer and Lindsay note that government provision is only one possible solution.[43,44] Prepayment schemes, such as HMO's, will also capture option demand. Culyer goes on to say that the choice of how to best capture option demand depends on which method best approximates the socially optimum capacity at the lowest cost.[45]

External Demands for Medical Care

The third case of external benefits concerning medical care is the consumption externality. Culyer believes this issue poses a key problem for health economics.[46] How does society deal with persons who do not consume adequate amounts of medical care by society's standards? These persons impose a disutility on others. This disutility is eliminated if they consume more medical care, regardless of who pays. Again, a free-rider problem exists which prevents market solutions, such as private charity, from obtaining optimum results. In attempting to solve the problem initial approaches regarded medical care as a merit good. Choice was imposed on those with inadequate consumption. But Culyer notes that this approach suffered from the inability to make interpersonal utility comparisons. There is no way to be sure if the gains outweigh the losses.

A more promising approach is based on the voluntary exchange theory of public finance. By assuming utility interdependence, the consumption-externality motive can be evaluated in a gains from trade framework without using interpersonal utility comparisons. In this approach the individuals consumption of medical care becomes a semi-public good, providing private benefits to the individual and public benefits to a significant segment of society. In the voluntary exchange framework, the level of consumption equating private marginal benefits with marginal costs falls short of the output where marginal social benefits equal marginal cost.[47] Potential gains from trade exist since both society and the individual can be made better off, if society subsidizes the individual to increase consumption of medical care to the point where marginal social benefits equal marginal social costs. But because the public good aspect of medical care is not subject to the exclusion principle, the free-rider problem will thwart attainment of a pareto-optimum unless potential beneficiaries are forced to contribute to the subsidy scheme. Public provision, then, provides the means of internalizing this externality.

Figure 3 presents Culyers argument, which he calls the traditional approach.[48] Society consists of the persons, R and P. R's utility is a positive function of P's consumption of medical care but P's utility is independent of R's. The private demands for each's medical care are D_r and D_p, respectively. If marginal social costs (MSC) are assumed constant and equal to private marginal costs (MC) each individual will consume M_r and M_p units of medical care respectively and the total amount provided by the private market will be M units, the horizontal sum of these private demands. But, this result is sub-optimal because it does not include R's demand for P's consumption (D_{pr}).

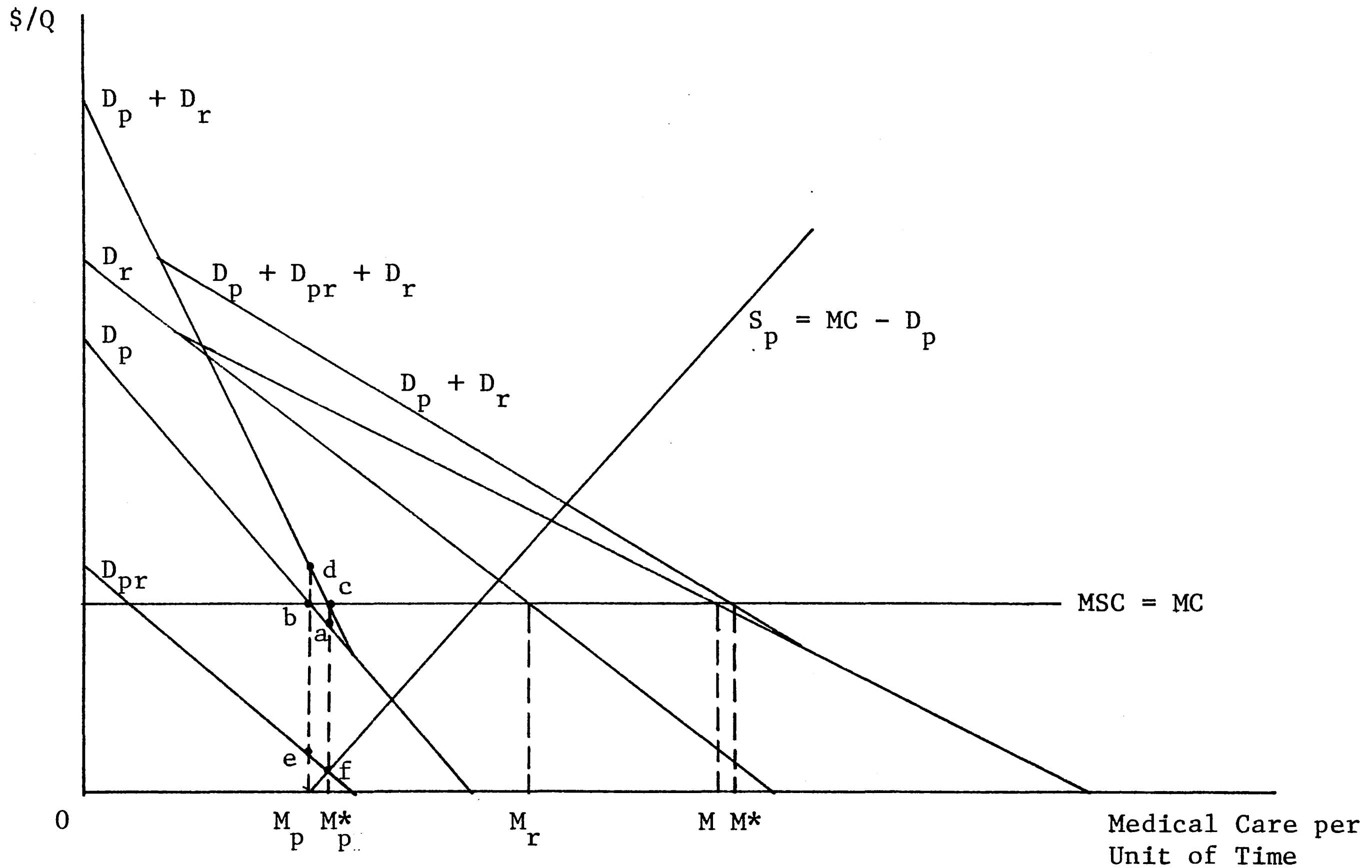

Figure 3. External Demands and Market Failure[49]

To determine the optimum amounts for P to consume and the market to provide (M_p^* and M*), the total demand for P's private demand (D_p) must be vertically added to R's demand for P's consumption (D_{pr}).[50] The total demand for P's consumption is ($D_p + D_{pr}$). P's optimal consumption is M_p^* where the "total marginal benefits" of P's consumption of the last unit of medical care equal society's marginal costs of providing that last unit. The total amount of medical care the market should provide is M* where the horizontal sum of ($D_p + D_{pr}$) plus D_r intersects MC.

The voluntary exchange occurs as follows. In the absence of coercion or subsidy, P will consume M_p units. At this consumption level, MSB equals M_pd and MSC equals M_pb. If P's consumption is increased to M_p^*, R's welfare will increase by area bdc. If this increased consumption is brought about by coercion, P would suffer a welfare loss equal to area abc. Since interpersonal utility comparisons cannot be made, we do not know the areas of these triangles so we cannot make a case for coercion. But the voluntary exchange approach shows that if R subsidizes P to voluntarily consume $M_pM_p^*$ units more medical care, R can increase his welfare by area $M_p$ef while P is made no worse off. In order to induce P to consume more than M_p units, R must pay the difference between MC and D_p, shown by line S_p. As long as S_p lies below D_{pr}, R will benefit by subsidizing an extra unit of consumption. P's optimum quantity of medical care consumption (M_p^*) occurs where D_{pr} intersects S_p.

Utility interdependency is necessary to make the voluntary exchange approach work in this context. Culyer's traditional approach introduces the absolute quantities of medical care consumed by P into R's utility function and shows how consumption subsidies can solve the problem.[51] But the reader should note that Lindsay has presented an alternative approach that uses a different interdependency term.[52] In Lindsay's model the equality of the distribution of medical care between R and P is the externality term in R's utility function. Faced with different means of achieving a more equal distribution of medical care among members of society, Lindsay's R's will choose a least-cost approach requiring both rationing for the R's and subsidies to the P's.

Lindsay and Buchanan argue that the type of externality which actually exists has important public policy implications.[53] If public concern lies in the absolute quantities of medical care consumed by the P's, then the appropriate form of government intervention is the provision of per-unit consumption subsidies to reduce the price of medical care to the P's.[54] By lowering the unit price of medical care, the P's voluntarily consume more medical care than they would if they were given a lump-sum voucher or income transfer.[55] This prescription, based on Pauly's article on consumption subsidies, requires that the size of the subsidy vary inversely with the income of the recepient. It is based on the assumption that lower income persons consume less health care and create a greater externality from under-consumption than higher income persons.[56]

If the relevant externality is the equality of the distribution of medical care among all members of society, government provision of per-unit consumption subsidies is not the least-cost method of achieving equality in the distribution. As noted earlier, Lindsay showed the least-cost approach to achieving distributional equality is a combination of subsidies to the P's and rationing for the R's. If administered correctly, this combination will promote equal access to society's medical care resources for both groups. Lindsay and Buchanan warn that equal-access preferences do not necessarily justify an "everything free" system.[57] Without funding the medical care system sufficiently to meet all demands, the relatively scarce resources must be rationed by time prices. Since time prices vary among individuals, equal access is not assured. Lindsay and Buchanan conclude that equal access can only be assured if medical care could be obtained by all at no cost, including the cost of time.[58]

ENDNOTES

[1]A. J. Culyer, "The Nature of the Commodity 'Health Care' and Its Efficient Allocation," Oxford Economics Papers, XXIII (1972), p. 191.

[2]Ibid.

[3]Ibid., pp. 192-193.

[4]Ibid.

[5]What number of consumers is significant is a value judgment that society must make. This issue is closely related to consumption externalities and will be discussed in more detail in subsequent sections.

[6]Ibid., p. 194. Moral hazard will be discussed in detail later.

[7]Martin Feldstein, "Quality Change and the Demand for Hospital Care," Econometrica, XLV (October, 1977), p. 1682.

[8]Ibid., p. 1683.

[9]R. M. Titmuss, "Ethics and Economics of Medical Care", in Commitment to Welfare (London, 1968), p. 255.

[10]Ibid., p. 260.

[11]Public provision is not the only way to avoid the erosion of the agency relationship resulting from a fee-for-service payment mechanism. Pre-paid medical practice, e.g., HMO's, is an alternative compatible with private market provision that avoids the distortions generated by the fee-for-service mechanism.

[12]Bernard Friedman, "On the Rationing of Health Services and Resource Availability," Journal of Human Resources, XIII (Supplement, 1978), p. 59.

[13]Ibid.

[14]Ibid., p. 61.

[15]Risk-averse persons, by definition, have a diminishing marginal utility for money income. Each additional dollar of income increases utility less than the previous dollar. For a more detailed explanation, see: Joseph P. Newhouse, The Economics of Medical Care (Menlo Park, CA, 1978), pp. 19-20.

[16]Kenneth J. Arrow, "The Welfare Economics of Medical Care," in Health Economics, M. H. Cooper and A. J. Culyer (eds.) (London, 1973), p. 16.

[17]Ibid.

[18]Ibid., p. 23.

[19]Ibid., p. 34.

[20]Culyer, 1972, p. 191. He notes that the presence of loading charges, in and of themselves, does not provide a case against market provision since alternative mechanisms would have administrative costs too.

[21]Mark V. Pauly, "The Economics of Moral Hazard: Comment," American Economic Review, LVIII (June, 1968), p. 535.

[22]Ibid., p. 531.

[23]Ibid., p. 532.

[24]Ibid., p. 534.

[25]Pauly's argument implies a certain income distribution. The ability to purchase insurance is obviously important. If the majority of the population "self-insure" because they cannot afford insurance, public provision at "affordable" prices may provide significant benefits. But this is a separate problem of income distribution and not a failure of the insurance market per se.

[26]Mark Pauly, "The Public Provision of Insurance," Quarterly Journal of Economics, LXXXVIII (February, 1974), p. 50.

[27]Ibid., p. 51.

[28]Ibid., p. 52.

[29]Ibid., p. 53.

[30]Ibid., p. 52.

[31]Martin Feldstein, "The Welfare Loss of Excess Health Insurance," Journal of Political Economy, LXXXI (March-April, 1973), p. 252.

[32]Ibid.

[33]Ibid., p. 277.

[34]Pauly, 1968, p. 534.

[35]Figure 1 and the following discussion is based on the analysis presented by Feldstein, 1973, pp. 266-268. Felstein concentrates his analysis to welfare loss while this analysis emphasizes the size of the medical care industry.

[36] Ibid.

[37] The proportion of the economy's resources devoted to the medical care sector is exaggerated for purposes of illustration only. In addition, the shapes of the indifference curves associated with a hypothetical welfare function for society are arbitrary and could be drawn differently without affecting the basic conclusions of the analysis.

[38] Individuals will equate private marginal benefits to private marginal costs. But at this level of output, private marginal benefits are less than social marginal benefits so gains in welfare would result from expanding output. It is ironic that the market failure arguments of Pauly and Felstein presented earlier might provide a solution, although a non-optimal one, to this problem. But Feldstein's argument involves increases in quality. The present discussion assumes quality is held constant.

[39] Culyer, 1972, pp. 199-200.

[40] If all persons pay the same price for immunizations, the expenditure will represent a larger portion of the poors' income than for the rich. Furthermore, decreasing the money price of immunizations to zero may not be sufficient to insure all persons will get their shots. Travel and waiting times will act as prices too. From a practical standpoint, authorities may inact laws to insure compliance. While the effects of zero money prices will be discussed in detail in a subsequent section, the reader should note that zero money prices are not sufficient to guarantee the program is not carried out without "regressive effects."

[41] Many states require children of school age to present a certificate indicating they have received vaccinations against certain communicable diseases. These immunizations can be obtained from a private physician at a positive money price or at a public health clinic at a zero money price. Although the public clinic may not charge a money price, the waiting time might be higher.

[42] Burton A. Weisbrod, "Collective-Consumption Services of Individual-Consumption Goods," <u>The Quarterly Journal of Economics</u>, LXXVIII (August, 1964), pp. 471-477.

[43] Culyer, 1972, p. 201.

[44] Cotton M. Lindsay, "Option Demand and Consumers Surplus," <u>The Quarterly Journal of Economics</u>, LXXXIII (May, 1969), p. 345.

[45] Culyer, 1972, pp. 201-202.

[46] See the previous discussion on consumer rationality.

[47] Culyer, 1972, pp. 203-205.

[48] Ibid.

[49]P's consumption of medical care is a semi-public good. Private benefits accrue to P but R benefits from P's consumption and cannot be excluded from benefiting. R's valuation (D_{pr}) must be vertically added to P's (D_p) to get the total demand for P's consumption of medical care.

[50]This analysis is fundamentally Culyer's. It is included to clarify the external-benefits argument for public provision. See Culyer, 1972, p. 204.

[51]A. J. Culyer, "Medical Care and the Economics of Giving," Economica, Vol. XXXVIII (August, 1971), pp. 295-303.

[52]C. M. Lindsay, "Medical Care and the Economics of Sharing," Economica, Vol. XXXVI (1969), pp. 351-362.

[53]C. M. Lindsay and J. M. Buchanan, "The Organization and Financing of Medical Care in the United States," Financing Health Services (London, 1970), p. 564.

[54]Ibid.

[55]Ibid.

[56]Ibid.

[57]Ibid., p. 567.

[58]Ibid., p. 568.

CHAPTER III

NEED, EQUITY, AND TIME PRICES

Chapter II discussed market failures associated with medical care resources allocation. The existence of sources of market failures not associated with externalities indicated areas where government intervention could improve economic efficiency. But the solutions did not necessarily require direct government takeover of the resource allocation mechanism and the factors of production. Even in the case of a significant consumption externality, coercive financing of medical care for any target group did not necessarily prescribe nationalization of the medical care sector, national health insurance being an alternative. Nevertheless, the British took this approach in founding the NHS.

By founding the NHS, the government partially replaced consumer sovereignty concerning medical care with a social welfare function which states that medical care resources should be allocated solely on the basis of medical need.[1] The NHS represents only a partial replacement of consumer sovereignty because the consumer still has some freedom of choice, e.g., which GP to see, to go to the doctor or not, or to consult a private practice physician.[2] On the other hand, the financial ability to pay a money price does not guarantee a medical care consumer immedicate treatment or treatment at all. The constraint imposed by the NHS's social welfare function requires that the individual's need rank sufficiently high on a needs priority basis to receive care, given the resources available to the NHS.

To evaluate how the NHS allocates resources while attempting to satisfy this equity objective, this chapter first discusses the concepts of demand and need. After developing the concept of need, the second part of the chapter discusses resource allocation using time prices. The final part of the chapter evaluates the suitability of time prices in helping the NHS achieve its equity objective.

Demand Versus Need

Health planners often use the term "need" to discuss what they think "ought to be consumed" in the context of medical care resource allocation. The term implies an external judgment that someone (or some group) needs something. The word demand, on the other hand, implies a choice made by a consumer. In making this choice, a consumer's perception of his need for medical care is only one factor among many that determine whether the consumer demands medical care. Feldstein points out that resource allocations based on the needs of a given group may vary considerably from allocations based on the group's demands.[3] Resources that are needed may not be demanded. To successfully allocate medical care resources on the basis of need and avoid waste, the demand for medical care must be understood. The following will first develop the concept of demand and then need. The interaction of these concepts will also be considered.

The Demand for Medical Care

While the demand for medical care has been viewed solely as the demand for the control and/or management of actual or potential diseases, health care economists point out that the demand for medical care is a derived demand.[4] Medical care is not the direct utility producing entity for most people. It is one of many inputs used to produce good health. Other key variables include diet, exercise, housing, environment, genetic endowment, and education.[5]

Good Health as a Capital Stock

Good health is similar to a capital stock.[6] The above "investment activities" increase the stock while aging, bad consumption habits, and a failure to invest will ultimately decrease the size of the stock. The stock of good health provides direct utility from feeling well and indirect utility by increasing the amount of healthy time available for other activities.[7] The stock is subject to depreciation whose rate eventually increases with age. To the extent that the individual has some control over the input to the key variables, one can argue that the individual chooses a health stock that maximizes utility, given the physical and economic constraints existing at the time.

The physical and economic variables interact to determine the rate of return on investment in the health stock. As aging increases the depreciation rate on the stock, a larger current investment is necessary to maintain a given health stock. Ceteris paribus, one would expect people to choose a lower stock as they age.

Education increases income by increasing the productivity of working time. The increase in income enables the person to invest in better housing and possibly a better environment. In addition, education makes the individual more aware of destructive consumption habits and enhances one's ability to combine health-promoting inputs efficiently.

Culyer argues that higher income persons will be healthier, even if all other factors are held constant. To the extent that higher incomes are correlated with higher real wage rates, the value of healthy time will increase. As time used in the production of good health increases in cost, individuals will have an incentive to devote less time to this activity. But the benefits from the investment have also increased and Culyer argues, on balance, that the increase in benefits should outweight the increase in costs.[8]

The Demand Process

The capital stock approach emphasizes that medical care is only one form of investment in good health. In order to analyze the demand for medical care, the degree of complementarity and substitutability with other inputs must be controlled for. Likewise, medical care is made up of many components (treatments) which allow various degrees of substitution. For example, to analyze the demand for hospital in-patient treatment of a disease, one must consider the other possible treatments of the disease. In order to move from the general investment in good health to the demand for specific forms of medical treatment, Feldstein has developed a model of the demand process.[9] Feldstein's model of the demand for medical care has three phases. The first phase involves the patient, the second the physician, and the third the demand for the various forms of medical care. It emphasizes the importance of the interaction between the patient and physician in the demand process.

The Patient. Feldstein divides the variables affecting the patient's demand for medical treatments into three groups. The first group is incidence of illness. The frequency and type of illnesses one experiences shapes the individuals perception of his need for medical care. Whether this perceived need is translated into a demand depends on the interaction of illness with the other variables.[10]

The second group of variables are cultural-demographic factors. Feldstein describes these factors as the physiological condition, perception of illness, and attitudes toward seeking medical attention.[11] Since these variables cannot be measured directly, Feldstein suggests that population characteristics be included as proxies. These proxies include age, sex, marital status, family size, education and residence.[12]

The third group of variables are the economic factors. These economic factors include the prices of the related forms of treatment (including the specific treatment), the patient's income, and the presence of health insurance.[13] These variables interact to determine if the patient demands care and what type of care is demanded.

<u>The Physician</u>. The second phase of the demand process involves the physician. Feldstein groups the factors affecting the physician's use of the various components of medical care into four groups: (1) patient characteristics; (2) institutional arrangements; (3) physician's knowledge; and (4) the relative costs of various treatments to the physician.[14] In this agency relationship, both the patient's interests and the physician's interests are accounted for. Groups (2) and (4) emphasize that factors other than the patient's interests may dilute the agency relationship. The physician's knowledge provides a constraint or the choice set of treatments offered the patient.

<u>Derived Demand</u>. The third phase of the demand process covers the actual derived demands for the various components of treatment, e.g.,

hospital care, physician care, referrals to specialists, etc.[15] The framework then shows how patient characteristics are translated into derived demands, subject to the distortions of institutional relationships, the physician's self-interest, and limitations of available treatments.

The Need for Medical Care

The economics literature contains several definitions of need. A representative definition, presented by Williams, states "a need for medical care exists when an individual has an illness or disability for which there is an effective and acceptable treatment or cure."[16] As long as a treatment yields either physical or psychological benefits, a need exists. A need does not exist if there is no treatment that provides either physical or psychological benefit.

While most physicans might agree with Williams' defintion Culyer finds fault with Williams' definition because it does not deal with the relative costs and benefits of treatment.[17] Williams' definition is based on a subjective third party evaluation of benefits only. Someone must decide if there are any positive benefits from treatment. Culyer argues that to ignore the costs in relation to the benefits is to avoid the resource allocation decision that is involved. According to Culyer, if the costs to society outweight the benefits, a need may not exist in the sense that it justifies an allocation of scarce medical care resources.

Most other definitions of need also involve a judgment by a third party, based on current medical knowledge, that someone should receive medical care.[18] The medical expert is usually the third party chosen, although Culyer and Williams both point out that society also should have a say.[19] In fact, serious resource allocation problems may result if the society does not participate in the decision or provide social and ethical guidelines for the medical expert to follow. In the absence of these guidelines, the medical expert may make decisions on technical medical criteria without considering the costs and benefits. If this happens, too many resources may be devoted to medical care in relation to other uses.

Jeffers, Bognanno, and Bartlett support this view.[20] They start their argument by distinguishing between need, a quantity determined by the medical experts, and wants, the total quantity of medical care demanded by consumers at a zero price. They conclude that consumer ignorance will generally prevent the quantity of medical care wanted from equaling the quantity needed. Furthermore, while wants may exceed needs for some persons, wants will generally fall short of needs for the popoulation as a whole.[21] Figure 4 adapts Jeffers et al.'s discussion to the NHS. In the current time period, medical care consumers will demand OW units of medical care if the price of medical care is zero. If medical care experts have their wishes, ON units will be allocated and the market price will fall to zero. Consumers will demand OW units. The remainder (WN), representing unmet needs, would not be used.[22] The demand curve (DD) represents society's total demand for medical care with the existing level of consumer ignorance. Due to competing wants, society will be unwilling to satisfy all quantities demanded at a zero price. Consequently, the vertical supply curve (SS) represents the quantity of medical care society will provide. At this level of provision, the market clearing price will be P_0. The quantity of

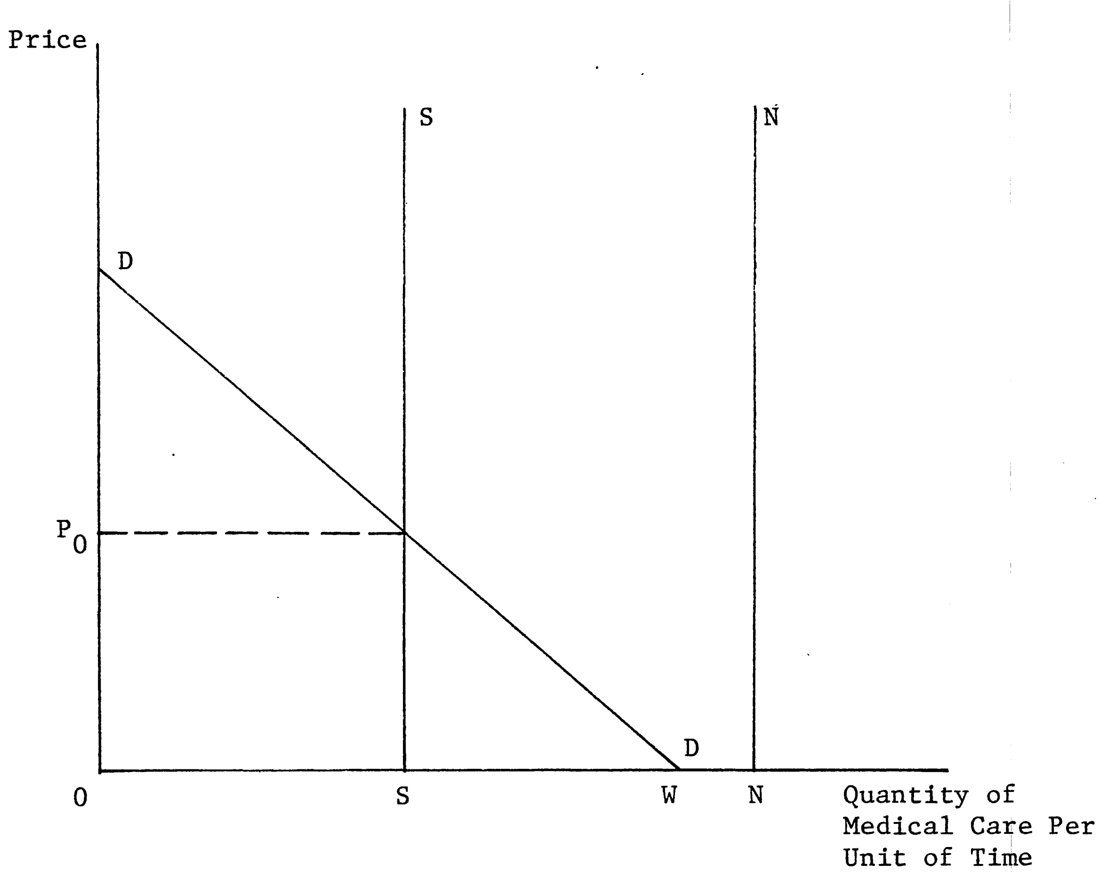

Figure 4. Demand Versus Need for Medical Care

unmet need will increase from WN to SN, since some needy consumers are priced out of the market.

Cooper argues that the NHS was founded on fundamental misconceptions concerning medical need.[23] First, the level of need (ON) was thought to be a finite quantity that could be determined with great certainty. While the founders of the NHS knew that society could not currently provide (ON) units of medical care, they believed that the quantity needed, being finite, could be reduced each year until the pre-NHS backlog was eliminated. Eventually, the quantity needed would fall below the quantity provided (OS), allowing the government to reduce the size of the NHS. By setting the money price of medical care equal to zero, the founders believed no one in need would be deterred from demanding care. Excess demands in the years of the NHS would be satisfied on a needs priority basis as resources become available until all demands which represented needs could be satisfied in the current year. But need is not a finite quantity. The quantity of need is a function of medical knowledge. As medical knowledge advances, more and more treatments exist which provide some positive marginal benefits. In fact, advances in medical technology may actually increase needs by increasing our life spans. For example, persons who previously died of pneumonia now die of cancer, a much more costly disease. This first misconception, then, is that society could afford a level of provision that would eventually reduce medical needs to a quantity that could be satisfied with current period allocations.

The belief that medical need could be determined with great certainty was another misconception. Uncertainty in diagnosis also exists. Consequently, it is difficult to rank needs on a priority basis. Even if society could allocate sufficient resources to reduce the absolute level of need faster than medical knowledge expanded it, doctors would still have to allocate scarce resources among competing demands. For example suppose the current level of provision (OS) equals the total estimated need of the population. Assume total demand still equals DD in spite of this reduction in need. If doctors cannot determine all needs with certainty and rank them accordingly, how can society be sure that all medical needs are satisfied? Uncertainty will also increase the chance that some consumers (not necessarily hypocondriacs) will demand care that is not needed and doctors will provide care to some persons with lower priority needs than others who are in greater need but are priced out of the market.

Allocating medical care on the basis of medical need, at best, then is allocating medical care resources among those who demand care on a needs priority basis. To the extent that uncertainty motivates doctors to provide treatment or referrals, when in doubt, doctors may provide rationing on a needs-priority basis for only easily diagnosed illnesses. Other diseases, for which the diagnosis and treatment are not so obvious, may be allocated resources on some other basis.

Resource Allocation and Time Prices

As discussed in the previous section, the founders of the NHS has a very naive view of medical needs. This led to another serious mistake. They granted NHS doctors clinical freedom.[24] Clinical freedom meant that doctors resource allocation decisions were not subject to accountability in terms of their efficient use of resources. No standardized concept of need could be imposed on the individual doctor.[25] The doctor's choice of treatments was a matter of concern to his patient only. Consequently, NHS managers had no assurance that the NHS's relatively scarce resources would be allocated on a needs priority basis.

Another mistake was to believe that abolishing money prices for most forms of medical care would ensure that all those who needed care would demand care. By abolishing money prices, the founders were implicitly adopting resource allocation by time prices, subject to the individual doctor's non-uniform views of relative medical needs.[26] The following sections first consider the role of time in allocating scarce resources and then discuss the rationale for favoring time prices for allocating medical care resources.

The Role of Time in Resource Allocation

Becker's work extended the theory of consumer behavior beyond the allocation of money income to maximize individual utility functions.[27] His theory also explains how individuals allocate their time among different activities in the process of utility maximization. Treating households as producers of utility-producing commodities, Becker's theory predicts that households will combine inputs of market goods and time according to the cost minimization rules of the theory of the firm.[28] The utility maximizing quantities of these commodities are determined by maximizing the household's utility function, subject to the relative prices of the commodities in the utility function and a total resource constraint.

Commodity prices in Becker's theory consist of two components: per-unit cost of the market goods used and the cost of the time requires to produce a unit of the commodity. For the commodity health care, then, the price can be represented by the following equation:

$$P_H = (P_m * M) + (W * t_m) \quad (1)$$

where P_H = the full commodity price of a unit of health care,

P_m = the price of a unit of medical care,

M = the quantites of medical care,

t_m = time required to produce a unit of health care, and

W = the foregone earnings rate of the household.

Beckers calls the market goods component $(P_m * M)$ the direct price and the time component $(W * t_m)$ the indirect price.[29]

The consumer's resources consist of money income (market goods) and time. Money income consists of earned income $(W * T_w)$ and unearned income (V). Time (T) consists of work time (T_w) and consumption time (T_c). Becker combines these two resources into a concept of total income equaling the sum of unearned income (V) plus the total value of the consumer's time (W * T).[30] Equation (2) represents the consumer's total resource constraint on utility maximization:

$$V + (W * T) - (P_{zi} * Z_i) = 0 \quad (2)$$ [31]

where P_{zi} and Z_i represent the full prices and quantities of the individual commodities. The constraint stresses the point that time not used in the production of earned income is used in some form of consumption. Increases in the foregone earnings rate will increase the value of time and change relative commodity prices while changes in unearned income leave relative commodity prices unaffected.

In maximizing utility, the consumer will equate his marginal rate of substitution of the commodity health care for other utility producing commodities to the ratio of the full commodity prices. This determines the utility maximizing quantities of these commodities.[32] At the same time, efficiency in the production of these commodities requires that the consumer equate the marginal rate of technical substitution of market goods for time to the ratio of their prices.

Figure 5 indicates how consumers R and P substitute market goods (medical care) and time in the production of a given quantity of the commodity health care. Consumer R has a relatively high foregone earnings rate (R). If each consumer pays the same money price for a unit of medical care (P_m), R will choose to use relatively more medical care (M_r) and relatively less time (T_r) to produce the same utility maximizing quantity of health care (H_0) than P, whose efficient quantities are M_p and T_p respectively.[33] Given the relative productivities of medical care and time in the production of health care, shown by the slope of the isoquant H_0, economic efficiency requires that both consumers be able to substitute medical care and time in the production of health care as technology permits. Becker's analysis predicts that consumers will substitute goods intensive production techniques for time intensive ones as their foregone earnings rates increase.[34]

The Rationale for Time Prices

If society considers medical care a merit good then the equity of the distribution of medical care among the members of society is important.[35] Allocating medical care by the financial ability to pay may not provide a sufficient degree of equity in the distribution.

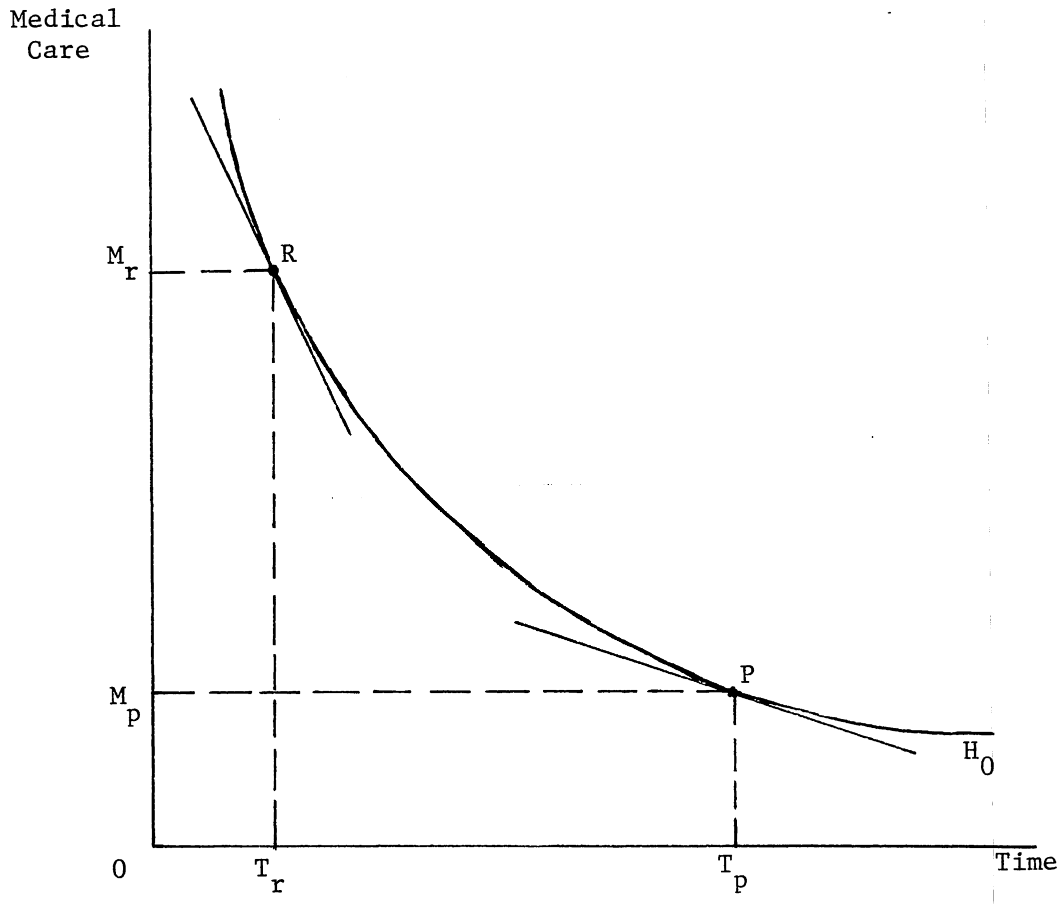

Figure 5. Technical Efficiency in Producing The Commodity "Health Care"

As shown in Chapter II, some form of subsidy may be desirable to induce target groups to consume more medical care. The question society must answer is how to increase the target population's consumption of the merit good at least cost without making these persons worse off. Generally speaking, two alternatives are available. Society can augment the income of the target group by providing income transfers or non-transferable vouchers for a certain quantity of medical care. Alternatively, the merit good can be offered to the target population at a lower (subsidized) price.

Clearly the income transfer will place the recipient on his highest possible indifference curve. The non-transferable voucher may provide an equal result but will never surpass the income transfer from the recipient's viewpoint.[36] Why, then, should society choose any other means of increasing the recipient's consumption of the merit good? Because the members of society who make the transfer have utility functions that are affected by the target groups consumption of medical care but not by the target group's utility level. And from the givers' viewpoint, the issue is to bring about the desired result at the lowest cost.

Returning to the gains from trade framework discussed in the previous chapter, redistribution of income from the givers (the R's) to the recipients (the P's) should make no one worse off while making someone better off. Faced with the political reality that some form of redistribution of wealth will take place, the R's may seek to impose their preferences on the P's in return for this redistribution. The R's, then may prefer to make the P's better off by giving them medical care as opposed to money. In a majority voting system, the redistribution forthcoming which addresses the interdependent element of the R's utility function may make the P's better off than a politically possible alternative redistribution of wealth that does not. In other words, the R's may be willing to redistribute more wealth in the form of medical care than in money income. The redistribution in kind (medical care) directly addresses the utility interdependence while a redistribution in the form of money only indirectly addresses it. The P's end up on a higher indifference curve than they would with a politically possible redistribution of income.

Figure 6 demonstrates this situation.[37] It compares P's choice between medical care (X_p) and a (numeraire) good (Y_p). For simplicity, the unsubsidized price of medical care equals the price of the numeraire, which is one. Initially, P is in equilibrium at point A. He purchases 5 units of X and 5 units of Y.

Assume that at P's current consumption level of good X, marginal social benefits exceed marginal social costs. R will be able to increase his own utility by subsidizing P to consume more good X. Suppose, given the price of medical care, marginal benefits equal marginal social costs when P consumes 20 units of good X. R can motivate P to conume 10 units of good X in two ways. He can purchase 10 units of X for P. This will move P to point D.[38] As a second approach, he can lower P's price of good X to the point where the ratio of the subsidized price of good X to the price of good Y just equals P's marginal rate of substitution between the two goods when P consumes 10 units of good X. The second approach will move P to point C.[39]

As the indifference curves indicate, P is unquestionably better off if R purchases 10 units of good X for P. Given P's preferences, he will

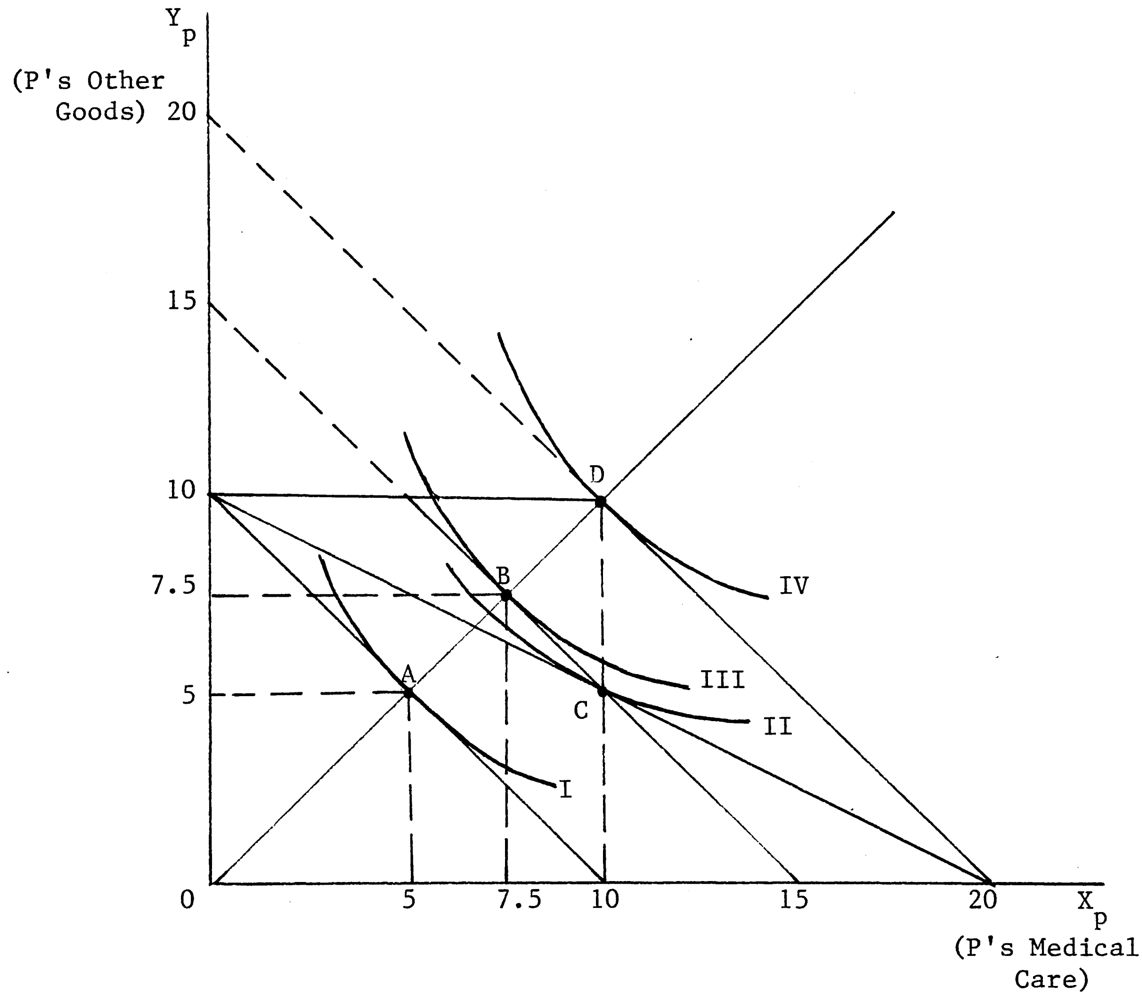

Figure 6. Per-Unit Consumption Subsidies for Medical Care

reduce his own purchases of good X to zero and spend his income on good Y. This places him on indifference curve IV.

But R does not look at it this way. R is concerned about the quantity of good X P consumes. How can he motivate P to consume 10 units of medical care (good X) while minimizing the redistribution of wealth that will occur? The first approach, which moves P to point D, requires a transfer of 10 units of wealth, since the price of medical care equals the price of the numeraire. The second approach involves reducing the price of medical care to P by one-half. This costs R five units of wealth and moves P to point C. P is clearly not as well off in this case but is better off than at point A.

One might ask another question. Why not transfer five units of good X to P as opposed to subsidizing the price P pays for good X? This transfer will move P from point A to point B on indifference curve III, a higher level of satisfaction than point C. The generalized transfer of good X will make P better off at no additional loss of wealth to R. But this approach places R on a lower indifference curve than when P's equilibrium is point C. Why? Because P consumes only 7.5 units of medical care at point B. R will gain utility as P increases consumption from 7.5 to 10 units of medical care at no additional cost to R.

As discussed in Chapter II, the subsidy method is appropriate when the absolute quantities of P's medical care consumed enter R's utility function. But Lindsay and Buchanan have noted that equality in the distribution of medical care may be the relevant interdependency. If so, Lindsay has shown that the subsidy method alone is not the least cost approach. The least cost approach requires both a subsidy to the P's and some form of rationing for the R's.

Nichols, Smolensky, and Tideman have shown that waiting time can play an important role in achieving the above objective.[40] Noting that merit goods such as medical care are often offered to the P's at zero money prices, they further point out that the quantity supplied falls short of total demand at a zero money price. In order to ration the good among demanders, the providing authority allows waiting times to increase to eliminate excess demand. This method has the advantage of directly discriminating by wage rates without imposing administrative costs on the providing authority. If wage rates are highly correlated with income, and the demand for medical care is higher for higher income individuals, time prices may actually increase the equality of the distribution of medical care between income groups. Now the price of medical care to each consumer equals the person's foregone earnings rate (W) multiplied times the demander's time input (t_m).[41] Those with higher incomes and higher foregone earnings rates will pay a higher time price for a given time input. The higher price will decrease the quantity of medical care demanded. On the other hand, those with lower incomes and lower foregone earnings rates will pay lower time prices for a given time input. This will increase the quantity of medical care they demand.

Time Prices, Money Prices, and Equity in the Distribution of Medical Care

Becker's theory provides a way to determine under what conditions zero pricing will increase equity in the distribution of medical care. According to Becker, the full price of consumption is the sum of the direct (money) and indirect (time) price and these direct and indirect prices are "symmetrical determinants" of total price. To emphasize this symmetry, Becker presents the following example which applies to the zero-money price provision of medical care.[42] Assume the commodity "health care" (H) is initially produced with medical care (M) and zero consumer time (t_m), so that the money expenditure for medical care (P_m * M) represents the full commodity price for H. If the money price of medical care (P_m) is reduced to zero but producers are subsidized to provide the same quantity of M as before, the price of H will not fall to zero. Holding the demand for H constant, relatively scarce M must now be rationed by time prices. Since the total quantity of M supplied is held constant, the full price of H is a time price which, given the average foregone earnings rate of health care consumers (W) equals (W * t_n). This, in turn, equals the price of H when the money expenditure for M reflected H's full price (P_m * M). In other words, the price of H changed from solely a money price to solely a time price but the full price of H remained constant.

To determine the conditions necessary for a time pricing system to increase the equality of the distribution of medical care among persons of equal medical need, consider a simple society consisting of two persons, R and P. Each have different foregone earnings rates, which are W_r and W_p respectively. Assume both persons have equal medical needs and that R's and P's independent utility maximization yields demand curves, D_r and D_p respectively. Further assume that both persons have equal time input requirements (t_m) for obtaining a unit of medical care. In addition, and for simplicity, assume one unit of medical care is used to produce a unit of health care.[43]

Figure 7 compares resource allocation with money prices and time prices and their respective abilities to increase the equality of the distribution of medical care between R and P, given the above assumptions. Following Becker's example, assume medical care is initially allocated between R and P by a single money price only. Although R and P have equal medical needs, individual differences in other determinants of demand lead R to demand a larger quantity of medical care than P for any given price.[44] Given the single money price (P_m), R will demand Q_r units while P will demand Q_p units of medical care. The total quantity of medical care supplied equals ($Q_p + Q_r$).

Now suppose the suppliers of medical care are subsidized to produce the same total amount and the money price of medical care is reduced to zero. At a zero money price, R and P now demand Q_{p_0} and Q_{r_0} units respectively. Since their total quantities demanded now exceed the subsidized level of provision ($Q_r + Q_p$), the time required to obtain a unit of medical care rises from zero until the cost of time to each individual constrains their total quantities demanded to within ($Q_r + Q_p$). The distribution of the medical care provided is now determined by the time prices R and P pay, which are ($P_{t_r} = W_r * t_m$) and ($P_{t_p} = W_p * t_m$) respectively.[45] As drawn, the differences in these time prices completely offset the other differences in the determinants of demand and both individuals consume Q_e units of medical care.[46]

In this simple two person world, the NHS equity objective is met. Both R and P, who have equal needs, demand the same amount of medical care. In this instance, time prices decrease the differences in R's and P's quantities demanded compared to a single money price for both. While the same result could have been achieved by charging R and P different

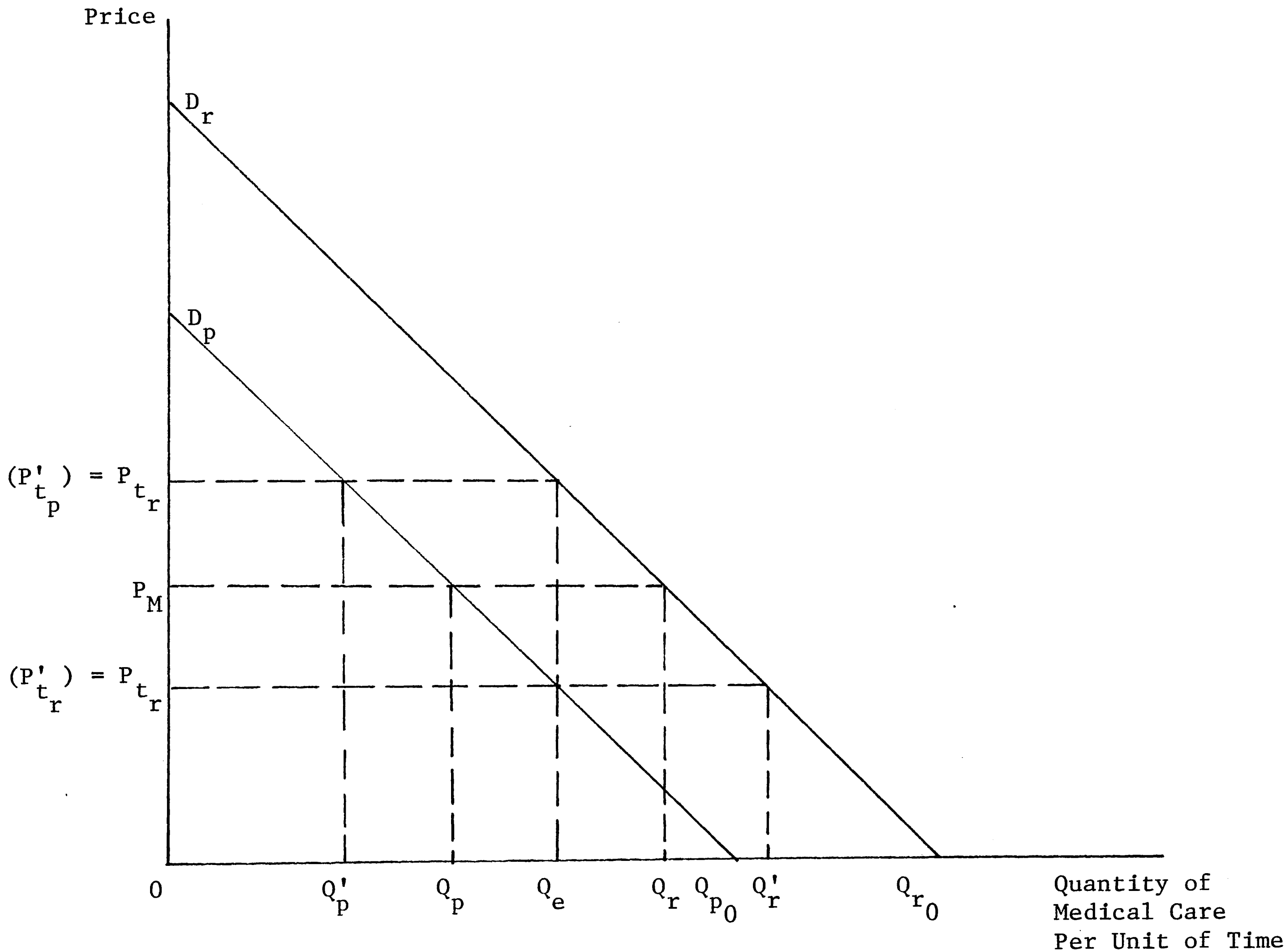

Figure 7. Time Prices, Money Prices, and Equity in the Quantity Demanded of Medical Care

money prices equal to P_{t_r} and P_{t_p}, respectively, the NHS would incur administrative costs in doing so.

In general, a pricing scheme will support the NHS's objective if it reduces the differences in quantities demanded between persons of equal medical needs. Up to a point, time prices which vary directly with the demand for medical care after controlling for medical need will provide greater equality than a single money price charged to all. However, if time prices increase more rapidly than demand, beyond some limit they may actually bring about greater inequality than a single money price.[47]

Figure 7 can also be used to describe a second case where time prices will increase inequality as opposed to decreasing it. Suppose foregone earnings rates vary inversely with the demand for medical care. A single money price led R to demand Q_r units while P demanded Q_p units. If P's foregone earnings rate (W'_p) is greater than R's (W'_r), P's time price ($P'_{t_p} = W'_p * t_m$) will now exceed R's ($P'_{t_r} = W'_r * t_m$). Faced with these time prices, R will now consume Q'_r while P will consume Q'_p. The inequality of quantities demanded among these persons of equal need increases.

Another potential source of increased inequality from time prices concerns the composition of R's and P_o's incomes. For illustration, assume R and P have identical total money incomes. In addition, all other determinants of demand are identical, giving R and P the same demand curve for medical care (D_r). The only difference between R and P is in the composition of their total incomes. Assume P's proportion of unearned income is larger than R's. For this to be the case when work times for both individuals are assumed equal, R's foregone earnings rate must exceed P's. Given that both individuals must invest the same amount of time to obtain a unit of medical care, R's time price (P_{t_r}) will exceed P's (P_{t_p}). Under a time pricing scheme, R will demand Q_e units while P will demand Q'_r units.[48] Alternatively, under a single money price scheme, both R and P would demand Q_r units.

If NHS doctors are likely to allocate a significant portion of scarce medical care resources on demand when medical needs are difficult to identify and rank on a priority basis, time prices will increase equity (within limits) between income groups when time prices vary directly with income (and demand).[49] On the other hand, time prices will increase inequality within the same income group if foregone earnings rates, and consequently time prices, vary among members of the same income group. Working poor persons will end up consuming less medical care than poor persons receiving some form of income maintenance.

ENDNOTES

[1]Michael H. Cooper, Rationing Health Care (New York, 1975), p. 8.

[2]For example, in 1971, paying inpatients accounted for 2.17 percent of all NHS discharges and deaths. "The three most popular specialties for private practice by NHS consultants were general surgery, gynecology, and orthopaedic surgery, the three specialties with the longest NHS waiting lists." This indicates that an alternative to public hospital care and its large waiting lists exists. See A. J. Culyer and J. G. Cullis, "Private Patients in H. H. S. Hospitals: Waiting Lists and Subsidies," in The Economics of Health and Medical Care, Mark Perlman (ed.) (New York, 1974), pp. 109-110.

[3]Paul J. Feldstein, "Research on the Demand for Health Services," Milbank Memorial Fund Quarterly, Vol. 44, No. 3, Part 2 (July, 1966), p. 130.

[4]Ibid., p. 132.

[5]A. J. Culyer, Need and the National Health Service (London, 1976), p. 24. Chapter 3 also contains a good summary of the "capital stock approach" in studying the demand for medical care.

[6]Michael Grossman, "On the Concept of Health Capital and the Demand for Health," Journal of Political Economy, Vol. 80 (March/April, 1972), pp. 223-255.

[7]Culyer, 1976, p. 24.

[8]Ibid., p. 25.

[9]Feldstein, 1966, pp. 133-139.

[10]Ibid.

[11]Ibid., p. 143.

[12]Ibid.

[13]Ibid., pp. 146-147.

[14]Ibid., p. 139.

[15]Ibid.

[16]A. Williams, "'Need'--An Economic Exegesis," in Economic Aspects of Health Services, p. 32.

[17]Culyer, 1976, p. 14.

[18]Ibid., p. 15.

[19]Ibid., p. 16, and Williams, p. 34.

[20]J. R. Jeffers, M. F. Bognanno, and J. C. Bartlett, "On the Demand Versus Need for Medical Services and the Concept of Shortage," American Journal of Public Health, Vol. 61, No. 1 (January, 1971), p. 48.

[21]Ibid.

[22]This amount of medical care resources remains unused due to consumers' ignorance of their medical needs.

[23]Cooper, 1975, pp. 8-9.

[24]Michael Cooper, "Economics of Need: The Experience of the British Health Service" in The Economics of Health and Medical Care, Mark Perlman (ed.) (New York, 1974), p. 90.

[25]Ibid., p. 99. Other work has been done on the problems of implementing a standardized definition of need. For a summary, see Culyer, 1975, pp. 30-48.

[26]Ibid., pp. 97-99.

[27]Gary Becker, "A Theory of the Allocation of Time," The Economic Journal, Vol. 75, No. 299 (September, 1965), pp. 493-517.

[28]Ibid., p. 516.

[29]Ibid., p. 500.

[30]Ibid., p. 497. Becker notes that this amount may overstate actual total income if the consumer's foregone earnings rates vary, e.g., the rate would drop after working hours.

[31]Ibid.

[32]The analysis assumes the second order conditions for utility maximization hold.

[33]For this analysis, assume that differences in other determinants of demand offset the differences in R's and P's foregone earnings rates. More generally, foregone earnings rates vary directly with income, R will have a larger utility maximizing quantity of health care than P, ceteris paribus.

[34]The effect of an increase in money income on the total quantity of health care demanded depends on two things. First, an increase in unearned income will increase demand if health care is a normal good and decrease demand if it is an inferior good. Second, an increase in the foregone earnings rate will increase the demand for health care if health care is relatively less time intensive than other commodities or can be produced with less time and more market goods (medical care).

Holtmann worked out the comparative statics of Becker's theory using a two-commodity model containing health care and other commodities. He found a "dual substitution effect" at work because a change in wages (foregone earnings rate) changes all relative prices. The net effect of these dual substitution effects depends on what happens to the price of health care relative to other commodities. If the relative price of health care increases, consumer's substitute away from health care. For the quantity of health care demanded to decrease as wages increase, health care must be either (1) an inferior commodity that is sufficiently time intensive or (2) a normal commodity that is sufficiently more time intensive for the negative substitution effect to overcome the positive income and substitution effects. For a more detailed presentation, see A. G. Holtmann, "Price, Time, and Technology in the Medical Care Market," Journal of Human Resources, Vol. 7 (Spring, 1972), pp. 179-190.

[35]For a general discussion of the concept of a "merit good", see Richard Musgrave, The Theory of Public Finance (New York, 1959), pp. 13-14.

[36]Richard H. Leftwich, The Price System and Resource Allocation, 6th Edition (Hinsdale, Ill., 1976), pp. 95-97.

[37]For a more detailed and more general analysis of consumption subsidies, the work on which this simple example is based, see Mark V. Pauly, "Efficiency in the Provision of Consumption Subsidies," Kyklos, Vol. 23, No. 1, pp. 33-57.

[38]The indifference curves are drawn so that both income and lump sum vouchers place P on the same indifference curve IV. This corresponds to point P4 in Figure 1 of Pauly's analysis. Ibid., pp. 42-43.

[39]Ibid., p. 45. This point corresponds to point P5 in Figure 1 of Pauly's analysis.

[40]D. Nichols, E. Smolensky, and T. Tideman, "Discrimination by Waiting Time in Merit Goods," American Economic Review, Vol. 61, No. 2, Part 1, pp. 312-323.

[41]Becker, p. 497. The market goods (direct) price term drops out of the expression for the commodity price.

[42]Ibid., pp. 515-516.

[43]For simplicity, this means the price of health care (P_H) will equal the price of medical care (P_m).

[44]In this example, the foregone earnings rates directly related to the magnitude of the other determinants of demand, such as income and education.

[45]Since real supply and real demand are held constant, the average time price $((P_{t_r} + P_p)/2)$ equals the average foregone earnings rate of both persons (W) multiplied by the time input required (t_m). Becker's example indicates this average time price must equal the original equilibrium price (P_m).

[46]Note that the quantity Q_e must equal one-half of the total care provided, given the individual demand curves and time prices.

[47]Using linear demand curves where R's demand is $(R_r = A - b * P_{t_r})$ and P's demand is $(Q_p = C - d * P_{t_p})$, time prices will result in greater inequality if $(P_t > P_{t_p})$ and $(P_{t_r}(3b-d)) - (P_{t_p}(3d-b)) > 4(A - C)$. Given that $(A > C)$, P's quantity demanded under a time pricing scheme will exceed R's quantity demanded by more than R's quantity demanded exceeded P's under a single money-pricing scheme.

[48]The limit on total provision $(Q_r + Q_p)$ does not apply to this example.

[49]This conclusion assumes the demand for medical care is directly related to income.

CHAPTER IV

EQUITY, NEED, AND THE REGIONAL ALLOCATION OF NHS RESOURCES

In Chapter III, equity, at the micro level, was defined as providing two individuals of equal medical need the same quantity of constant-quality medical care. On a broader regional basis, equity in resource allocation would require allocating equal amounts of medical care to regions with equal medical need. This implies that regions with unusual need should be allocated resources in direct proportion to their relative need.

In analyzing the issue of equity on a regional basis, this chapter has two objectives: first, to evaluate the actual allocation of NHS resource inputs among regions in accordance with indices of regional need, with primary attention given to non-psychiatric hospital in-patient services; and second, to examine regional hospital medical care output from the same perspective. Three indices of a region's medical need will be used in the evaluations. First, following the work of Cooper and Culyer, regional resource allocations are evaluated on the basis of equal per-capita allocations.[1] Next, to extend the analysis, two additional measures of need are considered. These measures, the region's age-sex adjusted population, and its age-sex-standardized mortality ratio (SMR) adjusted population, represent an attempt by Britian's Labor Party to develop an operational index of a region's medical need.[2]

The chapter is organized in the following way. First, a brief discussion of the three indices of need are presented. Second, using the three indices, NHS resource allocations across regions are evaluated for 1974 through 1977, the years following the reorganization of the NHS.[3] When data permits, these results are compared with Cooper and Culyer's 1966-1967 findings. Third, regional medical care output is evaluated for the same years, again, using the three indices of medical need.[4]

Three Measures of a Region's Medical Need

Three indices of a region's medical care need are used to evaluate NHS performance on an equity basis. The first measure of need is the region's crude population. The second measure is the region's crude population adjusted for its age-sex structure. The third measure of need is the region's age-sex adjusted population adjusted for the standardized mortality ratios (SMR) of the age-sex groups.

Equal Per-Capita Allocations Based on Crude Populations

Cooper and Culyer used this index of need to evaluate the NHS's performance in meeting its equity objective using 1966 and 1967 data.[5] In their view, this measure of need represents "the equality of opportunity of all people to consume the health services it has been decided they need, in the sense of physical provision of the services in a geographical area."[6] While demand-side variables make achieving real equality of opportunity essentially unattainable in their view, Cooper and Culyer believed this measure to be an objective and reasonable yardstick for evaluating the NHS's performance.

Equal Per-Capita Allocations Based on Age-Sex Adjusted Populations

While the preceeding measure may be a preferable equity criterion in the sense that all persons are given equal weight, medical needs are not equal across the population. Since expenditures on hospital care account for the major part of the budget and utilization of hospital in-patient care varies by age and sex, adjusting the crude population for expected utilization rates seems justified. This index of need takes into account the impact on hospital utilization of the age-sex structure of the regions' populations.

The age-sex adjusted population is calculated in the following way. First, the expected bed-days of hospital care for each age-sex group are computed and then summed over all age-sex groups in the region. Next, all regions' expected bed-days are summed to obtain a total for England. Third, the current total population of England is apportioned in proportion to the ratio of a region's expected bed days to the total expected bed days for England.[7]

Equal Per-Capita Allocations Based on Age-Sex SMR Adjusted Populations

The age-sex adjusted populations provided the starting point for developing this index of need. The age-sex categories used to determine utilization rates are divided into 17 International Classification of Diseases (ICD) conditions to account for variation in bed utilization rates by condition. For each ICD classification, the region's standardized mortality ratio (SMR) is multiplied by the age-sex weighted population for that condition. The results are then aggregated by region and scaled, as before, to match England's total population.[8]

Indices of Need and Regional Resource Input Allocations

The chapter considers the distribution of NHS resources among regions (RHA's) on the basis of medical care need from two perspectives. First, NHS expenditures per capita are compared with our indices of medical care need. Second, actual manpower and hospital bed distributions among RHA's are compared with the same indices of need.

NHS Expenditures Per-Capita

The Office of Health Economics provides two types of hospital expenditure data for RHA's.[9] Net revenue expenditures refer to current operating expenses. Capital expenditures refer to expenditures on plant and equipment. Both categories of expenditure are discussed within the context of each measure of medical care need.

Per-Capital Expenditures and Crude Population as an Index of RHA Need. Table I presents two measures of NHS performance concerning the regional allocation of resources using crude population as the index of regional need. These two measures are the coefficient of variation and the ratio of the highest RHA value to the lowest RHA value. The coefficient of variation equals the standard deviation of the RHA values divided by the mean of the RHA values. It is a measure of the overall equity of the distribution and is expressed as a percentage of the mean.

TABLE I

EQUITY OF NHS RHA EXPENDITURES PER CAPITA BASED ON
THE CRUDE POPULATION NEED MEASURE

Measures of Equality	Year				% Change
	1974	1975	1976	1977	1974-1977
Net Revenue Expenditure Per Capita					
Coefficients of Variation					
All RHA's	16.30	16.10	15.50	14.70	-9.81
Non-London RHA's	7.90	7.30	6.70	6.70	-15.18
London RHA's	3.90	3.40	3.40	3.80	-2.56
Ratio of High to Low					
All RHA's	1.60	1.55	1.52	1.46	-8.75
Non-London RHA's	1.33	1.29	1.25	1.23	-7.51
London RHA's	1.08	1.07	1.08	1.08	0.00
Capital Expenditure Per Capita					
Coefficients of Variation					
All RHA's	17.10	17.00	19.70	26.40	54.38
Non-London RHA's	19.90	14.90	19.80	25.40	27.63
London RHA's	4.30	24.30	21.80	22.90	432.55
Ratio of High to Low					
All RHA's	1.82	1.69	1.84	2.47	35.71
Non-London RHA's	1.82	1.61	1.82	2.47	35.71
London RHA's	1.10	1.63	1.66	1.55	40.90

Perfect equality produces a coefficient of variation of zero. The second measure, the high-low ratio, provides a measure of the difference in relative position of the "best-advantaged" RHA to the "least-advantaged" RHA. This measure would equal one with complete equality.

Table I presents the two measures of equity for three RHA groups: non-London RHA's, London RHA's, and all RHA's. First consider the coefficients of variation. Between 1974 and 1977, the coefficient of variation for net revenue expenditures per capita (NREPC) declined 15.2 percent for non-London RHA's and declined only slightly for the London RHA's. The combined result was a 9.8 percent decrease for all RHA's. For capital expenditures per capita the coefficients of variation increased 27.6 percent for non-London RHA's and 432.6 percent for the London RHA's. The combined result was a 54.4 percent increase for all RHA's.

While this latter change could represent a worsening of the regional distribution of capital resources, it could also represent a correction of past regional inequity if the capital expenditures were made in capital-shortage regions. To find support for either case, the correlation coefficients between non-psychiatric hospital beds per 1,000 population (NPB's) and capital expenditures per capita (CEPC) were calculated for each RHA group for 1974 through 1977. Only in 1974 was there a statistically significant negative correlation between NPB's and CEPC. In that year, there was a negative 0.58 correlation between NPB's and CEPC for non-London RHA's, statistically significant at the 5 percent level. Thus, the correlation analysis provides no strong support for the proposition that the increases in the coefficients of variation were the result of attempts to redress past RHA bed inequities through the current distribution of capital expenditures.[10]

Table I also shows that the ratios of high RHA values to low RHA values tell the same story. The ratios for NREPC decreased 7.5 percent for non-London RHA's and remained approximately constant for the London RHA's. The combined was an 8.8 percent decrease for all RHA's. At the same time, the ratios for CEPC increased 35.7 percent for non-London RHA's and 40.9 percent for London RHA's.

In conclusion, the data from 1974 through 1977 indicate some increase in equality in terms of current operating expenditures. At the same time, there was an increase in inequality in terms of capital expenditures per capita. While these results indicate only small improvement overall, the data reflect a period when both net revenue expenditures per capita and capital expenditure per capita did not keep pace with inflation. Improvements in equality are much more painful in periods when the real budget is declining. To make a disadvantaged RHA better off, a more advantaged RHA must incur a larger real budget cut.

Per-Capita Expenditures and the Age-Sex Adjusted Population as an Index of RHA Need. Correlation coefficients provide a simple and useful way to evaluate the NHS's performance concerning the equitable regional distribution of resources when one uses other indices for need, e.g., the age-sex adjusted population ratio (ASPOPR). Table II presents the simple correlations between NREPC and CEPC with ASPOPR. As the table indicates, there is very little simple correlation between NREPC and ASPOPR for non-London RHA's. The coefficients are very small and not statistically significant. In fact, only the 1974 correlation coefficient for the London RHA's is statistically significant (5 percent level) and it carries the "wrong" sign. Consequently, one can conclude that the variation of NREPC's among RHA's does not reflect the regions' age-sex adjusted populations.

TABLE II

SIMPLE CORRELATION COEFFICIENTS BETWEEN RHA NET REVENUE AND CAPITAL EXPENDITURES PER CAPITA AND THE AGE-SEX ADJUSTED POPULATION NEED MEASURE

Measures of Equality	Year			
	1974	1975	1976	1977
Net Revenue Expenditure per Capita				
Correlation Coefficients				
All RHA's	0.24	0.25	0.28	0.28
Non-London RHA's	0.00	-0.05	-0.02	0.03
London RHA's	-.92**	-0.77	-0.41	-0.52
Capital Expenditure per Capita				
Correlation Coefficients				
All RHA's	-0.32	0.01	-0.04	-0.15
Non-London RHA's	-0.37	-0.37	-0.21	-0.11
London RHA's	0.43	0.67	0.59	0.31

*Statistically significant at the 10 percent level.

**Statistically significant at the 5 percent level.

***Statistically significant at the 1 percent level.

To determine if any efforts to allocate net revenue expenditure among regions to reflect the age-sex structure of the region's population were masked by the necessity to fund NHS facilities and personnel already in the region, the partial correlation coefficients between NREPC and ASPOPR were calculated for each year. These coefficients represent the correlation between NREPC and ASPOPR after controlling for (holding constant) the number of hospital beds, doctors, and nurses in the region. After controlling for the effects of existing facilities and personnel, one finds statistically significant positive correlations between NREPC and ASPOPR for all RHA's combined for all years and for non-London RHA's in 1974 and 1977.[11] These correlations indicate some recognition of this need measure. Again, in this period of declining real expenditures, this is an indication that the NHS may have been attempting to use an operational measure of need other than the crude population or was using some other allocation scheme which is highly correlated with the ASPOPR measure.

Once again referring back to Table II, one finds the simple correlation coefficients between CEPC and ASPOPR. No statistically significant simple correlations between CEPC and ASPOPR exist for any RHA grouping for any year. When the partial correlation coefficients are calculated, one finds statistically significant negative correlations for non-London RHA's in 1974 and 1975. From these results, one must conclude that the distribution of capital spending among RHA's does not reflect the age-sex weighted population distribution.

<u>Per-Capita Expenditures and the Age-Sex SMR Adjusted Population as an Index of Need</u>. The analysis uses the same procedure to evaluate the regional distribution of expenditures in relation to the third index of need, the age-sex SMR adjusted population ratio (ASSMRPR). Table III presents the results. The findings indicate a statistically significant (5 percent level) simple positive correlation between NREPC

and ASSMRPR in all years for the London RHA's. While the study finds a negative correlation for the non-London RHA's, only the correlation coefficient for 1974 was statistically significant.

TABLE III

SIMPLE CORRELATION COEFFICIENTS BETWEEN RHA NET REVENUE AND CAPITAL EXPENDITURES PER CAPITA AND THE AGE-SEX SMR ADJUSTED POPULATION NEED MEASURE

Measures of Equality	Year 1974	1975	1976	1977
Net Revenue Expenditure per Capita				
Correlation Coefficients				
All RHA's	0.04	0.04	0.06	0.09
Non-London RHA's	.55*	.56*	.62*	.70*
London RHA's	-.90*	-0.79	-0.42	-0.54
Capital Expenditure per Capita				
Correlation Coefficients				
All RHA's	-.61**	-0.06	0.18	0.22
Non-London RHA's	-.71**	-0.33	0.07	0.17
London RHA's	0.47	0.69	0.63	0.35

*Statistically significant at the 5 percent level.

**Statistically significant at the 1 percent level.

The partial correlation coefficients used earlier to evaluate the ASPOPR need measure indicate a statistically significant positive correlation in 1974 and 1976 between NREPC and ASSMRPR. While these partial correlations are not as strong as the simple correlations reported in Table III, they do support the proposition that the NHS may recognize this concept of need in its resource allocation decisions.

Capital expenditures per capita and this need index show a similar pattern to the other capital expenditure comparisons. In only 1974 does one find a statistically significant simple correlation coefficient with the wrong sign for non-London RHA's. When existing NHS facilities and personnel are held constant, a statistically significant negative correlation occurs in 1975 for the non-London RHA's and for all RHA's combined. In addition, a statistically significant negative partial correlation exists for all RHA's in 1974. Again, the distribution of capital spending shows no direct relationship to the RHA's age-sex SMR adjusted population ratios.

NHS Resource Allocations Among RHA's

To evaluate the distribution of NHS facilities and personnel among the RHA's according to medical need the analysis uses five indices. They are: Nurses and midwives per 100,000 population (NMW's); Hospital doctors below consultant grade per 100,000 population (HMS's); Hospital medical consultants per 100,000 population (HMC's); general medical practitioners per 100,000 population (GP's); and non-psychiatric hospital beds available per 100,000 population (NPB's). The following narrative discusses the performance of the NHS in terms of each index of medical need.

NHS Resource Allocations and Crude Population Used as an Index of RHA Need. Table IV presents the coefficients of variation and the high-low ratios for the five indices for 1974 through 1977. The coefficients of variation for the manpower variables indicate general improvement in

TABLE IV

EQUITY OF NHS RESOURCE ALLOCATIONS BASED ON THE CRUDE POPULATION NEED MEASURE

Measures of Equality	Year 1974	1975	1976	1977	% Change 1974-1977
Coefficients of Variation					
Nurses and Midwives per 100,000 Population					
All RHA's	11.00	8.80	8.50	8.70	-20.90
Non-London RHA's	9.80	9.50	7.80	8.70	-11.22
London RHA's	4.00	2.90	2.70	2.70	-32.50
Hospital Medical Staff per 100,000 Population					
All RHA's	18.30	17.50	15.90	16.40	-10.38
Non-London RHA's	10.00	9.70	8.30	9.90	-1.00
London RHA's	16.40	14.70	13.60	13.40	-18.29
Hospital Med Consultants per 100,000 Population					
All RHA's	12.70	12.80	11.10	11.70	-7.87
Non-London RHA's	9.10	10.40	9.60	8.90	-2.19
London RHA's	10.30	8.70	6.90	7.40	-28.15
General Practitioners per 100,000 Population					
All RHA's	6.70	6.40	6.50	6.00	-10.44
Non-London RHA's	4.70	4.40	4.60	3.80	-19.14
London RHA's	3.20	3.20	3.30	3.90	21.87
Non-Psyc Hospital Beds per 100,000 Population					
All RHA's	9.4(E)	9.4(E)	10.10	10.1(E)	NA
Non-London RHA's	9.5(E)	9.5(E)	10.20	10.2(E)	NA
London RHA's	7.4(E)	7.4(E)	7.20	7.2(E)	NA

TABLE IV (Continued)

Measures of Equality	Year 1974	1975	1976	1977	% Change 1974-1977
<u>High RHA/Low RHA</u>					
Nurses and Midwives per 100,000 Population					
All RHA's	1.40	1.33	1.27	1.33	-5.00
Non-London RHA's	1.40	1.33	1.27	1.33	-5.00
London RHA's	1.10	1.06	1.06	1.06	-3.63
Hospital Medical Staff per 100,000 Population					
All RHA's	1.79	1.74	1.65	1.70	-5.02
Non-London RHA's	1.35	1.33	1.27	1.32	-2.22
London RHA's	1.47	1.38	1.36	1.36	-7.48
Hospital Med Consultants per 100,000 Population					
All RHA's	1.62	1.68	1.54	1.52	-6.17
Non-London RHA's	1.43	1.49	1.40	1.36	-4.89
London RHA's	1.25	1.21	1.17	1.16	-7.20
General Practitioners per 100,000 Population					
All RHA's	1.24	1.23	1.23	1.20	-3.22
Non-London RHA's	1.18	1.17	1.16	1.13	-4.23
London RHA's	1.08	1.07	1.08	1.09	0.92
Non-Psyc Hospital Beds per 100,000 Population					
All RHA's	1.34(E)	1.35(E)	1.37	1.38(E)	NA
Non-London RHA's	1.29(E)	1.29(E)	1.33	1.33(E)	NA
London RHA's	1.16(E)	1.16(E)	1.16	1.16(E)	NA

Note: E = Estimated; NA = Not Available.

per-capita equality in all years. For nurses and midwives (NMW's), the data show decreases of 11.2 percent for non-London RHA's and 32.5 percent for London RHA's. For hospital doctors below consultant grade (HMS's), the decrease of one percent for non-London RHA's is small while the decrease of 18.3 percent for the London RHA's is much larger, possible reflecting the relative attractiveness of the London RHA's to hospital doctors. This same pattern of improvement in the equality of the distribution occurs for hospital medical consultants (HMC's), with the coefficient of variation declining 2.2 percent for non-London RHA's and 28.2 percent for London RHA's. For general medical practitioners (GP's), the data indicate an improvement of 19.1 percent for non-London RHA's but an increase in the coefficient of variation of 21.9 percent for the London RHA's. In this latter case, the greatest increase in GP's occurred in the London RHA with the highest GP to population ratio, North West Thames.

South West Thames experienced the largest percentage increases in HMS's and HMC's during the same period. The increases in HMS's and HMC's possibly represent corrections for shortages since the RHA ranked last among London RHA's in these categories in 1974. But the changes still left South West Thames in last place, although the RHA improved its rank from third to second in terms of GP's.

Cooper and Culyer provided information on three of the above indices for 1966-1967.[12] While the reorganization of the NHS in 1974 slightly altered the boundaries, making "strict comparisons" between 1966-1967 and the 1974-1977 period impossible, a comparison between Cooper and Culyers' findings and the data for 1974-1977 provides an indication of NHS improvement in meeting this equity objective over time.[13]

Looking at our indices slightly differently, Cooper and Culyer found a coefficient of variation of 16.8 percent for population per hospital medical consultant for non-London RHA's while the high-low ratio was 2.02.[14] When HMC's are converted to population-doctor ratios, the coefficient of variation for non-London RHA's was 10.4 percent in 1974 and 9.5 percent in 1977. At the same time, the high-low ratios were 1.43 and 1.36 respectively. This comparison suggests significant improvement over time in the equity of the geographical distribution of HMC's. While one cannot attribute all of the change to an improvement in equity, it would also be difficult to attribute changes of this magnitude to the boundary changes alone.

For hospital doctors below consultant, Cooper and Culyer report a coefficient of variation of 16.86 percent for non-London RHA's. They also report high-low ratios of 1.83 for non-London RHA's and 2.69 for all RHA's.[15] After converting HMS's to population-doctor ratios, one finds a coefficient of variation of 9.6 percent in 1974, and 9.5 percent in 1977 for non-London RHA's. At the same time, the high-low ratio for non-London RHA's was 1.35 and 1.32 respectively while equalling 1.79 and 1.70 for all RHA's. Again, these data show significant improvements in equity using these measures.

A comparison of Cooper and Culyers' findings to data for 1976, the only year for which detailed RHA bed data were available, indicates no improvement in the equity of the distribution of hospital beds among regions. In 1966-1967, the coefficient of variation for non-London RHA's was essentially the same, 11.0 percent. The comparison of high to low values remains unchanged. The ratio for non-London RHA's in 1966-1967 was 1.43 and also 1.43 in 1976.[16] However, some improvement did occur in this measure for all RHA's combined, declining from 1.82 in 1966/1967 to 1.59 in 1976. The change suggests that bed reductions in the NHS occurred in the regions with the most beds. But since the coefficients of variation changed little, there is no indication that the bed reductions were systematic attempts to equalize the regional bed distribution on a per capita basis.[17]

NHS Resource Allocations and Age-Sex Adjusted Populations Used as an Index of Need. Table V presents the simple correlations between the five resource indices and the age-sex adjusted population ratios (ASPOPR's) for the RHA's for 1974 through 1977. Positive statistically significant correlations exist for at least one RHA grouping for NMW's and GP's. Statistically significant negative correlations exist for HMS's for the London RHA's. HMC's and NPB's demonstrate no statistically significant correlations with ASPOPR, although the correlation coefficients between HMC's and ASPOPR are fairly large and negative for the London RHA's. The correlations between NPB's and ASPOPR are small, indicating that the distribution of NPB's is almost completely independent of the RHA's populations' age-sex structures.[18] In conclusion, only the allocation of GP's among non-London RHA's reflects the age-sex need measure for all years, while the distribution of NMW's among the London RHA's does so only for 1977. No other resource index has the requisite positive correlation.

Table VI presents the correlations between the manpower indices and ASPOPR, controlling for the number of hospital beds in the region.[19] Generally, the distribution of GP's is the only index to demonstrate a statistically significant positive correlation, with non-London RHA's demonstrating a generally large and increasing correlation with ASPOPR

TABLE V

NHS RESOURCE INPUT ALLOCATIONS AND THE AGE-SEX ADJUSTED POPULATION NEED MEASURE

Correlation Coefficients	Year 1974	1975	1976	1977
Nurses and Midwives per 100,000 Population				
All RHA's	0.16	0.20	0.18	0.23
Non-London RHA's	0.03	0.05	-0.12	-0.05
London RHA's	-.82*	0.11	0.36	.90**
Hospital Medical Staff per 100,000 Population				
All RHA's	-0.07	-0.14	-0.12	-0.10
Non-London RHA's	-0.32	0.25	0.33	-0.44
London RHA's	-.83*	-.87*	-.90**	-.84*
Hospital Med Consultants per 100,000 Population				
All RHA's	0.04	0.06	0.12	0.13
Non-London RHA's	-0.11	-0.07	0.03	0.02
London RHA's	-0.64	-0.72	-0.71	-0.74
General Practitioners per 100,000 Population				
All RHA's	.48**	.48**	.56**	.48**
Non-London RHA's	.72***	.74***	.88***	.83***
London RHA's	-.90*	-.86*	-.81*	-0.78
Non-Psyc Hospital Beds per 100,000 Population				
All RHA's	.38*(E)	.38(E)	.37*	.37*(E)
Non-London RHA's	.38(E)	.38(E)	0.34	.34(E)
London RHA's	.19(E)	.19(E)	0.19	.19(E)

Note: E = estimated; * = statistically significant at the 10 percent level; ** = statistically significant at the 5 percent level; *** = statistically significant at the 1 percent level.

TABLE VI

CORRELATIONS BETWEEN NHS RHA MANPOWER RESOURCE INPUTS AND THE AGE-SEX ADJUSTED POPULATION NEED MEASURE (CONTROLLING FOR RHA PSYCHIATRIC AND NON-PSYCHIATRIC BEDS)

Partial Correlation Coefficients	Year 1974	1975	1976	1977
Nurses and Midwives per 100,000 Population				
All RHA's	-0.25	-0.26	-0.29	-0.03
Non-London RHA's	-0.28	-0.17	-.51*	-0.24
London RHA's	-0.86	0.12	0.37	0.91
Hospital Medical Staff per 100,000 Population				
All RHA's	-0.20	-0.24	-0.21	-0.25
Non-London RHA's	-0.35	-0.39	-0.28	-0.41
London RHA's	-0.85	-0.90	-0.88	-0.86
Hospital Med Consultants per 100,000 Population				
All RHA's	-0.01	-0.03	0.07	0.07
Non-London RHA's	0.10	0.19	0.35	0.43
London RHA's	-0.72	-0.79	-0.77	-0.80
General Practitioners per 100,000 Population				
All RHA's	0.36	0.36	.46*	0.37
Non-London RHA's	.65***	.67**	.85***	.81***
London RHA's	-1.00**	-1.00**	-.99**	-.99**

*Statistically significant at the 10 percent level.

**Statistically significant at the 5 percent level.

***Statistically significant at the 1 percent level.

over the period. However, at the same time, the London RHA's showed an even larger statistically significant negative correlation.[20]

NHS Resource Allocations and the Age-Sex SMR Adjusted Populations Used as an Index of Need. Table VII presents the simple correlation coefficients between the resource indices and the age-sex SMR adjusted population ratios (ASSMRPR's). These correlations tell a different story. NMW's exhibit the desired statistically significant positive correlations with this need index for all years in at least one RHA grouping. For non-London RHA's. the correlations are statistically significant in all years. In 1977, the London RHA's also exhibit a large statistically significant correlation.

Hospital doctors, at consultant grade and below (HMS's), have either positive coefficients that are not statistically significant or negative coefficients which are statistically significant. The coefficients for the non-London RHA's are generally small. At the same time, the coefficients for the London RHA's are large and negative.

While GP's exhibited a strong positive correlation with the ASPOPR need measure for non-London RHA's, there is essentially no correlation between GP's and the ASSMRPR need measure. This seems contradictory in light of the correspondence in the results between the two need measures for the London RHA's. But a check of the correlations between these two need measures shows a .48 correlation for non-London RHA's and a .998 correlation for the London RHA's. This explains away the apparently contradictory results.

The correlation between NPB's and ASSMRPR's is encouraging for the non-London RHA's. While actual data were available for 1976 only, a strong statistically significant positive correlation exists.[21] This

TABLE VII

NHS RHA RESOURCE INPUT ALLOCATIONS AND THE AGE-SEX SMR ADJUSTED POPULATION NEED MEASURE

Correlation Coefficients	Year 1974	1975	1976	1977
Nurses and Midwives per 100,000 Population				
All RHA's	0.22	.52**	.37*	.50**
Non-London RHA's	.54*	.70**	.62**	.68**
London RHA's	-0.79	0.15	0.41	.93**
Hospital Medical Staff per 100,000 Population				
All RHA's	-0.18	-0.08	-0.06	-0.13
Non-London RHA's	0.17	-0.38	-0.35	0.23
London RHA's	-.86*	-.86*	-.88*	-.87*
Hospital Med Consultants per 100,000 Population				
All RHA's	-0.27	-0.21	-0.15	-0.15
Non-London RHA's	-0.17	-0.07	0.02	0.05
London RHA's	-0.66	-0.74	-0.74	-0.77
General Practitioners per 100,000 Population				
All RHA's	-0.16	-0.16	-0.05	-0.15
Non-London RHA's	0.02	0.02	0.21	0.06
London RHA's	-.91*	-.87*	-.83*	-.80*
Non-Psyc Hospital Beds per 100,000 Population				
All RHA's	.44*(E)	.44*(E)	.45*	.45*(E)
Non-London RHA's	.70**(E)	.71**(E)	.72***	.72***(E)
London RHA's	.28(E)	.28(E)	0.29	.29(E)

Note: E = estimated; * = statistically significant at the 10 percent level; ** = statistically significant at the 5 percent level; *** = statistically significant at the 1 percent level.

indicates at least some direct correlation between this need measure and the historical distribution of hospital beds. Unfortunately, there is no large positive correlation between NPB's and ASSMRPR's for the London RHA's.

Table VIII presents the partial correlation coefficients for the manpower indices and ASSMRPR, after holding constant non-psychiatric and psychiatric hospital beds in the RHA.[22] No statistically significant positive correlations exist. Only in the case of the non-London RHA's was a large positive coefficient (.94) found for NMW's.

The correlations for both HMS's and HMC's are large and negative for the London RHA's while the non-London RHA's exhibit smaller negative correlations. One possible explanation for the lack of correlation is that the allocation of NHS hospital doctors among regions is mainly a function of the staffing of hospital beds and not medical need as measured by ASSMRPR. For all RHA's combined, the positive statistically significant (5 percent level) simple correlation of .50 or better which exists between hospital doctors (HMC's and HMS's) and NPB's bears this proposition out.

The partial correlation coefficients for GP's and ASSMRPR's present the same patterns observed earlier with the ASPOPR need measure. After controlling for both types of hospital beds in the RHA's, there is no statistically significant correlation between GP's and ASSMRPR for the non-London RHA's while a negative 1.00 correlation, statistically significant at the 5 percent level, exists for the London RHA's. These results agree with the findings concerning the partial correlations of GP's with ASPOPR. Since GP's and ASPOPR are highly positively correlated for the non-London RHA's but ASPOPR and ASSMRPR are not higher

TABLE VIII

CORRELATION COEFFICIENTS BETWEEN NHS RHA MANPOWER RESOURCE INPUTS AND THE AGE-SEX SMR ADJUSTED POPULATION NEED MEASURE (CONTROLLING FOR PSYCHIATRIC AND NON-PSYCHIATRIC BEDS)

Partial Correlation Coefficients	Year			
	1974	1975	1976	1977
Nurses and Midwives per 100,000 Population				
All RHA's	-.58**	0.23	-0.30	0.15
Non-London RHA's	-0.31	0.23	-0.11	0.18
London RHA's	-0.83	0.18	0.43	0.94
Hospital Medical Staff per 100,000 Population				
All RHA's	-.54**	-.56**	-.52**	-.54**
Non-London RHA's	-0.12	-0.06	-0.10	-0.03
London RHA's	-0.88	-0.92	-0.91	-0.89
Hospital Med Consultants per 100,000 Population				
All RHA's	-.68***	-.67***	-.59**	-.58**
Non-London RHA's	-0.50	-0.47	-0.30	-0.18
London RHA's	-0.76	-0.82	-0.80	-0.84
General Practitioners per 100,000 Population				
All RHA's	-.44*	-.44*	-0.34	-.42*
Non-London RHA's	-0.01	-0.01	0.22	0.16
London RHA's	-1.00**	-1.00**	-1.00**	-1.00**

*Statistically significant at the 10 percent level.

**Statistically significant at the 5 percent level.

***Statistically significant at the 1 percent level.

positively correlated with each other for this RHA sub-group, the high correlation of GP's with ASPOPR precludes its correlation with ASSMRPR. On the other hand, ASPOPR and ASSMRPR were highly positively correlated for the London RHA's, so correlation with one need measure would ensure correlation with the other.

In conclusion, the comparisons show the NHS improved the equality of the distribution of manpower among the RHA's, based on the crude population need measure. Furthermore, for the two manpower indices compared to Cooper and Culyers' study of 1966-1967 data, the 1974-1977 distributions of HMC's and HMS's represented significant improvements over the distributions existing in 1966-1967. The only index showing no improvement was the regional distribution of non-psychiatric hospital beds. In general, then, the NHS appears to have made an effort to improve the regional equity of the distribution of these manpower resources. But, as noted, regional per capita inequities still remain.

However, evaluation of the 1974-1977 regional distribution of NHS resources with respect to the alternative need measures finds mixed results at best. The desired large, statistically significant positive correlations occur in non-London RHA's for NMW's and GP's when ASPOPR is used as an index of need. In the London RHA's, the correlation for NMW's also show the desired result for this need index. However, at the same time, HMS's and GP's exhibit large statistically significant negative correlations for the London regions. The other indices, HMC's, HMS's and NPB's, did not possess large statistically significant correlations.

When ASSMRPR is used as the measure of regional medical care need, the regional distributions of NMW's and NPB's in the non-London RHA's exhibit the desired results. For the London RHA's, only NMW's in 1977 do so. At the same time, however, large statistically significant negative correlations occur for HMC's and GP's for the London RHA's. Also, the distribution of HMC's among all RHA's combined possess a large statistically significant negative correlation with ASSMRPR for 1974 and 1975.

One must conclude from these findings that, overall, the distributions of NHS resources among RHA's show no systematic relation to either the age-sex adjusted or age-sex SMR adjusted population ratios of the RHA's. This is not a devastating criticism of past NHS policy, in and of itself, for these measures of need were developed for future allocations (post 1977). But if these indices represent valid measures of past and current regional need, they point out the inadequacy of the past methods used to allocate NHS resources among the various geographic regions.

Indices of Need and Regional Output

After evaluating the distribution of NHS resource inputs among geographic regions, the final objective of this chapter is to evaluate the distribution of the RHA's hospital medical care outputs with respect to the same indices of need. Correlations of RHA resource inputs with the measures of need provide one means of evaluating the NHS's intent of achieving geographic equity. Alternatively, correlations of RHA hospital medical care outputs with need provide a more direct measure of NHS performance at the RHA level in actually achieving equality.

The analysis uses two indices of RHA hospital medical care output. The first index, which Cooper and Culyer also used in their study, is the number of all hospital discharges and deaths per 1000 population (QHDDP). The second index is the number of cases treated per available bed in non-psychiatric hospitals (QHCB). The following narrative presents the findings of the analysis, discussing each index within the context of each measure of need.

RHA Output and the Crude Population Need Measure

Table IX presents the coefficients of variation and the high-low ratios for QHDDP and QHCB. The coefficients of variation for QHDDP decreased 14.9 percent between 1974 and 1977 for the non-London RHA's but increased 18.3 percent for the London RHA's during the same period. The combined effect was a decrease in the coefficients of variation for all RHA's of 5.6 percent.

Decreases in the coefficients of variation were expected for both RHA groups since the regional variation of NHS manpower inputs generally decreased during this period. That made the increase for the London RHA's unexpected. From 1974 through 1977, the coefficients of variation decreased for the nurse, hospital doctor, and hospital medical consultant manpower categories for both the non-London RHA's and the London RHA's. In addition, the decreases in the coefficients of variation for the London RHA's were much larger than their non-London counterparts for nurses, hospital doctors, and consultants, while GP's, as noted earlier, experienced a 21.9 percent increase. At the same time, manpower to population ratios increased in all manpower categories for both non-London and London RHA's. Why, then did the coefficients of variation for QHDDP and QHCB increase for the London RHA's? The answer appears to

TABLE IX

EQUITY OF NHS RHA OUTPUT PER 1000 POPULATION AND PER BED
BASED ON THE CRUDE POPULATION NEED MEASURE

Measures of Output Equality	Year 1974	1975	1976	1977	% Change 1974-1977
RHA Hospital Cischarges and Deaths per 1000 Population					
Coefficients of Variation					
All RHA's	7.20	7.10	7.30	6.80	-5.55
Non-London RHA's	7.40	7.00	7.20	6.30	-14.86
London RHA's	6.00	5.40	5.90	7.10	18.33
Ratio of High to Low					
All RHA's	1.31	1.28	1.30	1.23	-6.10
Non-London RHA's	1.28	1.23	1.24	1.22	-4.68
London RHA's	1.15	1.14	1.15	1.17	1.73
RHA Non-Psychiatric Cases Treated per Available Bed					
Coefficients of Variation					
All RHA's	8.80	8.40	8.20	8.10	-7.95
Non-London RHA's	9.60	9.30	9.20	8.90	-7.29
London RHA's	4.80	5.20	4.50	6.00	25.00
Ratio of High to Low					
All RHA's	1.39	1.41	1.41	1.39	0.00
Non-London RHA's	1.39	1.41	1.41	1.39	0.00
London RHA's	1.10	1.12	1.10	1.15	4.54

lie with the explanation of the large relative increase in output for one London RHA. The data indicate that South East Thames experienced a larger relative increase in both output measures than the other London RHA's. QHDDP increased 8.28 percent while QHCB increased 14.2 percent, compared with 3.1 percent and 11.2 percent for the next closest RHA. In addition, North West Thames and South West Thames exhibited very small relative increases in QHDDP during the period.

When one compares these coefficients of variation with the 9.95 percent reported for QHDDP by Cooper and Culyer for 1966-1967 for non-London RHA's, one finds an improvement of 36.7 percent by 1977.[23] The high-low ratio decreased from 1.41 to 1.22, a 13.5 percent decrease, during this same time span.[24] Both measures, then, indicate improvements in the geographic distributions of hospital output for the non-London RHA's using crude population as the measure of need.

While Cooper and Culyer did not report the number of cases treated per bed in non-psychiatric hospitals in 1966-1967, this alternative measure of RHA output (QHCB) exhibits smaller but similar changes to QHDDP over the 1974 to 1977 period. The coefficients of variation for QHCB decreased 7.3 percent for non-London RHA's during the period while increasing 25.0 percent for the London RHA's. At the same time the high-low ratios for the non-London RHA's remained essentially constant and increased 4.6 percent for the London RHA's. For all RHA's combined, the coefficients of variation decreased 8.0 percent and the high-low ratios remained approximately constant.

RHA Output and the Age-Sex Adjusted Population Need Measure

Table X provides the simple correlation coefficients between the RHA output indices and ASPOPR. As the table indicates, the correlations are generally small and not statistically significant. Only for the London RHA's in 1974 and 1977 does one find somewhat larger coefficients. In 1974, the coefficient is -0.52, while the coefficient in 1977 is 0.44. While this change represents a definite improvement, neither coefficient is statistically significant.

To examine the possibility that output correlations with this need index may be masked by variations in case-mix or resource input allocations among RHA's, partial correlations for the output indices and ASPOPR were calculated. The results are presented in Table XI on the following page. When case-mix only is held constant, all RHA's combined exhibit positive and statistically significant coefficients of .51 or greater for all years for QHDDP and ASPOPR. While the coefficients for the non-London subgroup were similar in size, they were not statistically significant. The partial correlation coefficients between QHCB and ASPOPR were small, negative, and not statistically significant.

When average lengths of stay as well as case mix are held constant, the partial correlation coefficients between QHDDP and ASPOPR increase in size in all years for non-London RHA's and all RHA's combined. In addition, the partial correlation coefficients for 1974 and 1977 for the non-London RHA's are statistically significant. Additionally controlling for average length of stay produces large positive

TABLE X

SIMPLE CORRELATION COEFFICIENTS BETWEEN RHA HOSPITAL DISCHARGES AND DEATHS PER 1000 POPULATION AND RHA NON-PSYCHIATRIC CASES TREATED PER AVAILABLE BED AND THE AGE-SEX ADJUSTED POPULATION NEED MEASURE

Measures of Output Equality	Year 1974	1975	1976	1977
RHA Hospital Discharges and Deaths per 1000 Population				
Correlation Coefficients				
All RHA's	-0.06	0.04	0.05	0.13
Non-London RHA's	-0.11	-0.09	-0.09	0.03
London RHA's	-0.52	-0.31	0.09	-0.08
RHA Non-Psychiatric Cases Treated per Available Bed				
Correlation Coefficients				
All RHA's	0.06	0.11	0.11	0.20
Non-London RHA's	0.23	0.21	0.23	0.23
London RHA's	-0.09	0.18	-0.29	0.44

TABLE XI

PARTIAL CORRELATION COEFFICIENTS BETWEEN RHA HOSPITAL DISCHARGES AND DEATHS PER 1000 POPULATION AND RHA NON-PSYCHIATRIC HOSPITAL CASES TREATED PER BED AND THE AGE-SEX ADJUSTED POPULATION NEED MEASURE

Measures of Output Equality	Year			
	1974	1975	1976	1977
Partial Correlation Coefficients Controlling for RHA Case Mix				
RHA Hospital Discharges and Deaths per 1000 Population				
All RHA's	.57*	.51*	.58**	.62**
Non-London RHA's	0.60	0.37	0.52	0.58
London RHA's	N	N	N	N
RHA Non-Psychiatric Cases Treated per Available Bed				
All RHA's	-0.29	-0.19	-0.35	-0.32
Non-London RHA's	-0.19	-0.13	-0.37	-0.18
London RHA's	N	N	N	N
Partial Correlation Coefficients Controlling for RHA Case Mix and Average Lengths of Stay				
RHA Hospital Discharges and Deaths per 1000 Population				
All RHA's	.73**	.59**	.67**	.72**
Non-London RHA's	.74*	0.46	0.65	.73*
London RHA's	N	N	N	N
RHA Non-Psychiatric Cases Treated per Available Bed				
All RHA's	0.41	0.47	0.44	0.07
Non-London RHA's	0.51	0.45	0.46	0.17
London RHA's	N	N	N	N

TABLE XI (Continued)

Measures of Output Equality	Year 1974	1975	1976	1977
Partial Correlation Coefficients Controlling for RHA Resource Inputs				
RHA Hospital Discharges and Deaths per 1000 Population				
All RHA's	0.11	0.05	0.17	.65**
Non-London RHA's	-0.04	0.00	-0.50	0.38
London RHA's	N	N	N	N
RHA Non-Psychiatric Cases Treated per Available Bed				
All RHA's	0.12	0.05	-0.01	0.46
Non-London RHA's	0.44	0.13	-0.53	0.15
London RHA's	N	N	N	N

Note: N = figure cannot be calculated due to insufficient degrees of freedom; * = statistically significant at the 10 percent level; ** = statistically significant at the 5 percent level.

coefficients for QHCB and ASPOPR but these coefficients are not statistically significant.

Controlling for RHA resource input allocations, by holding constant NMW's, HMS's, NPB's, and GP's, results in a positive statistically significant partial correlation between QHDDP and ASPOPR in 1977. A partial correlation coefficient of 0.65 results. But while a similar pattern emerges between QHCB and ASPOPR, the coefficient for 1977 is 0.46 and not statistically significant.

RHA Output and the Age-Sex SMR Adjusted
Population Need Measure

Table XII provides the simple correlation coefficients between the same RHA output indices and ASSMRPR. When this index of need is used, one finds a positive statistically significant simple correlation of 0.56 between QHDDP and ASSMRPR in 1977 for the non-London RHA's. At the same time, however, the simple correlation between QHCB and ASSMRPR is -0.54 and also statistically significant. The respective correlations for the London RHA's are not statistically significant. Only in 1977 does one find a positive coefficient of any size.

To further analyze the results to again determine if any important relationships are masked by variations in case severity or resource input allocations, partial correlation coefficients were computed. Table XIII presents the findings. First, controlling for variations in case mix only, positive coefficients are found between QHDDP and ASSMRPR in all years. At the same time, statistically significant negative coefficients between QHCB and ASSMRPR result for all RHA's combined from 1974 through 1976, while the coefficients for the non-London RHA group are negative, small, and not statistically significant.

TABLE XII

SIMPLE CORRELATION COEFFICIENTS BETWEEN RHA HOSPITAL DISCHARGES AND DEATHS PER 1000 POPULATION AND RHA NON-PSYCHIATRIC CASES TREATED PER AVAILABLE BED AND THE AGE-SEX SMR ADJUSTED POPULATION NEED MEASURE

Measures of Output Equality	Year 1974	1975	1976	1977
RHA Hospital Discharges and Deaths per 1000 Population				
Correlation Coefficients				
All RHA's	0.15	0.22	0.21	0.33
Non-London RHA's	0.35	0.42	0.41	.56**
London RHA's	-0.53	-0.31	-0.09	-0.10
RHA Non-Psychiatric Cases Treated per Available Bed				
Correlation Coefficients				
All RHA's	-.47**	-.45*	-.46**	-.39*
Non-London RHA's	-.57**	-.57**	-.56**	-.54*
London RHA's	-0.08	0.19	-0.28	0.43

*Statistically significant at the 10 percent level.

**Statistically significant at the 5 percent level.

TABLE XIII

PARTIAL CORRELATION COEFFICIENTS BETWEEN RHA HOSPITAL DISCHARGES AND DEATHS PER 1000 POPULATION AND RHA NON-PSYCHIATRIC CASES TREATED PER AVAILABLE BED AND THE AGE-SEX SMR ADJUSTED POPULATION NEED MEASURE

Measures of Output Equality	Year			
	1974	1975	1976	1977
<u>Partial Correlation Coefficients Controlling for RHA Case Mix</u>				
RHA Hospital Discharges and Deaths per 1000 Population				
All RHA's	0.16	0.04	0.01	0.30
Non-London RHA's	0.52	0.37	0.53	0.59
London RHA's	N	N	N	N
RHA Non-Psychiatric Cases Treated per Available Bed				
All RHA's	-.56*	-.48*	-.54*	-0.40
Non-London RHA's	-0.15	-0.19	-0.33	-0.03
London RHA's	N	N	N	N
<u>Partial Correlation Coefficients Controlling for RHA Case Mix and Average Lengths of Stay</u>				
RHA Hospital Discharges and Deaths per 1000 Population				
All RHA's	0.26	0.06	-0.04	0.35
Non-London RHA's	0.65	0.49	.69*	.77*
London RHA's	N	N	N	N
RHA Non-Psychiatric Cases Treated per Available Bed				
All RHA's	-0.27	-0.16	-0.22	-0.19
Non-London RHA's	0.61	0.50	.69*	0.57
London RHA's	N	N	N	N

TABLE XIII (Continued)

Measures of Output Equality	Year 1974	1975	1976	1977
<u>Partial Correlation Coefficients Controlling for RHA Resource Inputs</u>				
RHA Hospital Discharges and Deaths per 1000 Population				
All RHA's	0.33	0.29	0.26	.62**
Non-London RHA's	-0.65	0.16	-0.18	0.59
London RHA's	N	N	N	N
RHA Non-Psychiatric Cases Treated per Available Bed				
All RHA's	-0.02	0.09	0.00	0.26
Non-London RHA's	-0.42	0.26	-0.21	0.27
London RHA's	N	N	N	N

Note: N = figure cannot be calculated due to insufficient degrees of freedom; * = statistically significant at the 10 percent level; ** = statistically significant at the 5 percent level.

When average length of stay as well as case mix is controlled for, the positive coefficients for the non-London RHA's between QHDDP and ASSMRPR increase in size and become statistically significant in 1976 and 1977. Furthermore, the coefficients for QHCB and ASSMRPR for this RHA subgroup become positive and increase considerably in absolute size. In 1976, the partial correlation coefficient of 0.69 is statistically significant.

If the variation of resource input allocations among RHA's are controlled for, a positive, statistically significant partial correlation results between QHDDP and ASSMRPR for all RHA's combined in 1977. However, the coefficient for the non-London RHA's is no longer statistically significant. At the same time, the partial correlations between QHCB and ASSMRPR are not statistically significant and demonstrate no clear trend.

In conclusion, the two RHA output indices demonstrate general improvements in the equality of provision of health care for the non-London RHA's, depending upon the need measure used. When need is measured by the RHA's crude population, the decline in the coefficients of variation for both output indices show this improvement. When need is measured by the regions' age-sex adjusted population ratios, the simple correlation coefficients do not reflect the desired improvement. But when case severity is controlled for by holding case mix and average length of stay constant, the non-London RHA's have statistically significant positive correlations between QHDDP and ASPOPR in 1974 and 1977. When the age-sex SMR adjusted population ratio is used to measure RHA need and case severity is controlled for, statistically significant partial correlations with QHDDP of 0.69 and 0.77 result in 1976 and 1977. Additionally, the partial correlation of 0.69 with QHCB is also statistically significant.

Controlling for variations in resource inputs among RHA's does result in a statistically significant partial correlation with QHDDP and the two adjusted population need measures for all RHA's in 1977. But the procedure provides no statistically significant coefficients for the non-London RHA's. This result should not be entirely unexpected since the distribution of hospital doctors among the non-London RHA's showed little improvement in terms of equality.

ENDNOTES

[1]M. H. Cooper and A. J. Culyer, "An Economic Assessment of Some Aspects of the Operation of the National Health Service," in Financing Health Services, I. Jones (ed.). (London, 1970), pp. 208-209.

[2]Department of Health and Social Security, Priorities for Health and Personal Social Services in England (London, 1976), pp. 98-100. Note that neither measure was operationally implemented for the years covered (1974-1977). But since they represent measures of need, use of these measures still provide a means of evaluating NHS performance with respect to a given concept of need.

[3]Data for this comparison are located in the following publication: Office of Health Economics, OHE Compendium of Health Statistics, 3rd Edition (Luton, 1979), Ch. 1, pp. 18-19; Ch. 2, p. 14; Ch. 3, pp. 8-9, 13, 25-28; and Ch. 4, p. 8.

[4]While demand also plays an important role in the NHS system which relies on time prices, consideration of the demand side will be postponed to the next chapter.

[5]Cooper and Culyer, pp. 208-214.

[6]Ibid., p. 208.

[7]Department of Health and Social Security, 1976, pp. 97-98.

[8]Ibid., p. 100.

[9]Office of Health Economics, pp. 18-19.

[10]During the period from 1974 through 1977, the medical care price index increased 67.7 percent, based on data supplied by the Office of Health Economics, p. 4. During the same period, capital spending per capita for non-London RHA's increased 24.1 percent while increasing 3.6 percent for London RHA.s Consequently, since real expenditures per capita are declining, it would be difficult to redress regional inequities without neglecting legitimate needs elsewhere. For example, Trent, a RHA which ranked very low in terms of NPBP, received a 59.3 percent increase in CEPC during this period. On the other hand, Mersey, which ranked highest among non-London RHA's in terms of NPBP, received a 57.7 percent increase during the same period.

[11]Comparable partial correlations for the four London RHA's could not be calculated because the number of observations was less than the five variables held constant.

[12]M. H. Cooper and A. J. Culyer, "Equity in the NHS: Intensions, Performance and Problems in Evaluation," in The Economics of Medical Care, M. M. Hauser (ed.) (London, 1972), p. 56.

[13]Ruth Levitt, The Reorganized National Health Service (London, 1976), pp. 22-51.

[14]Cooper and Culyer, 1972, p. 55.

[15]Ibid.

[16]Ibid.

[17]Department of Health and Social Security, 1977, pp. 84-86. To make the 1974-1977 years comparable to Cooper and Culyer's measure, total beds per 1000 population, non-psychiatric and psychiatric beds per 1000 population were combined and the coefficients of variation computed.

[18]One would not necessarily expect any correlation since the distribution of hospital beds was "inherited" by the NHS and has remained largely unchanged. The only way the current age-sex distribution would match the historical distribution of hospital beds would be for the distributions of both to not have changed significantly over time and for the original distribution of beds to reflect the need for hospital beds originating from this age-sex structure.

[19]The partial correlation coefficient can be used to locate relationships where none appear to exist. This is appropriate if there exists a possibility that some variable or variables act to hide or suppress the relationship. For example, since the manpower indices are highly correlated with hospital beds, but hospital beds are not highly correlated with ASPOPR, then any attempt to reallocate manpower resources by ASPOPR may not show up. For more information on partial correlation coefficients, see Robert Pindyck and Daniel Rubinfeld, Econometric Models and Economic Forecasts, 2nd Edition (New York, 1981), pp. 91-94.

[20]The London RHA partial correlations are computed holding only non-psychiatric hospital beds constant due to the limited number of observations.

[21]The results for the other years represent bed to population ratios estimated using the assumption that the RHA bed shares in 1976 were equal to the bed shares for the other years. Total beds in each of these years were allocated among the RHA's on this basis.

[22]The closest partial correlation which could be calculated for the London RHA's holds only non-psychiatric hospital beds constant. With only four London RHA's, the partial correlation holding both types of beds constant could not be calculated.

[23]Cooper and Culyer, 1972, p. 55.

[24]Ibid.

CHAPTER V

REGIONAL CHARACTERISTICS, PHYSICIAN BEHAVIOR, AND HOSPITAL EFFICIENCY

This chapter presents the second and third areas of empirical investigation. The second area of investigation examines the relative efficiency of the RHA's hospital sectors and attempts to find explanations for these efficiency differences. The third area of investigation examines how hospital doctor behavior affects relative regional efficiency through the impact on average length of stay.

To examine relative regional efficiency, the analysis uses linear programming techniques to estimate an RHA hospital sector production function. The estimated production function is then used to create indices of relative efficiency. These indices are then to be used to determine causes for the efficiency differences among the RHA's.

The chapter organizes the presentation into two major sections. The first section discusses the linear programming estimation procedure. After doing so, a discussion of the data follows. The estimated production functions are then presented. Last, the efficiency indices created from these production function estimates are discussed within the context of the RHA medical care need measures.

The second section uses the efficiency indices to determine possible causes for relative regional efficiency differences. Medical care demand variables are examined in an excess demand framework. The study hypothesizes that variations in excess demand will explain variations in RHA hospital sector efficiency. As part of this analysis, this section seeks to determine if utility maximizing behavior of hospital doctors reduces hospital sector response to excess demand pressure.

RHA Hospital Sector Production Functions and Relative Efficiency

The concept of technical efficiency in economics refers to producing the maximum possible output from a given set of inputs.[1] Following Wilson and Jadlow and others, this study employs linear programming to estimate the parameters of an RHA hospital sector production function.[2] As Wilson and Hadlow note, this approach is desirable because it constrains all observations to lie either on or below the estimated production surface.[3]

After choosing the estimation technique, one must choose the type of production surface to estimate. Basically, there are two choices and Forsund, Lovell, and Schmidt attribute both to Farrell.[4] The first surface is called a deterministic non-parametric frontier and the second a deterministic parametric frontier. Forsund et al. note the former has the advantage of not imposing a functional form, such as that of the Cobb-Douglas function, on the data. But it has the disadvantage of assuming constant returns to scale. The latter has the advantages of expressing the frontier in a simple mathematical form, e.g., the Cobb-Douglas form, and the ability to handle non-constant returns to scale. But it has the disadvantages of imposing a possibly incorrect form on the production surface and limiting the number of observations that can lie on the production surface to the number of parameters estimated. Additionally, both approaches suffer from the disadvantages of being very sensitive to extreme observations and the fact that the coefficient estimates have no statistical properties for which they can be evaluated.[5]

Due to the theoretical and computational problems of the non-parametric approach, Forsund et al. note that this approach has "won few adherents."[6] Consequently, this study follows the path of others by choosing the parametric approach and using a Cobb-Douglas production function.[7] In addition, like Timmer, once the deterministic frontier is estimated, the efficient observations are discarded, one at a time in this case, to determine if the estimated coefficients are affected by outliers.[8]

The Data and Estimated Production Functions

Hospital bed data were available on a regional basis for 1976 but were needed for 1974, 1974, 1975, and 1977. To provide bed data for the missing years, it was necessary to allocate yearly total beds for England to the regions, based on the assumption that RHA bed shares remained constant. This assumption is supported by the finding in Chapter IV that the coefficients of variation for hospital beds had not changed much from Cooper and Culyer's study.[9]

The Office of Health Economics provided RHA data on nurses and midwives (NMW), hospital medical staff (doctors below consultant grade) (HMS), and hospital medical consultants (HMC).[10] But each of these manpower categories combines NHS activities in addition to non-psychiatric hospital care. Consequently, it was necessary to separate non-psychiatric hospital manpower from this total. To do so, the study uses average staff ratios per bed.[11]

Two measures of each RHA's non-psychiatric hospital sector output are calculated. The first measure is total non-psychiatric cases treated. The second measure is non-psychiatric cases treated, weighted by an index of the relative cost of each broad case type and a cost index converting hospital outpatient visits to weighted case equivalents. Each measure was used to estimate an alternative form of the Cobb-Douglas production function.[12]

The Office of Health Economics provided data on non-psychiatric cases treated per available bed.[13] To obtain total non-psychiatric cases treated, cases treated per bed were multiplied by the RHA non-psychiatric bed stock.

The second output measure is based on the first. To capture the multiproduct nature of the hospital sector's outputs, the cases are weighted by an index based on national average cost per case. Data on four broad case types were available for each year.[14]

The form of the Cobb-Douglas production function used to estimate unweighted non-psychiatric cases has the following functional form:

$$\begin{aligned} Ln(Q) \leq\; & Ln(A) + B_1\, Ln(NMW) + B_2\, Ln(HMS + HMC) + B_3\, Ln(BEDS) + \\ & t\,(YEAR) + L\,(LONDON) + m\,(CASEMIX - MEDICAL) + \\ & s\,(CASEMIX - SURGERY). \end{aligned} \qquad (1)$$

In addition to variables discussed in the preceding section, a time trend variable (YEAR), two casemix variables (MEDICAL and SURGERY), and a dummy variable for London (LONDON) are added. The exponential time trend variable provides a simple control for technical progress and implies technology is neutral in its effect on the productivity of capital and labor.[16] The casemix variables are also entered in exponential form. This implies that they have proportional effects on the number of cases treated.[17] They also provide a control to reflect the multiproduct nature of the hospital sector. Finally, a dummy variable for the London RHA's is included to reflect the special characteristics of the London hospitals. These characteristics include complexity of the cases treated, the advanced technology available that may not be available in other areas, and medical personnel with the latest skills.

The Cobb-Douglas production function used to estimate weighted non-psychiatric cases has the following form:

$$\begin{aligned} \mathrm{Ln}(Q_w) \leq \mathrm{Ln}(A) &+ B_1\ \mathrm{Ln}(NMW) + B_2\ \mathrm{Ln}(HMS + HMC) + B_3\ \mathrm{Ln}(BEDS) + \\ &t\ (YEAR) + L\ (LONDON) \end{aligned} \qquad (2)$$

This model differs from the first by adjusting the hospital output to reflect variations in casemix. The other variables serve the same purposes as they do in the first model.

Table XIV presents the estimated production function frontiers for the model in (1) above. The coefficients for NMW's are negative. They indicate that an increase of one percent in a region's nurse and midwife staff will lead to approximately a 0.23 percent decrease in the number of non-psychiatric cases treated. The coefficients for doctors and hospital beds have the expected positive signs. They infer that a one percent increase in doctors and hospital beds will lead to increases in the number of cases treated of approximately 0.40 percent and 0.83 percent respectively.[18]

The sum of the elasticities for the labor and capital inputs equal 0.9823 or larger. This indicates approximately constant returns to scale. The positive coefficients for the time trend indicate increasing productivity over time. The negative coefficients for the London dummy variable agree with the proposition that the London RHA's treat more complex and severe cases than their counterparts elsewhere, decreasing the rates of output for these RHA's, given the level of inputs. Lastly, the negative coefficients for the two casemix variables support the data which indicate that the national average lenghts of stay for medical cases exceed the national average lengths of stay for all acute care cases combined while surgery cases have average lengths of stay that are only slightly longer than the average for all cases combined.[19]

TABLE XIV

PRODUCTION FRONTIER ESTIMATES FOR DEPENDENT VARIABLE NON-PSYCHIATRIC CASES

Parameter	100%	98%	96%	95%
	1974-1977			
Ln(CONSTANT	1.8393	2.1009	2.0858	2.0893
Ln(NMW)	-0.2763	-0.2334	-0.2345	-0.2309
Ln(HMS + HMC)	0.4093	0.3909	0.3906	0.3849
Ln(BEDS	0.8529	0.8249	0.8264	0.8295
YEAR	0.0436	0.0395	0.0397	0.0393
LONDON	-0.0462	-0.0433	-0.0438	-0.0419
MEDICAL CASES	-3.4844	-3.4727	-3.4800	-3.4590
SURGICAL CASES	-0.1249	-0.0314	-0.0295	-0.0360
Number of Observations	56	55	54	53

Table XV presents the estimated production function frontiers for the model in (2) above. The coefficients for nurses and midwives (NMW) again are negative and quite large.

TABLE XV

PRODUCTION FRONTIER ESTIMATES FOR DEPENDENT VARIABLE NON-PSYCHIATRIC CASES WEIGHTED BY CASEMIX AND OUTPATIENT VISITS CASE EQUIVALENTS

Parameter	100%	98%	96%	95%
	1974-1977			
Ln(CONSTANT)	1.3195	0.8979	0.1944	0.1755
Ln(NMW)	-0.3404	-0.3360	-0.4076	-0.4096
Ln(HMC + HMS)	0.1727	0.0843	0.0762	0.0761
Ln(BEDS)	1.0576	1.1147	1.2397	1.2431
YEAR	0.0457	0.0520	0.0552	0.0553
LONDON	-0.0700	-0.0555	-0.0552	-0.0552
Number of Observations	56	55	54	53

The elasticities for doctors are now much smaller while the elasticities for beds are considerably larger. Furthermore, the large negative elasticity for nurses implies that a one percent increase in NMW's will result in a 0.4076 percent decrease in weighted output. In addition, the sum of the elasticities for the labor inputs equals -0.3314 while the sum of the elasticities for both labor and capital inputs equals 0.9083 in the 96 percent frontier and less in the 98 percent and 100 percent frontiers. The coefficients for YEAR and LONDON are the same sign but somewhat larger.[20]

The Indices of RHA Hospital Sector Efficiency

From the two production function frontiers estimated using linear programming, indices of RHA hospital sector efficiency were constructed. Since the estimated coefficients for a given set of observations represent the production surface, actual RHA hospital sector output must be less than or equal to estimated output. The objectives of this section of the chapter are to describe the procedure of picking the set of coefficients for a given production frontier; to describe the construction of the resulting efficiency indices; and to examine the simple correlations between the efficiency indices and two measures of RHA medical care need.

Forsund et al. note that the estimated frontier obtained by linear programming techniques is very sensitive to extreme observations.[21] In attempting to avoid this problem, Timmer suggests deleting efficient observations, a small fraction at a time, to determine the level at which the coefficients stabilize.[22] When the coefficients stabilize, the resulting frontier, which represents a certain percentage of the observations, is called a "probablistic frontier."[23] Forsund et al. note that the usefulness of this approach depends on the rate of change in the estimates decreasing rapidly as additional observations are deleted.

Table XIV presented the results of this procedure for the frontier using unweighted non-psychiatric cases. While the coefficients never all attain a constant set of values, the values roughly stabilize at the 98 percent frontier. No added degree of stability results from deleting additional observations. In fact, on run 11, after deleting 25 efficient observations from the previous 10 runs, the coefficients attained similar values to those of the 98 percent frontier. Consequently, the coefficients associated with the 98 percent frontier were chosen for this particular efficiency index (DQ).

Table XV presented the results of this procedure for the frontier using weighted non-psychiatric cases. Again, the coefficients never attain a constant set of values. However, the coefficients attain relative stability at the 96 percent frontier so this set of coefficients was used for the efficiency index (IQW).

Since the actual levels of RHA hospital sector output must be less than or equal to the estimated levels of output, the efficiency index takes the following form:

$$IQ_{ij} = (Q_{e_{ij}} - Q_{ij}) / Q_{e_{ij}} \qquad (3)$$

Where $Q_{e_{ij}}$ = the estimated output of type i for RHA hospital sector j,
Q_{ij} = the actual output of type i for RHA hospital sector j, and
IQ_{ij} = the associated index.[25]

For observations on the frontier, the index will be zero so the larger the index the larger the relative inefficiency of the particular hospital sector.

The indices of relative efficiency ranged in value from zero to 2.22 percent for the weighted output index and zero to 1.44 percent for the unweighted output index. The largest average index value was 0.560 percent for IQW while the average index value was 0.367 percent for IQ. As these figures indicate, the estimated relative efficiency differences are small. But the coefficients of variation equal 105 percent for IQW and 101 percent for IQ. This means that, although average relative inefficiency is small, considerable variation among the RHA hospital sectors exists.

Table XVI presents the simple correlations between the two efficiency indices and the age-sex and age-sex SMR adjusted population ratios. None of the correlations between the measures of need and the unadjusted output index (IQ) are statistically significant. However, the weighted output index correlations with these need indices are statistically significant at the five percent level. The negative coefficient for the age-sex adjusted population ratio indicates a small direct association between this need measure and increased RHA hospital sector efficiency. This might possibly reflect increased demand pressure on hospitals that result in increased rates of output. The positive coefficient for the age-sex SMR adjusted population ratio implies a small direct relation between this more complex need measure and hospital sector inefficiency. A possible explanation of this result may be that the second need measure better captures the severity of need and indicates that regions with greater need also have more cases requiring longer lengths of stay. The longer lengths of stay would show up in lower rates of output and might not be captured by the casemix controls built into the production functions.

Relative RHA Hospital Sector Efficiency: Two Hypotheses

This section of the chapter uses the efficiency indices developed above to test two hypotheses concerning relative regional efficiency. The first hypothesis states that excess demand will put pressure on a region's hospital sector to produce more output from a given amount of inputs and this result will show up as follows. Treating the supply of medical care as fixed in a given time period, RHA characteristics which increase the demand for medical care will be negatively related to the efficiency indices. The second hypothesis concerns the fixed nature of supply in a given time period. It states that hospital doctors will regulate their workload to some professionally designated level since their compensation is not a function of output. Consequently, as the number of hospital beds per hospital doctor increases, the average length of stay will increase to regulate the doctors' workload. The following sections discuss the hypotheses in more detail and present the empirical findings.

TABLE XVI

SIMPLE CORRELATIONS BETWEEN THE RHA HOSPITAL SECTOR EFFICIENCY INDICES AND MEASURES OF RHA MEDICAL CARE NEED

Efficiency Index	IQW	IQ
Measure of Need		
Age-Sex Adjusted Population Ratio	-0.3103 (.011)	0.1088 (.215)
Age-Sex SMR Adjusted Population Ratio	0.2608 (.028)	-0.0045 (.487)

Note: Significance levels in parentheses.

Excess Demand and Increased Relative Hospital Sector Efficiency

Waiting lists have been used as a traditional means to observe excess demand. As discussed in earlier chapters, the concept of medical care need is a technological and relative concept so even though the output of the NHS increased over time, waiting lists have not decreased in size. In fact, they have increased from 460,000 persons in 1949 to 607,000 in 1976.[26] Consequently, waiting lists do not reflect all excess demand that exists at a particular point in time.

To explain the behavior of waiting lists, Lindsay developed a model that attempts to explain the increase in waiting lists in response to increases in system capacity.[27] According to Lindsay, the decision to join the waiting list is a demand decision originating with the individual. The decision to join the list depends on the expected value of the hospital treatment to be obtained and the cost of joining the list. The expected value of hospital treatment decreases over time at a rate which depends on the type of disease. For diseases which can be treated at home or by outpatient care, the value of treatment decreases quickly over time so these individuals will not join a waiting list that is too long. For diseases that must be treated in the hospital and do not deteriorate over time, the value of treatment decreases slowly so they will join a longer queue than those in the first category.[28]

Lindsay regresses available beds per population and the change in available beds per population on mean waiting time by disease category and finds that diseases whose value of treatment decreases quickly over time to be highly responsive to changes in bed capacity.[29] He also finds that diseases whose values of hospital treatment decrease slowly over time are responsive (in terms of mean waiting times) to higher bed capacities. As he hypothesizes, the NHS appears to be aware of resulting needs for this latter category.

On the other hand, Lindsay reasons that the excess demand reflected by the waiting lists may be only a small part of total excess demand.[30] As noted earlier, besides the value of the hospital treatment to be obtained, the other factor in the individual's decision to present a demand to the NHS for treatment, or at least placement on the waiting list, is the cost of registering this demand with the NHS. The higher this cost, the less aware the system will be of the existing demands. The lower this cost, the more aware the system will be and the more likely that it will respond to the demands by attempting to increase the output from a given endowment of resources.

To test the hypothesis that the RHA hospital sectors respond to visible excess demands by increasing output from a given level of inputs, this study regresses variables related to RHA demand characteristics on the efficiency indices. Factors that decrease the cost of registering a demand with the system or factors which lower the cost of care obtained from the system will be directly related to relative hospital sector efficiency (negatively related to the efficiency indices). Those factors which raise the value of care will increase demand and efficiency.

Demand Variables Affecting the Cost of Registering Demands and Receiving Medical Care. Since the NHS relies on time prices as a rationing device, factors increasing the time involved in registering demand or receiving medical care will decrease the number of demands presented to the system and vice versa. Additionally, factors which raise the cost of the time spent in registering demand or receiving care will have the same effect. Variables included in the model which are associated with time inputs are population density (POPDEN), population per GP (PGP), and average length of stay (LOS). The variable directly associated with the cost of time is wages not covered by employer guarantees to pay the individual when sick (WFP).

As discussed in Chapter III, the time required to obtain a unit of medical care is composed of travel, waiting, and consumption times. While the cost of registering a demand with the system may involve no current consumption of hospital services, the prospective patient must spend time with both a GP and a consultant to do so. This involves some of each time component. Additionally, the longer the lengths of stay in the region, the longer the waiting times required to obtain hospital treatment, and possibly the lower the value of that treatment, depending of the nature of the condition.[31]

The model the study employs to test the hypothesis uses POPDEN as a proxy for travel distance to the doctor. Acton found in a study of New York City's free health clinics that distance to the clinic acted as a price in determining the quantities of care demanded.[32] Since this information is not available for the NHS, population density will be used as a proxy for travel distance to the doctor. As population density increases, the distance to the doctor should decrease. One problem, however, in using POPDEN for this purpose is that higher population densities may also be associated with longer waiting times. While shorter travel times should increase quantities demanded, longer waiting times will have the opposite effect.

To anticipate this possibility, two additional variables are included in the model. PGP enters as a proxy for waiting time and the percentage of the RHA's wage and salary income from the agriculture, forestry, and fishing industries (PAFFWS) enters as a proxy for travel distance to the doctor, i.e., travel time. The reasoning behind including PAFFWS is that these industries are associated with rural areas and, hence, larger distances to the doctor.

The unit cost of the time required to register a demand with the system is the RHA wage and salary income per capita that is not covered by employer guarantees to pay the employees when they are sick. The General Household Survey provided the percentage of employees that responded they "got paid when sick," by industry.[33] These percentages were multiplied by wage and salary income in each industry to obtain wages that would not be lost due to sickness. This amount was subtracted from total wages to obtain non-covered wages for each industry. These amounts were added together and then divided by RHA population to obtain non-covered wages per capits (WFP). This figure provides a crude measure of the average cost of time to prospective patients in a given region.

However, since the above foregone earnings variable is fairly crude, the proportions of wage and salary incomes associated with other industry groups are also included in the model. The percentage of non-covered wages varies by industry group. Industries accounting for larger shares of WFP will be associated with higher time prices. The variables are: construction industry wages share (PCWS); mining, quarrying, gas, electricity, and water industry wages share (MQGEWS); and service industry wages share (SERVWS). They are listed in decreasing order of their contributions to the RHA's non-covered wages.[34]

Two demand variables reflect the value of treatment to potential patients. The first variable, already discussed, is average length of stay (LOS). The second variable is a proxy for quality. It is the number of teaching districts in the region. Since teaching hospitals are associated with higher quality care, the demand for medical care should increase. However, to the extent that the case weighting in the production function estimates does not adequately reflect the variation in the severity of cases among regions, the likelihood of this proxy performing well decreases.

Three variables in the model attempt to control for the availability of substitute forms of medical care. The first two, the percentage of fulltime medical consultants (PFTHMC) and the percentage of fulltime hospital doctors below consultant grade (PFTHMS), are proxies for the availability of private practice sources of medical care. Acton found that the free clinics in New York City, with their long waits, were economically inferior goods while private practice physicians were normal goods.[35]

The third proxy for substitute forms of medical care is the ratio of general practitioners to hospital doctors exists in a region, general practitioners may not refer patients to consultants, knowing that the chance they will be placed on a waiting list is small. Instead, they may choose to provide treatment for which hospital care is not absolutely required.

The two Working Party indices of regional medical care need enter the model to determine if these indices explain any variation in regional efficiency. The first index, the age-sex adjusted population ratio (ASPOPR), is derived from hospital sector utilization rates for different age and sex groups. The second index, the age-sex standard mortality ratio adjusted population ratio (ASSMRPR), weights the age-sex groups' utilization rates by the regions' standard mortality ratios for different diseases. The latter should more correctly measure regional medical care need than the first index. In addition, it may also account for variations in case severity not previously accommodated. While the expected sign for ASPOPR is unknown, one might argue that the second variable should be directly correlated with the efficiency indices if it correctly takes into account variations in case severity not previously controlled for. On the other hand, if the index represents conditions that demanders are generally aware of, then the correlation would be negative, since persons in need are motivated to demand medical attention.

The Empirical Findings: RHA Hospital Sector Efficiency and Excess Demand. Table XVII presents the regressions of the demand variables on the two efficiency indices. Four time price variables, WFP, PAFFWS, PGP, and LOS, have the correct statistically significant positive signs in both regressions. The results imply that as the time price increases and the value of the care decreases, the demands presented to the system decrease and this shows up as relative inefficiency.

The conflicting influences of POPDEN on travel time and waiting time appear as small, non-statistically significant positive coefficients in both regressions. But PAFFWS does exhibit the expected positive sign in three of the four regressions. This lends support to its use as a proxy for travel distance but this cannot be stated conclusively since the percentage of wages paid when sick in this industry equals the percentage of wages paid when sick in the construction industry which also exhibits a large positive statistically significant coefficient. Both industries rank lower than the others in terms of the percentage of wages paid when sick, implying that time prices are higher for their

TABLE XVII

RHA HOSPITAL SECTOR EFFICIENCY AND MEDICAL CARE DEMAND

Efficiency Index	IQW	IQ
Independent Variables		
Constant Term	-15.0053***	-10.5660***
	(-4.161)	(-3.154)
Non-Covered Wages (WFP)	0.0082***	0.0041**
	(4.704)	(2.492)
Population Density (POPDEN)	0.0002	0.0004
	(0.458)	(0.903)
% AG-FOR-FISH Wages (PAFFWS)	0.1882***	0.2541***
	(3.672)	(5.056)
% Construction Wages (PCWS)	0.4443***	0.0942
	(6.071)	(1.3392)
% MIN-QUAR-UTIL Wages (PMQGEWS)	-0.0487	0.0007
	(-0.775)	(0.011)
% Service Sector Wages (PSERVWS)	0.0945***	0.0323**
	(5.981)	(2.139)
Population Per GP (PGP)	0.0012*	0.0014**
	(1.990)	(2.312)
Length of Stay (LOS)	0.4485***	0.2425***
	(6.889)	(3.966)
% Full-Time HMC's (PFTHMC)	0.0452***	0.0318**
	(3.173)	(2.301)
% Full-Time HMS's (PFTHMS)	-0.0230	-0.0211
	(-1.070)	(-1.024)
GP's per Hosp. Doctor (GPHD)	0.0321***	0.0045
	(3.319)	(0.482)
Age-Sex Pop. Ratio (ASPOPR)	-0.1192***	0.0235
	(-4.994)	(1.027)
Age-Sex SMR Pop. Ratio (ASSMRPR)	0.0555***	-0.0211
	(3.083)	(-1.203)
# Teaching Districts (TDIST)	0.0486	0.0637
	(0.841)	(1.125)
Adjusted R-Square	.883	.715
Number of Observations	54	55

Note: T-values in parentheses; * = statistically significant at the 10 percent level; ** = statistically significant at the 5 percent level; *** = statistically significant at the 1 percent level.

workers. In any case, the performance of each of these variables is consistent which work to decrease the quantities of medical care demanded.

The performance of the other two industry variables is not as satisfactory. The PMQGEWS variable has the hypothesized sign in the weighted output regression but the coefficient is not statistically significant. A different problem occurs with the PSERVWS variable. It is statistically significant in both regressions but has an unexpected sign. Since most service industry employees reported they got paid when sick, their foregone earnings rate should be low, resulting in a lower time price. Multicollinarity may be at fault here, because both variables are highly correlated with each other (-0.662), as well as with several other variables in the model. For example, the correlation between PSERVWS and PFTHMC is -0.745. This implies that service industry employment is highly correlated with the availability of substitute private practice alternatives, whose influence may offset the effect of the higher percentage of wages not lost due to sickness.

For the three variables entered as proxies for substitute forms of medical care, GPHD has the correct sign and a statistically significant coefficient in the regression on the weighted output efficiency index. On the other hand, PFTHMS has the correct sign but the coefficients are not statistically significant. The third proxy for substitute medical care, PFTHMC, has a positive sign and is statistically significant in both regressions. The coefficients indicate that as the percentage of hospital medical consultants that work for the NHS full-time decreases, efficiency increases. One plausible interpretation of this result is that the existence of private practice, by providing a substitute to NHS care, motivates the NHS to try harder.

The coefficients for the two need variables were statistically significant in the weighted output efficiency index regression but not in the other. ASPOPR appears to act as a demand variable and not as a case severity variable. This should not be surprising since it is based on utilization rates only. Also, ASSMRPR seems to be measuring case severity in the regression on IQW.

TDIST may be a poor proxy for the influence of quality on the demand for medical care. On the other hand, the increased complexity of cases treated by teaching hospitals would tend to increase LOS. This would tend to decrease the demand for medical care. These offsetting influences, like the case of the POPDEN variable, tend to explain the poor performance of this variable.

Lastly, the adjusted R-squares indicate that these demand variables explain a significant portion for the interregional variations in hospital sector efficiency. The generally consistent performance of the demand variables is encouraging. Their performance lends strong support to the excess demand hypothesis.

Hospital Doctors, Workload, and Variation of Average Lengths of Stay

The significance of average length of stay (LOS) in explaining interregional variations in relative hospital sector efficiency leads to the second hypothesis to be tested. This hypothesis states that NHS hospital doctors will regulate their workload to some professionally acceptable level since their compensation is not directly a function of output.[36] To maximize their utility, they will be motivated to meet only the standards set by their profession and, to some extent, NHS administrators. While clinical freedom allows each doctor discretion in using NHS resources to treat his patients, it will bias behavior toward meeting professional standards and away from any NHS objectives that may be in conflict with those standards. Consequently, given the uncertainty often associated with the treatment of illness, clinical freedom, coupled with utility maximization, will bias doctors toward longer lengths of stays as opposed to higher output rates.[37]

To test this hypothesis, the following general model will be used:

$$LOS_j = F[(BEDS/HMC)_j, (BEDS/NMW)_j, (HMS/HMC)_j, PFTHMC_j, (GP/HMC)_j, IQ_{ij}, ASSMRPR_j, YEAR] \qquad (4)$$

where the variables are defined as before. The model concentrates on hospital medical consultants because junior grade hospital doctors are responsible to a consultant and presumably, it is the consultants that make the final decisions concerning the disposal of cases. No management hierarchy exists among consultants so only informal controls, such as consultants distinction awards, exist among consultants.[38] Consequently, it is at the consultant grade that true clinical freedom exists.

The hypothesis to be tested maintains that as the number of beds per hospital doctor (consultant) increases, the average length of stay will increase. The largest part of the workload for a consultant probably occurs at the beginning of the hospital stay. An increase in the number of beds available per consultant, holding length of stay constant, will increase the workload of the consultant and decrease the amount of leisure time consumed. If the consultant receives no direct compensation for this extra workload, then he will not be as motivated to accept the higher workload.

The consultants' ability to shift workload to other NHS employees depends on the availability of these persons. If the nursing staff bears the responsibility for most of the work following the initial treatment phase, the ability of the HMC's to shift workload to the nursing staff by increasing average lengths of stay will depend on the ratio of hospital beds to nurses (BEDS/NMW). The lower this ratio, the greater capacity for the nursing staff to handle a given workload. In fact, the workload for the nursing staff may also decrease as LOS increases. If this is the case, the nursing staff may also favor longer average lengths of stay. Consequently, the expected sign for BEDS/NMW is negative.

Another variable that must be controlled for in testing the hypothesis is the ratio of junior hospital doctors to hospital medical consultants (HMS/HMC). As this ratio increases, the more supervisory workload a consultant has. One would expect this type of workload, which must be performed, to limit the capacity to take on new cases. Therefore, this proxy for supervisory workload should be positively associated with LOS.

The percentage of full-time consultants in the region (PFTHMC) is entered in the model to control for private practice demands placed on the consultant. If private practice work is fee for service, the consultant may give it priority over salaried NHS work. Using NHS work to fill in for a lack of private work, the consultants may increase lengths of stays for NHS patients when private workload increases. If this is the case, then the expected sign for PFTHMC is negative.

The consultants must also spend time determining who should be put on waiting lists for future hospital treatment. Since the number of general practitioners in the region are directly associated with the number of referals to consultants, the ratio of general practitioners to hospital medical consultants (GP/HMC) enters the model to control for the referal workload of consultants. The consultant may respond to this workload in two ways. First, since this workload takes time away from treating hospital patients, consultants may respond by increasining average lengths of stay. Alternatively, if referals place pressure on consultants to increase output by making them aware of large amounts of unmet needs, they may decrease lengths of stay in order to increase output. This latter motivation would seem less likely if consultants are workload minimizers, as hypothesized. If the first motivation dominates, then the expected sign of GP/HMC is positive. If the second motivation dominates, the sign will be negative.

In addition to the above variables, LOS will be affected by the general efficiency of the RHA hospital sectors. As previously demonstrated, LOS is directly associated with relative hospital sector inefficiency. To control for this variable, the index of relative efficiency (IQ_{ij}) will be included in the model. As before, LOS should increase directly with the size of the efficiency index.

The age-sex SMR adjusted population ratio measure of regional medical care need (ASSMRPR) is included in the model to determine if consultants recognize medical need as indicated by this index. As discussed earlier, if the need index measures the severity of cases treated that has not been accounted for in the efficiency indices, then it should be directly associated with average length of stay.

Lastly, a time trend variable (YEAR) is included to control for improvements in medical technology and treatment methods. Since improved techniques will affect doctors attitudes concerning the appropriate treatment of cases, this knowledge will act to restrain consultants from increasing average lengths of stay. Combined with the increased productivity offered by advances in medical care technology, the average length of stay should decrease as time advances, implying a negative regression coefficient for time.

The Empirical Results: Consultant Behavior and Average Length of Stay. Table XVIII presents the results for the LOS regressions using the four efficiency indices discussed earlier and the ASPOPR need index. As the adjusted R-squares indicate, the model explains a large proportion of the variation of average lengths of stay among the regions. In addition, most of the variables are statistically significant and have the expected signs.

The coefficients for the BEDS/HMC variable is statistically significant in both regressions and have the expected positive sign. These results support the hypothesis that HMC's do use LOS to regulate workload. The small size of the coefficient may be deceiving. This implies an elasticity coefficient of 0.23 to 0.50. This indicates that a one percent increase in the BEDS/HMC ratio will increase LOS from 0.23 to 0.50 percent.

The BEDS/NMW variable has the expected negative sign in both regressions. The negative sign indicates that as the ratio of beds to nurses falls, average lengths of stay increase. While these coefficients are not statistically significant, their negative signs are consistent with the hypothesis that doctors shift workload to nurses. The coefficients indicate that a one percent decrease in the

TABLE XVIII

HOSPITAL DOCTORS, LENGTHS OF STAY, AND THE ASSMRPR NEED MEASURE
DEPENDENT VARIABLE: AVERAGE LENGTH OF STAY (LOS)

Efficiency Index	IQW	IQ
Constant	38.1862*** (5.297)	33.4322*** (4.732)
Beds/Consultants (BEDS/HMC)	0.0810* (1.860)	0.1749*** (5.420)
Beds/Nurses (BEDS/NMW)	-1.1282 (-1.370)	-0.7920 (-0.951)
Doctors/Consultants (HMS/HMC)	1.5255*** (3.335)	1.8211*** (4.083)
% Full-Time Consultants (PFTHMC)	-0.0458*** (-4.233)	-0.0530*** (-4.910)
GP's/Consultants (GP/HMC)	-1.5508*** (-2.885)	-2.8724*** (-8.694)
Efficiency Index (IQ_{ij})	0.6190*** (3.550)	0.7147*** (4.425)
Age-Sex SMR Pop. Ratio (ASSMRPR)	0.0088 (0.628)	0.0140 (1.087)
Year	-0.3383*** (-3.826)	-0.2839*** (-3.276)
Adjusted R-square	.808	.820
Number of Observations	54	55

Note: T-values in parentheses; * = statistically significant at the 10 percent level; ** = statistically significant at the 5 percent level; *** = statistically significant at the 1 percent level.

BEDS/NMW ratio will increase LOS by 0.120 to 0.084 percent, a relatively small response.

The HMS/HMC variable has the expected positive sign and is statistically significant at the one percent level in both regressions. The elasticities range from 0.28 to 0.34 percent. This indicates that a one percent increase in the supervisory workload of HMC's will lead to a 0.28 to 0.34 percent increase in average lengths of stay.

The proxy for referral workload (GP/HMC) has a relatively larger potential impact on LOS than the variables discussed above. In elasticity terms, a one percent increase in the ratio, ceteris paribus, will decrease LOS from 0.37 to 0.68 percent. Additionally, the negative sign indicates that HMC's respond to referral workload from GP's by decreasing LOS, presumably to enable the system to operate at a higher rate of output.

The efficiency indices (IQW and IQ), PFTHMC, and ASSMRPR have very small relative impacts on LOS. The elasticities indicate that a one percent increase in relative efficiency, ceteris paribus, will lead to only a 0.0003 to 0.0004 percent decrease in LOS. For a one percent decrease on the number of full-time HMC's, LOS will increase 0.0039 to 0.0045 percent. And for a one percent increase in ASSMRPR, LOS increases 0.0009 to 0.0014 percent. Although two of these variables are statistically significant at the one percent level in both regressions, their quantitative impact is negligible.

The time trend (YEAR) variable's coefficient is statistically significant at better than the one percent level in all regressions. The results indicate that lengths of stay have been falling by 0.28 to 0.34 days per year. This implies that improvements in medical technology and treatments have increased the productivity of hospital stays and resulted in lower average lengths of stay.

An Integration of the Empirical Findings

The estimated production function frontiers provide relative efficiency indices which indicate that considerable variation in relative efficiency of RHA hospital sectors exists. As hypothesized, interregional variation in these efficiency indices can be explained in large part by variation in demand factors from one region to another. The study finds that those factors which increase the demand or quantity demanded of hospital care are directly associated with relative efficiency and vice versa.

The study also finds, however, that hospital sector response to excess demands is affected by the behavior of hospital doctors. The study hypothesizes that hospital doctors will regulate their workloads by varying average lengths of stay. The regressions of hospital beds per hospital doctor and hospital beds per nurse and midwife on average length of stay support this hypothesis. As the ratio of hospital beds to hospital doctors increases, increasing doctor workload, average lengths of stay increase. Since longer lengths of stay are associated with relative hospital sector inefficiency, workload regulation is one source of relative inefficiency and makes NHS hospitals less sensitive to excess demand pressures. Workload regulation, then, helps explain why the output elasticities for nurses and midwives were negative in both production function estimates. An increase in the number of nurses in the region provides more workload shifting capacity.

ENDNOTES

[1]James M. Henderson and Richard E. Quandt, Microeconomic Theory, 2nd Edition (New York, 1971), p. 54.

[2]George W. Wilson and Joseph M. Jadlow, "Competition, Profit Incentives, and Technical Efficiency in the Provision of Nuclear Medicine Services," The Bell Journal of Economics, Vol. 13, No. 2 (Autumn, 1982), pp. 472-482. Also see C. P. Timmer, "Using A Probabilistic Frontier to Measure Technical Efficiency," Journal of Political Economy, Vol. 69 (July-August, 1971), pp. 776-794.

[3]Wilson and Jadlow, p. 474.

[4]Finn R. Forsund, C. A. K. Lovell, and Peter Schmidt, "A Survey of Frontier Production Functions and of Their Relationship to Efficiency Measurement," Journal of Econometrics, Vol. 13 (May, 1980 Supplement), pp. 8-10.

[5]Ibid.

[6]Ibid.

[7]One limitation of using the Cobb-Douglas production function in studying the NHS regional hospital sectors should be noted. It concerns the assumption that the elasticity of substitution is one. An elasticity of substitution of one means that a one percent relative change in the ratio of the price of labor to the price of capital would lead to a one percent relative change in the capital-labor ratio. Given the low level of capital spending in the NHS and the large number of persons employed, free substitution of capital for labor may be politically impossible. Consequently, a one percent relative increase in the labor-capital price ratio may result in less than a one percent increase in the capital-labor ratio. This means the elasticity of substitution may be less than one.

[8]Timmer, p. 784.

[9]Chapter IV, p. 76.

[10]Office of Health Economics, OHE Compendium of Health Statistics, 3rd Edition (Luton, 1979), Ch. 2, p. 14; Ch. 3, pp. 25-28.

[11]OHE data provided total psychiatric and non-psychiatric beds and manpower for England. From this data, average manpower per bed ratios for England were constructed for each manpower category for both bed types. These ratios were then multiplied by the number of beds of each

type in a region to determine regional manning requirements based on these national averages. To determine the proportion of a region's actual manpower belonging to non-psychiatric hospitals, the fraction of total regional manning requirements accounted for by non-psychiatric hospital beds was multiplied times actual regional manpower in the category to obtain non-psychiatric manpower. This technique, then, implicitly assumed that any shortages or surpluses of manpower were allocated to each hospital type in proportion to the number of hospital beds of that same type in the region. For the data used in these calculations, see Office of Health Economics, Ch. 3, pp. 15-31.

[12]Both forms attempt to control for the multiproduct nature of hospital output. The model using unweighted output as the dependent variable introduces casemixes as independent variables which have a proportional effect on hospital output. The model using weighted output controls for different casemixes in the dependent variable instead.

[13]Office of Health Economics, Ch. 3, pp. 8-12.

[14]To compute weighted output, a weighted average cost per broad case type was calculated. This weighted averaged was then divided by the average cost per case for all case types. The resulting figure was then multiplied by the number of cases of that type to obtain its contribution toward total weighted output. The same procedure was also used to obtain a weighted output equivalent for out-patient visits. For the cost data used to compute weighted output, see Department of Health and Social Security, Priorities for Health and Personal Social Services in England (London, 1976), p. 107.

[15]In initial runs, casemix variables for geriatric and obstetrics-maternity cases were also included, along with outpatient visits. But the linear programming estimation could not obtain an optimal solution with these variables included. Consequently, they were omitted from subsequent runs.

[16]A. A. Walters, "Production and Cost Functions: An Econometric Survey," Econometrica, Vol. 31 (January-April, 1963), pp. 1-66.

[17]Martin Feldstein, Economic Analysis for Health Service Efficiency (Amsterdam, 1967), p. 95.

[18]Ibid., p. 98. Feldstein, in his study of 177 acute, general, non-teaching NHS hospitals, using ordinary least squares (OLS), found elasticities for nurses, doctors, and hospital beds of 0.037, 0.331, and 0.500 respectively. While the elasticities for doctors and beds were statistically significant, the elasticity for nurses was not.

[19]Office of Health Economics, Ch. 3, p. 19.

[20]Feldstein, 1967, p. 98. Feldstein found a small but not statistically significant coefficient for nurses using ordinary least squares. He found statistically significant elasticities for doctors and beds of 0.387 and 0.465 respectively.

[21]Forsund et al., p. 10.

[22]Timmer, p. 784.

[23]Ibid.

[24]Forsund et al., p. 10.

[25]Wilson and Jadlow, p. 476.

[26]Office of Health Economics, Ch. 3, p. 18.

[27]Cotton M. Lindsay, National Health Issures: The British Experience (Santa Monica, 1980), pp. 34-50.

[28]Ibid.

[29]Ibid.

[30]Ibid.

[31]Ibid.

[32]Jan Paul Acton, "Nonmonetary Factors in the Demand for Medical Services: Some Empirical Evidence," Journal of Political Economy, Vol. 83 (June, 1975), pp. 595-614.

[33]Office of Population Censuses and Surveys, The General Household Survey-1975 (London, 1978), p. 111.

[34]The impact of manufacturing wages is included in the intercept to avoid a matrix singularity which would result if all sources of wage and salary income were included as independent variables. The wage and salary figures were allocated to the regions based on the percent employment in those industries in 1976. See Central Statistical Office, Regional Statistics-1977 (London, 1976), pp. 2-21, 166-169.

[35]Acton, p. 609.

[36]Hospital doctors are paid a salary. To receive additional pay from the NHS, a doctor must work over 80 hours in a single week. Another possible source of additional NHS compensation for consultants is the distinction award. This award is based on a non-specified form of peer review. See Economic Models Limited, The British Health Care System (Chicago, 1976), pp. 85-91.

[37]For a good discussion of clinical freedom, see Michael Cooper, Rationing Health Care (New York, 1975), pp. 53-58, 107-109.

[38]Economic Models Limited, pp. 85-88.

CHAPTER VI

SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS

This dissertation has two purposes. The first is to consider the contributions to economic theory related to the public provision of health care, the zero-money price rationing of health care, and the separation of the concept of health care need from health care demand. The second major purpose of the dissertation is to empirically investigate the NHS's performance in three areas. The first area examines the allocation of health care resources among geographic regions on the basis of three criteria of equity. The second area evaluates the relative efficiency of the regions' hospital sectors and the third explores the relationship between the behavior of hospital doctors and hospital sector efficiency. By accomplishing these purposes, the dissertation seeks to provide insight into the difficulties of meeting the conflicting goals of simultaneously insuring access to health care, eliminating the financial risk of serious illness, and controlling health care costs. The British National Health Service represents an ongoing attempt to accomplish these tasks.

This final chapter first brings together the relevant contributions to the economic literature. It then summarizes the findings from the empirical research. Finally, it presents the conclusions of the study along with some recommendations for future research.

Summary

Chapter II discusses medical care and market failure. Market failure can occur both with and without externalities related to the consumption of medical care. Three possible causes of market failure exist, even in the absence of externalities. They are: the existence of irrational health care consumers; uncertainty in the diagnosis and treatment of disease; and the competitive behavior of insurance companies and resulting overinsurance.

Culyer notes that what is widely perceived as irrationality on the part of health care consumers is usually not irrationality at all but consumer ignorance.[1] Since health care information is costly to obtain, rational consumers will only obtain information to the point where their marginal benefits equal their marginal costs. Behavior attributed to ignorance, then, does not imply market failure per se. Only if an individual is insane, unconscious, or knows with certainty that the marginal benefits of health care exceed marginal costs and refuses treatment would that person be considered irrational. Culyer concludes that the existence of these persons is not a sufficient condition to justify replacement of consumer sovereignty with some other allocation mechanism.[2]

Market failure due to uncertainty in the diagnosis and treatment of disease concerns the issue of whether the consumer can accurately equate marginal benefits to marginal costs. Two institutions exist to deal with this problem. The first is the doctor-patient relationship and the second is health insurance.

When the doctor-patient relationship is complete, the consumer's demand for medical care through the doctor will be identical to his own demand if he had the doctor's knowledge. But the literature notes that fee-for-service compensation mechanisms may create a conflict of interest and result in the consumption of health care past the point where marginal benefits equal marginal costs. The literature concludes that replacing fee-for-service compensation with the kinds of profit incentives built into pre-paid health care plans would help prevent abuse of the doctor-patient relationship.

Health insurance prevents significant financial losses associated with serious illnesses and increases access to health care. It helps the consumer equate marginal benefits and marginal costs by eliminating the uncertainty associated with the costs of treating a serious illness. Market failure can result when either there is too little or too much health insurance. The former results when actual insurance premiums exceed actuarily fair premiums due to loading changes, moral hazard, and risk-pooling. The latter results from the non-profit structure of the hospital industry and the competitive behavior of insurance companies. The presence of persons who choose to self-insure does not provide a sufficient case for market failure. The literature notes that only when advantages from increased risk-spreading or economies of scale in administration exist would public provision of insurance be justified.[3] On the other hand, the government can take a positive role in curbing over-insurance by providing insurance companies with information concerning the amounts of health insurance purchased by individuals. This information would allow companies to adjust premiums according to expected losses and motivate individuals not to over-insure.[4] In addition, the government could set coinsurance rates and deductibles to provide consumers with economic incentives to reduce over-consumption of health care.[5]

Chapter II also considers market failure associated with three types of externalities concerning health care. The first type of externality concerns public health programs and little disagreement in the literature exists concerning the case for government. The second type of externality concerns what Weisbrod calls option demand.[6] Society benefits from the hospital sector's ability to satisfy peak demands. Consequently, excess capacity provides external benefits that will not be realized unless payment is received from potential customers as well as current customers. The NHS represents one possible solution to the problem but the private market alternative of pre-paid health care also exists as an alternative.[7]

The third type of externality Chapter II discusses concerns the external benefits generated by an individual's consumption of health care. This is essentially a case of interdependent utility functions where one person's or a group's utility is affected by another's consumption of medical care. The literature discusses two types of interdependency. What Culyer calls the traditional argument treats the object person's or group's consumption of medical care as the source of the externality.[8] The interdependency is the object's consumption of medical care. Per-unit subsidies provide a least-cost approach to motivating increases in medical care consumption.[9] The second type of interdependency concerns the equal distribution of medical care among all members of society. Lindsay and Buchanan note that if this type of externality is the relevant one, per-unit subsidies are not the least-cost approach to internalizing the externality. The least-cost approach requires subsidies to the indigent and some form of rationing for the other members of society.[10]

Chapter III presents the concepts of demand and need for medical care and discusses how these concepts work in the NHS framework. The demand for medical care connotates consumer sovereignty while the need for medical care refers to a third party's evaluation of a person's condition.[11] Jeffers, Bognanno, and Bartlett point out that the concept of medical care need used by health care professionals will not generally act as a constraint on the quantity of health care demanded.[12] And, as noted in the same discussion, the founders of the NHS did not recognize this fact and falsely assumed that need was simple to determine and rank on a priority basis. Since this is not the case, the zero-money pricing system adopted by the NHS to ensure equal access to health care by all persons of equal need faces serious difficulty in meeting this objective since demand would act to constrain the actual needs presented to the system. In practice, the NHS ranks those needs it can categorize among those persons that demand health care from the zero-money pricing system.

As noted above, if the relevant externality is the equality of the distribution of health care, then the least-cost way to internalize this externality is to provide consumption subsidies to the indigent while rationing the other members of society. Time prices provide a means of accomplishing both objectives. To the extent that they vary directly with income, they will increase the number of demands presented to the health care system by the indigent while decreasing the quantities demanded by the higher income groups. As the time pricing analysis demonstrates, time pricing can increase consumption equality between income groups and promote the NHS's objective of allocating health care on the basis of medical care need. But the analysis also notes that time prices can increase inequality between income groups if time prices do not vary directly with income. Additionally, time prices can increase consumption inequality within the same income group if the proportions of total income accounted by unearned income vary significantly within income groups.

Chapter IV seeks to determine if the NHS is attempting to allocate health care among geographic regions on the basis of need. As Culyer and Cooper note, it is unrealistic to expect the NHS to allocate resources among regions so as to compensate for all demand variables and ensure equal access on this basis.[13] The simple objective of this chapter is to determine if the NHS is attempting to meet its equity objective in terms of the allocation of resources among the geographic regions and in terms of the outputs of the hospital sectors of those regions.

The chapter evaluates the NHS's performance using three measures of equity. The first and simplest measure is equal per-capita allocations based on crude populations. The second measure is based on each region's population's age-sex structure's hospital utilization rates. The third measure adjusts the second by standardized mortality ratios (SMR) for specific disease categories.

In terms of expenditures per capita, the equality of the distribution of operating expenditures improved during the 1974-1977 period using both the crude population index and the age-sex SMR population based index. The equality of the distribution of capital expenditures decreased. The study also finds no evidence that this increased inequality represents attempts to redress past regional inequalities in the distribution of hospital beds.

The research also indicates a general improvement in the regional distribution of nurses and midwives, junior hospital doctors, hospital medical consultants, and general practitioners using the crude population index. But there was no improvement in the distribution of hospital beds among the regions. Using the age-sex adjusted population need measure, the findings indicate a very high positive correlation between this index and the number of general practitioners per capita but not for the other manpower categories or for hospital beds.

When the age-sex SMR adjusted population need measure is used, the study finds significant positive correlations for nurses and midwives and for hospital beds. The latter correlation for hospital beds indicates that the historical distribution of hospital beds inherited by the NHS corresponds fairly well with the geographic distribution of medical care need, as measured by this index. This finding is encouraging, given the low level of capital spending by the NHS.

Two measures of regional medical care output indicate some improvement in equality using the crude population need index. The findings indicate decreases in regional inequality in both the number of discharges and deaths per 1000 population and in the number of cases treated per available hospital bed. In addition, there is evidence of improved performance in terms of the second need measure when one controls for casemix and average lengths of stay. A statistically significant, positive, partial correlation exists between this need index and the number of discharges and deaths per 1000 population. Unfortunately, there is no evidence of increased equity among regions using these output measures and the age-sex SMR adjusted population need measure.

Chapter V presents the empirical research concerning the relative efficiency of the regions' hospital sectors and the behavior of NHS hospital doctors concerning average lengths of stay per case treated. With respect to relative efficiency, the study hypothesizes that hospitals will directly respond to demand pressure by increasing output and this will show up as increased efficiency. To test this hypothesis, demand variables are regressed on efficiency indices constructed from production function estimates for the regions' hospital sectors.

To construct the indices, the study uses two Cobb-Douglas production functions, estimated using linear programming techniques. The first model controls for the multiproduct nature of hospital output by including casemix proportions as independent variables where the dependent variable is the number of cases treated. The second model controls for the multiproduct nature by weighting hospital sector output by average cost per case. Both estimates indicate that hospital beds have the greatest impact on hospital sector output with the number of hospital doctors having the next largest impact. They also indicate that the number of nurses employed has a negative effect on output, suggesting that NHS doctors may shift workload to NHS nurses in order to regulate workload to some desired level. The third area of empirical research tests this hypothesis.

The regressions of the demand variables on the efficiency indices support the hypothesis that hospitals respond to demand pressures. Adjusted R-squares indicate that the demand variables explain approximately 72 to 88 percent of the interregional variation in the efficiency indices. The findings also indicate that the variables associated with time prices, including proxies for potential lost earnings due to illness and waiting times, vary directly with the level of relative hospital sector inefficiency. This means that as time prices increases, the number of demands presented to the NHS decreases, reducing the demand pressure on hospitals with a resultant decrease in output. The two measures of regional medical care need included in the model, the age-sex and age-sex SMR adjusted need measures, were statistically significant in the regression on the weighted output efficiency index but not in the other. They have opposite signs and their small elasticity indicates that need, as measured, has little impact on relative efficiency.

The second hypothesis tested states that NHS doctors will regulate their workload to some professionally acceptable level since their compensation is not directly a function of output. The study hypothesizes that they will do this by varying average lengths of stay. Consequently, hospital doctor behavior affects the response of the region's hospital sector to excess demands. The study finds, after controlling for other variables that would affect doctor workload and behavior, that average lengths of stay vary directly with the number of beds per hospital doctor. Consequently, varying lengths of stay to regulate workload represents one source of relative inefficiency. In addition, this finding helps explain why the output elasticities for nurses and midwives were negative in both production function estimates. An increase in the number of nurses in the region provides more workload shifting capacity.

Conclusions and Recommendations for Future Research

The empirical findings indicate that fixed salary schemes that do not reflect differences in workload result in adaptive behavior that can result in inefficiency. If fixed salaries are retained, then health care planners should recognize increased workloads by paying higher salaries in under-doctored regions. The use of a single contract rate for doctors and other personnel ignores the interregional variation in potential workloads and working environments.

In addition, more work needs to be directed toward determining the causes of relative efficiency differences among regions. The data used in this study necessitated several simplifying assumptions concerning the distribution of hospital resources among the regions. The production function estimates could be biased to the extent that the estimated resource distributions differ from actual resource distributions. A preferable approach would use individual hospitals of a given type in a region to estimate a stochastic production frontier for the region. The stochastic frontier would allow for the possibility that the regional efficiency differences calculated in this study may be, at least partially, accounted for by factors outside the NHS's control.[14] By estimating this type of production function for each region, one could be more confident of the efficiency differences among the regions and consequently, have a better index for evaluating the causes of these differences.

As the theoretical and empirical work of this dissertation indicates, attempting to simultaneously attain the three goals of insuring access to health care, eliminating financial risks associated with serious illness, and limiting the growth of health care costs is certainly a formidable task. If one pays no attention to supply, attaining the first two objectives compromises the third objective of controlling health care costs. The third party payment mechanisms used to attain the first two objectives motivate inefficient use of resources. Over-insurance, by driving a wedge between the price received by the supplier and the price paid by the consumer, encourages a continued and undesirable expansion of the health care sector. Withou controls which limit the expansion of this sector, we pay a larger and larger price for the ideal that all persons who need care should receive care.

As the review of the literature indicates, the NHS solution represents only one alternative to this end. Supply-side constraints thought to allocate scarce health care resources on the basis of need are only partially effective for two reasons. As currently applied, the concept of need used by health care professionals lacks effective benefit-cost criteria incorporated into the decision making process. The work of Culyer in this area should be extended to develop operational definitions of need that can be applied uniformly by NHS personnel.[15] Second, the NHS appears to allocate resources on the basis of need only among the subset of the population willing to pay the time price. And while time prices provide one possible means of promoting equal access, there is little direct evidence on whether they do increase equality in the consumption of medical care among different income groups. Future research on the NHS should seek to determine how effective the time-price only system is in promoting equity. If the theoretical literature on the subject is valid, society may be best served by a system that offers many different time and money price combinations, allowing the individual consumer to choose his own least-cost alternative and avoiding some of the potential inequities that can result in a time price only system where foregone earnings rates do not necessarily vary directly with demands for health care. This multiprice system, combined with effective operational definitions of need, would promote the equal access objective of the NHS as well as a more uniform distribution of NHS resources on the basis of medical care need.

ENDNOTES

[1]A. J. Culyer, "The Nature of the Commodity 'Health Care' and Its Efficient Allocation," Oxford Economics Papers, XXIII (1972), p. 191.

[2]Ibid., pp. 192-193.

[3]Mark V. Pauly, "The Economics of Moral Hazard: Comment," American Economic Review, LVIII (June, 1968), p. 535.

[4]Ibid.

[5]Martin Feldstein, "The Welfare Loss of Excess Health Insurance," Journal of Political Economy, LXXXI (March-April, 1973), p. 277.

[6]Burton A. Weisbrod, "Collective-Consumption Services of Individual-Consumption Goods," The Quarterly Journal of Economics, LXXVIII (August, 1964), pp. 471-477.

[7]Culyer, pp. 201-202.

[8]A. J. Culyer, "Medical Care and the Economics of Sharing," Economica, XXXVI (1969), pp. 351-362.

[9]C. M. Lindsay and J. M. Buchanan, "The Organization and Financing of Medical Care in the United States," Financing Health Services (London, 1970), p. 564.

[10]Ibid., p. 567.

[11]A. J. Culyer, Need and the National Health Service (London, 1976), p. 15.

[12]J. R. Jeffers, M. F. Bognanno, and J. C. Bartlett, "On the Demand Versus Need for Medical Services and the Concept of Shortage," American Journal of Public Health, LXI (January, 1971), p. 48.

[13]M. H. Cooper and A. J. Culyer, "An Economic Assessment of Some Aspects of the Operation of the National Health Service," in Financing Health Services, I. Jones (ed.) (London, 1970), p. 208.

[14]Finn R. Forsund, C. A. K. Lovell, and Peter Schmidt, "A Survey of Frontier Production Functions and Their Relationship to Efficiency Measurement," Journal of Econometrics, XIII (May, 1970 Supplement), pp. 13-14.

[15]Culyer, 1976, pp. 30-80.

BIBLIOGRAPHY

Abel-Smith, Brian. National Health Service: The First 30 Years. London: Her Majesty's Stationery Office, 1978.

Acton, Jan Paul. "Nonmonetary Factors in the Demand for Medical Services: Some Empirical Evidence." Journal of Political Economy, LXXXIII (June, 1975), pp. 595-614.

Arrow, Kenneth J. "The Welfare Economics of Medical Care." In Health Economics, M. H. Cooper and A. J. Culyer (eds.). London: Penquin, 1973, pp. 13-42.

Becker, Gary. "A Theory of the Allocation of Time." The Economic Journal, LXXV, No. 299 (September, 1965), pp. 493-517.

Central Statistical Office. Regional Statistics-1977. London: Her Majesty's Stationery Office, 1976, pp. 2-21, 166-169.

Cooper, Michael. "Economics of Need: The Experience of the British Health Service." In The Economics of Health and Medical Care, Mark Perlman (ed.). New York: John Wiley and Sons, 1974, pp. 89-107.

__________. Rationing Health Care. New York: John Wiley and Sons, 1975.

Cooper, M. H. and A. J. Culyer. "An Economic Assessment of Some Aspects of the Operation of the National Health Service." In Financing Health Services, I. Jones (ed.). London: British Medical Association, 1970, pp. 187-250.

__________. "Equity in the NHS: Intensions, Performance and Problems in Evaluation." In The Economics of Medical Care, M. M. Hauser (ed.). London: Allen and Unwin, 1972, pp. 47-60.

Culyer, A. J. "Medical Care and the Economics of Giving." Economica, XXXVIII (August, 1971), pp. 295-303.

__________. Need and National Health Service. London: Martin Roberton, 1976.

__________. "The Nature of the Commodity 'Health Care' and Its Efficient Allocation." Oxford Economics Papers, XXIII (1972), pp. 189-211.

Culyer, A. J. and J. G. Cullis. "Private Patients in N. H. S. Hospitals: Waiting Lists and Subsidies." In The Economics of Health and Medical Care, Mark Perlman (ed.). New York: John Wiley and Sons, 1974, pp. 108-116.

Davis, Karen. National Health Insurance: Benefits, Costs, and Consequences. Washington, D.C.: Brookings Institution, 1975.

Department of Health and Social Security. Health and Personal Service Statistics. London: Her Majesty's Stationery Office, 1977.

__________. Priorities for Health and Personal Social Services in England. London: Her Majesty's Stationery Office, 1976.

Economic Models Limited. The British Health Care System. Chicago: American Medical Association, 1976.

Feldstein, Martin. Economic Analysis for Health Service Efficiency. Amsterdam: North Holland Publishing Company, 1967.

__________. "Quality Change and the Demand for Hospital Care." Econometrica, XLV (October, 1977), pp. 1681-1702.

__________. "The Welfare Loss of Excess Health Insurance." Journal of Political Economy, LXXXI (March-April, 1973), pp. 251-280.

Feldstein, Paul J. "Research on the Demand for Health Services." Milbank Memorial Fund Quarterly, XLIV, No. 3, Part 2 (July, 1966), pp. 128-165.

Forsund, Finn R., C. A. K. Lowell, and Peter Schmidt. "A Survey of Frontier Production Functions and of Their Relationship to Efficiency Measurement." Journal of Econometrics, XIII (May, 1980 Supplement), pp. 5-25.

Friedman, Bernard. "On the Rationing of Health Services and Resource Availability." Journal of Human Resources, XIII (Supplement, 1978), pp. 57-75.

Grossman, Michael. "On the Concept of Health Capital and the Demand for Health." Journal of Political Economy, LXXX (March-April, 1972), pp. 223-255.

Henderson, James M. and Richard E. Quandt. Microeconomic Theory, 2nd Edition. New York: McGraw-Hill Book Company, 1971.

Holtmann, A. G. "Price, Time, and Technology in the Medical Care Market." Journal of Human Resources, VIII (Spring, 1972), pp. 179-190.

Jeffers, J. R., M. F. Bognanno, and J. C. Bartlett. "On the Demand Versus Need for Medical Services and the Concept of Shortage." American Journal of Public Health, LXI (January, 1971), p. 48.

Klarman, Herbert E. "Requirements for Physicians." American Economic Review, XLI (May, 1951), pp. 631-645.

Leftwich, Richard H. The Price System and Resource Allocation, 6th Edition. Hinsdale, IL: Dryden Press, 1976.

Levitt, Ruth. The Reorganized National Health Service. London: Croom Helm, 1976.

Lindsay, Cotton M. "Medical Care and the Economics of Sharing." Economica, XXXVI (1969), pp. 351-362.

__________. National Health Issues: The British Experience. Santa Monica: Roche Laboratories, 1980.

__________. "Option Demand and Consumer's Surplus." The Quarterly Journal of Economics, LXXXIII (May, 1969), pp. 344-346.

Lindsay, Cotton M. and Buchanan, J. M. "The Organization and Financing of Medical Care in the United States." Financing Health Services. London: British Medical Association, 1970, pp. 535-585.

Musgrave, Richard. The Theory of Public Finance. New York: McGraw-Hill, 1959.

Newhouse, Joseph P. The Economics of Medical Care. Menlo Park, CA: Addison-Wesley, 1978.

Nichols, D., E. Smolensky, and T. Tideman. "Discrimination by Waiting Time in Merit Goods." American Economic Review, LXI (1970), pp. 312-323.

Office of Health Economics. OHE Compendium of Health Statistics, 3rd Edition. Luton: White Crescent Press Ltd., 1979.

Office of Population Census and Surveys. The General Household Survey-1975. London: Her Majesty's Stationery Office, 1978.

Pauly, Mark V. "Efficiency in the Provision of Consumption Subsidies." Kyklos, XXIII, No. 1 (1970), pp. 33-57.

__________. "The Economics of Moral Hazard: Comment." American Economic Review, LXIII (June, 1968), pp. 531-537.

__________. "The Public Provision of Insurance." Quarterly Journal of Economics, LXXXVIII (February, 1974), pp. 44-62.

Pindyck, Robert and Daniel Rubinfeld. Econometric Models and Economic Forecasts, 2nd Edition. New York: McGraw-Hill Book Company, 1981.

Timmer, C. P. "Using a Probablistic Frontier to Measure Technical Efficiency." Journal of Political Economy, LXIX (July-August, 1971), pp. 776-794.

Titmuss, R. M. "Ethics and Economics of Medical Care." Commitment to Welfare. London: Allen and Unwin, 1968, pp. 247-261.

U. S. Department of Commerce, Bureau of the Census. Statistical Abstract of the United States. Washington, D.C.: Government Printing Office, 1984.

Walters, A. A. "Production and Cost Functions: An Econometric Survey." Econometrica, XXXI (January-April, 1963), pp. 1-66.

Weisbrod, Burton A. "Collective-Consumption Services of Individual-Consumption Goods." The Quarterly Journal of Economics, LXXVIII (August, 1964), pp. 471-477.

Williams, A. "'Need': An Economic Exegesis." In Economic Aspects of Health Services, A. J. Culyer and K. B. Wright (eds.). London: Martin Robertson and Co. Ltd., 1978.

Wilson, George W. and Joseph M. Jadlow. "Competition, Profit Incentives, and Technical Efficiency in the Provision of Nuclear Medicine Services." The Bell Journal of Economics, XIII, No. 2 (Autumn, 1982), pp. 472-482.

VITA

Kurt Harlan Jackson

Candidate for the Degree of

Doctor of Philosophy

Thesis: THE ZERO PRICING OF HEALTH CARE: A REGIONAL ANALYSIS OF THE BRITISH NATIONAL HEALTH SERVICE

Major Field: Economics

Biographical:

Personal Data: Born in Pauls Valley, Oklahoma, June 5, 1947, the son of Mr. and Mrs. E. L. Jackson.

Education: Graduated from Pauls Valley High School, Pauls Valley, Oklahoma, in May, 1965; received the Bachelor of Science degree in Accounting from Oklahoma State University in 1969; received the Master of Science degree in Economics from Oklahoma State University in 1977; completed the requirements for the Doctor of Philosophy degree in December, 1984.

Professional Experience: Graduate Assistant, Department of Economics, Oklahoma State University, 1969-1971; Graduate Teaching Associate, Department of Economics, Oklahoma State University, 1974-1978; Research Associate, Department of Economics, Oklahoma State University, 1980-1982; Assistant Professor of Economics, East Central Oklahoma State University, 1982-Present.