PHYSICIAN MOTIVATION AND THE STRUCTURE OF INCENTIVES IN PREPAID GROUP PRACTICE: A THEORETICAL AND EMPIRICAL STUDY

by

SARAH ELIZABETH TINKLER

A DISSERTATION

Presented to the Department of Economics
and the Graduate School of the University of Oregon
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy

December 1989

Approved: [redacted]

Dr. Christopher J. Ellis, co-chair

[redacted]

Dr. Joe A. Stone, co-chair

An Abstract of the Dissertation of

Sarah Elizabeth Tinkler for the degree of Doctor of Philosophy

in the Department of Economics to be taken December 1989

Title: PHYSICIAN MOTIVATION AND THE STRUCTURE OF INCENTIVES IN PREPAID GROUP PRACTICE: A THEORETICAL AND EMPIRICAL STUDY

Approved: [redacted]

Dr. Christopher J. Ellis, co-chair

[redacted]

Dr. Joe A. Stone, co-chair

This dissertation explores the issue of how institutional structures in the medical industry influence physician behavior. The analysis is particularly concerned with prepaid group practices, an organizational type that has tripled its enrollment since 1981. By 1988 more than thirty million Americans were enrolled in prepaid group practices.

The theoretical analysis predicts that physician labor supply is sensitive to the way in which physicians are paid. Specifically, salaried physicians supply less labor than wage earning physicians because they do not receive payment based on marginal work effort. An "ethical" physician will have a greater supply of both labor and medical

care than a "standard optimizing" physician, although still not necessarily the "appropriate" level. Finally, usage of nonphysician medical inputs also differs depending on the way in which the physician is paid. Under some incentive schemes the physician tends to overuse inputs other than his own time. The model predicts that only profit-sharing physicians (both "ethical" and "standard optimizing") and "standard optimizing" salaried physicians will use medical inputs efficiently.

The dissertation also reports a test of one of the important theoretical results: Salaried physicians supply less labor than physicians working under incentive based reimbursement. Estimation of simultaneous labor demand and supply functions using data on American primary care physicians for the year 1984 confirms the theoretical prediction. A further result of the estimation is that salaried status does not affect the hourly wage of physicians.

VITA

NAME OF AUTHOR: Sarah Elizabeth Tinkler

PLACE OF BIRTH: Preston, England

DATE OF BIRTH: February 21, 1961

GRADUATE AND UNDERGRADUATE SCHOOLS ATTENDED:

University of Oregon
Cambridge University, England

DEGREES AWARDED:

Doctor of Philosophy, 1989, University of Oregon
Master of Science, 1987, University of Oregon
Bachelor of Arts, 1982, Cambridge University, England

AREAS OF SPECIAL INTEREST:

Health Economics
Labor Economics
International Economics

PROFESSIONAL EXPERIENCE:

Assistant Professor of Economics, Weber State College, Ogden, Utah, 1988 -

Graduate Teaching Fellow, University of Oregon, Eugene, Oregon, 1982-88

ACKNOWLEDGEMENTS

Chris Ellis and Joe Stone deserve many thanks for their good natured hard work as co-chairs of my committee. In addition, Kelly Eakin and Greg Hundley provided valuable advice while serving as committee members. The many friends I made while in Oregon and Utah made my graduate school experience most enjoyable. Finally, my parents, Anne and Sam Tinkler; my aunt and uncle, B. and Harry Hulme; and my sisters and brother, Jane, Mary, and William were very supportive.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

In order to understand the production of medical care it is necessary to understand the physician since it is he who determines "the bulk of resource use, output quantities and characteristics, and total costs in the health sector" (Pauly, 2). This dissertation examines the behavior of physicians working in prepaid group practices, an innovative organizational structure that has tripled its enrollment between 1981 and 1988 to 30 million. (Kittrell, 32)

The Theoretical Model

Fee-for-service payment (cost-based reimbursement) may encourage physicians to provide excessive amounts of medical care, particularly when combined with third-party reimbursement. With the growth in importance of other methods for compensating physicians, such a capitation, attention has turned to the possibility that physicians will provide insufficient care.

A theoretical model of physician behavior in the context of a prepaid group practice (Health Maintenance Organization) is developed. It is found that a physician in a Health Maintenance Organization (HMO) will supply differing amounts of care to patients depending on the way in which he is paid. A salaried physician will always supply less care than a wage earning physician. The assumption of ethics,

defined as a concern for patient welfare, may moderate, but does not alter, the result that payment schemes are an important determinant of the level of care provided. This result is important because it means that if we believe that there is only one "appropriate" amount of care--one correct way to treat the patient--then the assumption of ethics is not sufficient to ensure that patients receive that level of care. The input mix is also a function of the incentives provided to physicians. In an HMO where medical inputs are available to the primary care physician at zero cost only the profit sharing physician will always choose an efficient combination of inputs; the salaried physician may use inputs efficiently provided he is assumed to be a standard optimizer (i.e., not "ethical").

The Empirical Work

It is clear that since a salaried physician receives no reward for marginal work effort he will supply less labor than physicians working under incentive based reimbursement. This result is, of course, applicable to all salaried physicians whether they are employed by HMOs or any other organization. Simultaneous labor demand and supply functions are estimated using 1984 data on individual physicians. It is found that employee status (which is a self-reported condition of salaried employment) has a negative effect on the weekly hours of physicians, and no significant effect on weeks worked per year or wages. It is suggested that this reduction in labor supply coupled with no reduction in wages is consistent with a

hypothesis that salaried positions arise to accommodate physician preferences over lifestyle.

Overview of Dissertation

In chapter II some of the general issues associated with the study of physician behavior are addressed. Principle among these are the institutional arrangements common in the industry and the notion of agency particularly as it relates to medical agency. In chapter III a model of an HMO physician is developed. Chapter IV examines some empirical issues associated with physician labor supply in both prepaid and conventional practices, describes the data set and reports descriptive statistics that highlight differences in employed versus self-employed physicians. Chapter V provides estimates of labor demand and supply functions for primary care physicians in prepaid and conventional practice. Chapter VI concludes this dissertation and suggests possible directions for future research.

CHAPTER II

MEDICAL INSTITUTIONS AND THE NATURE OF THE PHYSICIAN-PATIENT RELATIONSHIP

This chapter introduces some of the basic concepts used in medical economics and some of the fundamental theoretical issues that have been recognized in this field.

Medical Institutions

Fee-for-Service

The most common method for remunerating physicians is fee-for-service. The patient is billed for the value of medical services provided. Fee-for-service is accompanied by third-party reimbursement (health insurance), government assistance in the form of Medicare and Medicaid, and the charitable provision of care.

Health Maintenance Organizations

The term "HMO" is used to cover a wide variety of institutional arrangements. HMOs are "prepaid health insurance plans where the organization and participating physicians accept contractual responsibilities for the delivery of a stated benefit package of covered health services available to the enrollees" (Eastaugh, 142). A staff model HMO hires physicians as salaried employees. A group

model HMO involves an independent physician group that provides patient care to the membership of an HMO under contract with the HMO management and insurance entity. An Independent Practice Association (IPA) is a "separate legal entity that contracts with individual professionals who practice out of a traditional setting" (Eastaugh, 142). A group model HMO differs from an IPA in the legal requirement that membership of the group model HMO constitutes an individual physician's "principal professional activity" (Eastaugh, 142). Individual physicians in an IPA maintain separate practices, are organized on an open panel basis (are not responsible to a closed group of patients), and are reimbursed under fee-for-service arrangements. The growth of IPAs is seen largely as a response to the threat of HMO encroachment on markets since they enable physicians to preserve an independent practice style. I will be primarily discussing a staff model HMO, although this organization is contrasted with a profit-sharing HMO where the physician is a residual claimant.

The Nature of the Physician-Patient Relationship

Previous studies correctly identify the physician-patient relationship as being one of agency. Agency relationships arise when there are gains to specialization of knowledge and function. The physician is a highly specialized provider of medical advice and care and coordinator of other medical inputs. Since the patient is frequently ill-informed about the exact nature of his medical condition and the potential benefit to be derived from purchasing health care he hires the physician to act as his agent. There would

be no agency problem if "agents could costlessly be induced to internalize the principal's objectives" (Hart and Holstrom, 8). Arrow (1963) regards one unique feature of the medical industry as being that the degree of asymmetry of information between the principal and agent is particularly large. Furthermore, the exact nature of the product (health) is uncertain until treatment is complete. Therefore, the need for an agent and the potential for difficulty in forcing the agent to internalize the principal's objectives are both very large in the medical industry. Feldstein (1974) suggests that the physician can appropriately be regarded as the perfect agent of the patient (i.e., the physician selects the level of health care that the patient would himself choose were he fully informed) since professional ethics and standards are so ingrained in physicians it is unlikely that they would act otherwise.

According to Pauly (Pauly, 6) we can expect that in general the physician will not act as the perfect agent of his patient for the following reasons:

1. The patient alone knows how much health he desires--the asymmetry of information works both ways. The physician is better informed about the exact nature of the patient's condition and the medical options available but only the patient knows how much health he desires. As Pauly says, "on balance the patient may be better informed than the physician". (Pauly, 6)

2. While own utility maximization on the part of the physician may be tempered by concern for the patient it is unrealistic to suppose that physicians, alone among economic agents, fail to obey the laws of economic rationality.

Medical Agency under Fee-for-Service Arrangements

The Physician under Fee-for-Service

The incentives inherent in fee-for-service systems are well documented. In the absence of perfect agency fee-for-service arrangements lead to "demand creation" since the physician's income is directly tied to the level of care provided. Demand creation occurs when the physician, using his superior knowledge of the health status and medical options available to his patient, advises the purchase of medical care in excess of the amount that would maximize the patient's utility. Ferguson, agreeing with Pauly, says: "It is clearly in the physician's interest to exaggerate the gain in health that will result from treatment, since by increasing the value of H [the increment in health expected by the patient as a result of treatment]...he reduces the apparent shadow price of health and so induces the consumer to purchase more medical services." (Ferguson, 37)

The Patient under Fee-for-Service

An insured patient obtains health care at a marginal cost that does not reflect the true cost of resources used in production of that care, and so he will have an incentive to demand excessive amounts of care. It is usually assumed in the literature that the patient obtains at least as much care as he wishes and in fact physician income objectives may lead to the overprovision of care--care in excess of that which a fully informed patient would choose. To the extent that "excessive" care is reflected in rising insurance costs

insured individuals as a group may prefer a lower level of care than individual patients. Group agency is discussed in more detail below.

The Insurer under Fee-for-Service

The insurance company contracts to provide medical benefits to insured patients. The physicians and hospitals that provide care to insured patients are therefore the agents of the insurance company. Large expenditures on care are obviously contrary to the interests of the insurer. Recently, insurance companies have sought to encourage a more conservative approach to treatment by, for example, requiring second opinions for procedures that are commonly overused (such as tonsillectomy and hysterectomy).

Medical Agency in Prepaid Group Practice

The Physician in Prepaid Group Practice

In an HMO the incentives facing an individual physician will depend on the way in which the physician is remunerated. A salaried physician will have no incentive to provide more than some minimal level of care (defined as C subsequently) since his income is not directly tied to patient consumption of medical services. However, since the provision of physician time is onerous to the physician the incentive exists to provide the patient with too little care.

A physician paid a wage related to either the number of hours worked or the number of patients seen will have some financial incentive to increase the supply of his own time in the provision of

medical care. The profit-sharing physician (residual claimant) will face incentives similar to a profit maximizing HMO--to minimize the level of care provided on a per patient basis and to use an efficient combination of medical inputs.

The incentive to use excessive amounts of nonphysician inputs will exist in an HMO. (The term "nonphysician inputs" will refer to all inputs into the medical production function other than the provision of time by a primary care physician. Thus, nonphysician inputs include the use of hospital beds, nurse time, diagnostic procedures, and the time of all physicians other than the primary care physician.) Opportunities exist within the medical setting for substitution among medical inputs. For example, a physician may delegate more responsibility to nurses or order more complex diagnostic tests as an alternative to providing his own time. A salaried physician who is either required, or wishes, to provide the patient with a particular level of care can do so by using various combinations of medical inputs. Since the physician values his leisure time he will tend to substitute nonphysician inputs for his own time.

A wage earning physician may also "overuse" nonphysician inputs if he is ethical. An ethical physician is defined as one who has internalized patient utility in his utility function, and hence the coefficient on patient utility is greater than zero. Since in the simple staff model HMO nonphysician inputs are available at zero cost to the physician they will tend to be "overused" by an ethical physician. The assumption that the physician is ethical is discussed

in more detail below. In the profit-sharing model the physician incurs the cost of supplying nonphysician inputs to the patient and so the problem of inefficient usage of nonphysician inputs does not arise.

The Patient in Prepaid Group Practice

The rational patient will, given the cost associated with joining an HMO, choose an organization that, ex-ante, provides a level of care that ensures that the expected total cost is less than or equal to the expected total benefit. However, once enrolled the problem of "moral hazard" arises since the patient seeks to equate the marginal user cost of obtaining care to the marginal benefit of care. In HMOs the user cost is typically nominal, $1-$5 per visit is not uncommon, which means that in order for the organization to cover costs care must be rationed in some way.

The Hospital in Prepaid Group Practice

For the purposes of this paper I will be limiting my discussion to profit maximizing HMOs where, except in the case of the profit-sharing physician, doctor-employee stock ownership is not significant. One feature of HMOs that deserves comment is that, in effect, the insurance company becomes the provider of health care. The insurance company no longer faces an agency problem with the institutions they contract with to provide the necessary care. However, they still face an agency problem in the internal organization of the hospital since it is necessary to control physicians in order to control costs.

As in other firms, the shareholders of the hospital hire agents, in this case physicians, to coordinate the production of an output. The principal (shareholder) is concerned to maximize profit, and under capitation with total revenue fixed, this implies cost minimization. Costs are minimized at a level of output where the expected marginal cost saving associated with providing one less unit of care is equal to the expected marginal cost of any lawsuit incurred from so doing. (For a more complete discussion of this point see appendix A.) Alternatively, there may be some minimal level of care defined by professional standards below which the hospital will be unable to attract staff.

The Double Principal Agent Problem

Under capitation a double principal agent problem arises because the physician is hired by a profit maximizing HMO but he remains the agent of the patient. The objectives of hospital and patient conflict. The hospital will be concerned to maximize profits which, under capitation in a one period model, implies cost minimization. The patient, who pays a nominal user fee, will have an incentive to drive the marginal benefit of care very low, thus consuming care in excess of the amount implied by the subscriber fee paid to join the HMO.

Group Agency and the "Appropriate" Amount of Care

It is extremely difficult to identify the appropriate amount of care, at either a theoretical or an empirical level. At a theoretical level the problem has been dealt with in several different ways.

Woodward and Warren-Boulton (1984) assume that an "ethical" physician will be able to determine the appropriate amount of care for a particular patient. In their model an ethical physician is, by definition, one who selects the "appropriate" amount of care for a patient. There are two problems with this approach: (1) the appropriate amount of care is partially determined by the preferences of the patients and (2) in an HMO the problem of moral hazard will confuse the definition of the appropriate amount of care. The patient may, given the cost associated with obtaining medical coverage, choose an HMO that provides, by physician standards, insufficient care. However, once enrolled in the HMO the patient will try and obtain as much care as possible. Pauly notes that in the context of an HMO the notion of agency may be ambiguous:

> With insurance that is not individual experience rated, the patient is best off by using care as long as the value he places on the care exceeds the user cost he pays. Since all patients must pay collectively the full cost of care, all may be better off if they keep use below this level. But which of the two levels of output is the one the physician should choose in his role as agent? One of the advantages of the HMO may be precisely that it does cause the physician to behave not as the agent of his own patients, or of the fraction of the membership he treats, but rather as agent for the entire membership group. (Pauly, 60)

However, it is not clear that individual physicians are familiar with the notion of group agency. The medical school culture

tends to foster the view that it is immoral to allow cost considerations to enter into medical decision making. If the physician's own patients constitute a small part of the total number of subscribers then an individual "ethical" physician may attempt to supply more care than has, in effect, been paid for. So in any setting we will be reluctant to use physician judgment as the primary determinant of the appropriateness of a particular level of care.

In a competitive economy, ignoring distributional considerations and externalities, the appropriate level of care is the one that ensures that the marginal private cost of the last unit of care consumed is equal to the marginal private benefit. At an empirical level the appropriate level of care is unobservable for it requires that a physician be fully informed of individual patient preferences and that he select treatments without regard to his own leisure or income considerations. In chapter III we will be restricted to ranking alternative payment schemes in terms of how much care each provides without being specific about what the ideal level of care would be.

Physician Ethics

In the medical economics literature there is concern about whether or not physicians are ethical. According to Farley (1986) there are two possible checks on physician behavior: (1) ethics and (2) consumer information. If the second check on physician behavior is imperfect, as it surely must be, great importance is attached to physician ethics (or altruism). Pauly defines an ethical physician as

follows: "Given a set of actions which yields equal money income the physician will choose that course of action which most benefits his patients" (Pauly, 59). This also requires that the physician's leisure time is constant across the alternative courses of action. Woodward and Warren-Boulton (1984) assume that an ethical physician knows the "appropriate" level of care for a particular patient and seeks to provide that care.

I will define an ethical physician as one who has patient utility as an argument in his utility function. This means that the physician experiences increased utility if patient utility rises. Physician utility is assumed to be always increasing in patient utility which is proxied by the amount of medical care received by the patient. This assumption is only tenable in the context of capitation because the incentives inherent in the system mean that underprovision, not overprovision, of care is possible. Any additional care in an HMO leads to an increase in patient utility.

Woodward and Warren-Boulton's definition of ethics can be criticized on two grounds. There are great difficulties with an assumption that the physician knows patient preferences. Furthermore, it is less restrictive to assume, as I do, that physicians display <u>some</u> degree of ethical feeling and will trade-off between patient utility and other arguments of the utility function. This approach allows the "marginal cost" of additional patient utility to increase as more patient utility is supplied, and for different degrees of ethical feeling to imply different levels of care. Woodward and

Warren-Boulton's definition is all or nothing--if the physician is ethical he provides a certain level of care.

Conclusion

In this chapter some of the general issues associated with the study of medical economics were discussed. In chapter III a model of an HMO is developed to evaluate the role of ethics and remuneration schemes in the provision of care within this organizational structure.

CHAPTER III

A ONE PERIOD MODEL OF A HEALTH MAINTENANCE ORGANIZATION UNDER ALTERNATIVE REMUNERATION SCHEMES

This chapter develops a simple one period model of an HMO in order to examine the impact of different incentive schemes on the amount and type of medical care provided to a representative patient. In addition, the importance of an assumption of ethical behavior on the part of the physician is examined. Comparisons can be drawn between the different remuneration schemes postulated. It is found that different methods of payment result in differing levels of care and differing input mixes. Furthermore, an ethical physician supplies more care than a standard optimizing physician who is similarly compensated--but ethical physicians supply differing amounts of care depending on the method used to compensate them. (A standard optimizing physician is defined as a physician who does not have patient utility as an argument in his utility function.) Previous work in the area of modeling physician behavior is first discussed and a model of an HMO is then developed.

Review of the Literature

I begin with an examination of some of the most important contributions to the literature in this area. Of particular interest are previous attempts to model adequately the notion of agency and the

concept of physician ethics. Since fee-for-service is the dominant payment method it is the system most commonly modeled. The literature on HMOs is less developed--HMOs, while growing, still constitute a small segment of the market. However, articles that discuss institutional arrangements such as prospective payment are included since all "flat fee" arrangements share some of the same incentives.

Arrow (1963), looking at the role of uncertainty in shaping the structure of the medical industry, concludes that "virtually all the special features of this industry, in fact, stem from the prevalence of uncertainty" (Arrow, 946). Uncertainty is great in the medical industry for reasons that Arrow identifies as follows:

1. The demand for medical care for an individual is typically "irregular and unpredictable" which makes a consumer less able to plan such expenditures.

2. The consumer is unable to test or evaluate the product before consuming it.

3. There is no guarantee as to what the final product will be--will the consumer recover or not? Furthermore, the physician will often have more information about the likely outcome than the patient.

Arrow attributes several peculiarities of the medical industry to the above conditions. The first is the emphasis on ethics. It is widely believed that doctors are likely to put consumer interests before their own private gain more frequently than other producers. The restrictions on medical school entrance and stringent control of licensing mean that the private benefits of a medical school education exceed the private cost (although this is less true today than it was in 1963). The quality and quantity of doctors is not market determined and, at least officially, the consumer does not have any

choice concerning quality--all doctors are certified as being competent to practice and of equal quality. According to Arrow strict licensing and educational standards "are designed to reduce the uncertainty in the mind of the consumer as to the quality of the product" (Arrow, 966). Arrow asserts that another feature of the medical industry, price discrimination by income, arises in part due to a desire to provide care based on need rather than simply ability to pay. However, it also serves to assure all patients of the physician's concern for patient welfare above private gain. More recently authors have seen price discrimination under fee-for-service as a profit maximizing strategy.

There have been many changes in the medical industry since 1963. The doctor shortage has, for the most part, disappeared and rapid cost inflation has focused concern on the problems of cost control in an industry where third-party reimbursement and other institutions inhibit the operation of the market mechanism. Recent authors are less willing to take doctor ethics as a sufficiently powerful check on doctor behavior to prevent either the overprovision of care under fee-for-service or the underprovision of care when flat fees are in effect.

"The-Not-For-Profit Hospital as a Physician's Cooperative" (Pauly and Redisch 1973) sets out a model of physician behavior in a hospital that draws on models of producers' cooperatives in Yugoslavia and the USSR. The hospital is assumed to be under the control of the staff physicians who are able to adjust patient load and staff size in order to maximize the net income per member of the physician staff.

The patients either pay the full cost of care or are covered by a comprehensive insurance plan. Physicians are remunerated under fee-for-service arrangements. One implication of this model is that if a physician's income is related to his own work effort and not group income then he will tend to overprovide inputs other than his own time. Furthermore, if every patient is fully insured then the physician will order medical inputs up to the point where the marginal contribution of each to the physician's revenue is zero. In the context of physician controlled hospitals and widespread comprehensive health insurance the authors assert that the word "quality" when applied to medical procedures might be regarded as "a synonym for the application of nonphysician labor and capital in physician-income-enhancing ways". (Pauly and Redisch, 98)

One of the most hotly debated topics in the literature on fee-for-service arrangements is whether or not physicians engage in "demand creation" in order to enhance their income. If we assume that physicians are evenly distributed across institutional arrangements in terms of their degree of ethical feeling then evidence from the fee-for-service arena that physicians will trade patient well-being for income gains can be regarded as evidence that physician ethics provide an inadequate check on physician behavior under <u>any</u> institutional arrangement.

Reinhardt (1985) sees the problem of demand creation as having three discrete parts. Firstly, is it possible for physicians to control patient demand for physician services? Secondly, if physicians are able to manipulate patient demand do they do so for

financial reasons? Thirdly, if they do manipulate patient demand do they do so consistently or is there a variable level of exploitation (with a rise in exploitative behavior when physicians are faced with increased economic pressure)?

Reinhardt says the evidence suggests that physicians are engaged in satisficing behavior: physicians seem to manipulate patient demand for their services but they are not income maximizers--rather they try to maintain a target income. Therefore, demand inducement is usually evident only in times of economic pressure. In support of this contention he quotes Newhouse (1970) who finds a positive correlation between physician fees and physician density. This might be regarded as evidence that physicians respond to increased competition by raising fees.

Pauly (1980) in his book Doctors and Their Workshops gives a comprehensive survey of the theory of physician behavior in a hospital setting up until that date. One of the most important observations he makes is that while there may be variations in the level of care provided, if the physician is a residual claimant he will always have an incentive to minimize cost. The physician:

> will not order a high cost input where a low-cost input will do...The physician may not act as his patients' agent in choosing the level of health or well-being but he will act as such an agent in choosing the level of other inputs for a given level of health. (Pauly, 113)

However, the incentive to minimize cost will be weaker in prepaid practice since user costs do not affect physician income so directly.

Ellis and McGuire (1986) assert that Pauly's conclusion, that the physician under fee-for-service always minimizes cost, derives

from his assumption that the patient and the physician agree in advance what the total expenditure on an episode of treatment will be, and the physician then has the task of allocating this expenditure between the different medical inputs. Several authors, among them Reinhardt (1972), find that physicians do not seem to use medical inputs efficiently. Pauly says this is because physicians "do not want to take the effort to seek the most efficient method of practice" (Pauly, 13).

Pauly introduces the concept of a minimum level of health that the physician cannot allow the patient to fall below. This minimum is determined by either: (1) the threat of malpractice actions or (2) "undesirable competitive repercussions" or (3) the consumer's prior knowledge of what type of outcome he can expect. He proposes that physicians may be observed acting as "partially benevolent oligopolists." Such a physician raises his income by increasing treatment and not by raising price. If the physician raises price he may face increased competitive pressure from other physicians. In addition, if the physician has some ethical feelings he can justify this strategy on the grounds that it provides some additional benefits to the patient.

In discussing the different types of incentive schemes Pauly agrees with most authors by concluding that a fee-for-service physician will typically provide too much care, and the existence of third-party reimbursement means the patient will normally concur with the suggested treatment plan. He suggests that the problem arises because the fees physicians receive are incorrect--an office visit is

undervalued relative to a high-tech procedure or a hospital stay. Fee-for-service can be made efficient simply by rearranging the fee schedule to make certain procedures less lucrative.

At a more general level, Pauly says the ideal situation would be one where "the physician makes the same amount of income no matter how he spends his time with patients [then] he might as well choose the way which most benefits his patients" (Pauly, 59). However, the amount of leisure time would also need to be fixed over all possible treatment strategies for this to be true--and it is obvious that some treatments, such as an office visit, use more physician time than others.

Ferguson (1985) looks at efficiency of input use under alternative institutional arrangements. Pauly asserts that any physician who is a residual claimant will use his own input as efficiently as any purchased input, a point that Ferguson feels is most applicable to the case of a physician who is a residual claimant in a prepaid group practice and who therefore "faces the same sort of incentives as does any neoclassical producer" (Ferguson, 38).

Ferguson's model incorporates the Pauly assumption that the patient and physician agree in advance on the amount that will be spent on treatment of a particular episode of illness. Obviously with fee-for-service under third-party reimbursement this amount is potentially very high. He incorporates a malpractice constraint into his model but since he is looking at fee-for-service arrangements he assumes (rightly) that it is not binding. However, the prior

agreement on expenditure is a binding constraint in his model. He models health as being produced in the following way:

$$S = S(M_1, M_2, S_0)$$

Where: S_0 = initial endowment of health

M_1 = an input produced using physician time

M_2 = an input that does not entail the use of physician time

M_1 is produced in the following way:

$$M_1 = M_1(T, N)$$

Where: T = physician time per patient

N = other inputs per patient (such as office time, nurse time)

This distinction between the production of physician time and the application of other medical inputs would seem appropriate in the fee-for-service setting where physicians typically maintain their own offices.

The physician maximizes the following utility function by choice of T, N, A, and M_2:

$$U = U(A[P_1M_1(T,N) - WN], V - AT)$$

Where: A = number of patients treated

W = payment to input N

V = total available hours

The physician's problem is to maximize utility subject to the expenditure constraint:

$$E = P_1M_1(T,N)+P_2M_2$$

P_1 and P_2 are the prices of the two inputs, which are treated as exogenous. Ferguson shows that a physician will choose an efficient mix of T and N in his production of M_1. In the absence of the malpractice constraint the doctor will select N, T and A efficiently, and then order M_2 up to the point where the previously agreed expenditure limit is exhausted. The most likely outcome would see the physician:

> overusing M_2 when the patient has no trouble financing expenditures...The implications for a medical system where most patients have third party insurance seem to be that when insurance companies tighten up on payment, the impact should be chiefly on M_2-type inputs. (Ferguson, 41)

Ferguson shows that if a physician incurs the full cost of his choice of inputs then he will produce care efficiently. However, he will not select an efficient level of "free" inputs. This result carries over into the HMO setting in my model.

Pauly, responding to Ferguson, reiterates his point that a physician will adopt cost minimizing behavior in a 1985 editorial in the Journal of Health Economics. He states that "the identical cost minimizing behavior will be chosen by the physician even if he does not pay for some of the inputs because by selecting the cost minimizing level of inputs the patient buys from others, the physician

maximizes the total net income he can get from his own services" (Pauly, 80).

Ellis and McGuire (1986) analyze the double principal agent problem that arises when payment is determined using Diagnostic Related Groups (DRGs), a flat fee system for reimbursing for care. The physician is seen as maximizing a utility function that includes both a hospital profit variable and a patient utility variable, physician consumption and leisure drop out of the physician's utility function because they "ignore any possible substitutability or complementarity between physician inputs and hospital services, and assume that the physician's input for a given episode is fixed" (Ellis and McGuire, 132). Hospital profits are justified in the physician's utility function on the grounds that the hospital can penalize the physician for not taking hospital interests into account by the loss of admitting privileges. A perfect agent is one who values "one dollar of benefit to the hospital (i.e., profit) equally with one dollar of benefit to the patient" (Ellis and McGuire, 131). If a physician fails to be a perfect agent (by the authors' criteria) then, according to Ellis and McGuire, it will usually be in the direction of favoring the hospital because of its economic power over the physician.

This definition of perfect agency seems unsatisfactory for two reasons. With a fixed fee per treatment episode the hospital has an incentive to minimize cost and the patient, who pays no marginal user cost, seeks additional treatment if the marginal benefit is greater than zero. The perfect agent should ensure that treatment decisions

reflect the shadow price of health which may mean a level of care that is greater or less than the care that would obtain under the perfect agency criteria which they suggest. Secondly, it seems a great weakness of this paper that they analyze the physician's problem by first assuming physician leisure and consumption out of the physician's utility function. Ellis and McGuire note that, unlike cost based reimbursement where the objectives of all actors except insurance companies are compatible, under any fixed fee system the objectives of the hospital and patient will conflict.

Woodward and Warren-Boulton (1984) look at the physician's treatment decision under three possible incentive schemes: (1) fixed, (2) time-based and (3) output-based income. Fixed income corresponds to the salaried physician in my model except they assume the physician incurs the cost of nonphysician inputs. This is an unlikely situation since salaried physicians normally work in an environment where other inputs are provided. The time-based physician is compensated per number of hours worked--again bearing the full cost of nonphysician inputs. Output-based income corresponds to fee-for-service.

Physicians are assumed to be ethical, and utility is a function of the production of leisure activities and the provision of medical care on a per patient basis up to an "appropriate" amount of care. (The "appropriate" level of care is discussed in chapter II in the section "Group Agency and the Appropriate Amount of Care.") Medical output is a function of expenditure on practice inputs, the physician's own time and the number of patients. Woodward and Warren-Boulton find that the output-based physician will "overprovide"

medical care; fixed and time-based physicians will necessarily "underprovide" care--even if the physician is assumed to be ethical. Underprovision of care arises in this model because physicians are assumed to bear the full cost of other inputs. This same assumption also results in care being produced efficiently. One weakness of this paper is the absence of a minimum level of care constraint which is necessary wherever the incentive exists to underprovide care.

Farley, in a 1986 survey article, stresses the problem of controlling provider behavior in a market where the asymmetries of information are large. According to Farley, the literature identifies two possible constraints to physician profits. The first is ethics--poor treatment enters negatively into the physician's utility function. The second is patient knowledge--the physician's practice size is a function of the quality of service provided. Typically, the patient acts alone in deciding whether to seek treatment for a particular condition and in choosing a physician but once a physician has been contacted the patient is normally expected to concur with whatever the physician suggests. This allows physicians to manipulate their income under fee-for-service by their control over both prices and the level of care, which creates a dilemma for the government:

> If physicians set prices (subject to market or ethical constraints) the conflict of interest between patient and physician is confined to the issue of the physician's financial compensation and does not systematically influence the process of care. If prices are set by the government or some other authority, by contrast, physicians can only affect their incomes through their treatment decisions. (Farley, 329)

Farley says that by virtue of their superior knowledge physicians are cast in the role of social planners and as such should be judged against "the twin standards of equity and efficiency" (Farley, 331).

Of the articles discussed above, Woodward and Warren-Boulton (1984) is most closely related to my work. The points of departure are: (1) the HMO setting, (2) the definition of ethical behavior (which in my work is seen as the provision of additional care since in an HMO the issue of the overprovision of care typically does not arise) and (3) the assumption that physicians in an HMO, unless they are residual claimants, do not bear the cost of the provision of nonphysician inputs into the medical production function.

The Theoretical Model

The institutional setting that will be described here is that of a staff model HMO that employs physicians either on a salary or pays its physicians a wage. In addition, a profit-sharing physician who is a residual claimant in an HMO that is entirely physician owned will be described. The components of the model are described below.

Total Revenue

The HMO has a fixed number of subscribers who each pay a membership fee of P_h. In theory, once P_h has been paid the member has access to all "necessary" medical care during the period of enrollment. The one period nature of this model precludes the possibility of examining the interrelationship between HMO behavior

and the number of subscribers, and hence revenue, in subsequent time periods.

Physician Utility

Following standard practice in this literature the physician is assumed to determine the amount and type of medical care provided to the patient. This assumption is realistic since physicians have a great deal of control over output and input mix. Therefore, the amount and type of medical care provided is selected by the physician to maximize physician utility. The physician utility function is assumed, for simplicity, to be of the Cobb-Douglas form:

$$U_p = (T-M_1)^{\alpha}(X)^{\beta}(U_s)^{\tau}$$

Where: U_p = physician's utility function

T = total discretionary time available to the physician

M_1 = input of physician time into the medical production function

X = composite good available to the physician

U_s = the representative patient's utility function

α, β and τ are coefficients on the physician's utility function. α, $\beta > 0$, and $\tau \geq 0$. U_p is increasing in leisure, income and patient utility and is concave. Assuming the Cobb-Douglas functional form simplifies the analysis although it does impose certain restrictions on the results. The one that is most important to this work is that

it embodies constant budget shares and thus does not permit a backward bending supply curve of labor.

Physician Income

Three possible schemes for remunerating the physician are identified:

1. The "salaried" physician receives a fixed income regardless of the number of hours worked above some minimal requirement, M_1, set by the employing HMO. The physician has discretion to work more than the minimum if, for ethical reasons, he regards the provision of care associated with M_1 as being inadequate. This additional supply of hours or effort may be regarded as charitable in this context.

2. The "wage earning" physician receives an hourly wage, w_1, total income is thus equal to w_1M_1.

3. The "profit sharing" physician is a residual claimant in the HMO. Income per patient is equal to $(P_c-P_2M_2)/N$, where P_2 is the cost of a unit of M_2, the nonphysician input into the production function. N is the number of physician-owners of the HMO. P_c is the average revenue per treatment episode. This income scheme incorporates the assumption that all costs and revenues are shared equally among the physicians in the group.

In the case of the salaried and wage earning physician notice that the HMO covers the cost of supplying M_2 to the patient.

Patient Utility

U_s is defined as follows:

$$U_s = (C)$$

Where C = units of care provided to the patient.

U_s is increasing in care and is concave. Given the incentives inherent in a capitated system the issue of overprovision of care does not

arise. Furthermore, in a one-period model where a fixed subscriber fee has been paid no consideration need be given to patient demand for other goods. Note that it is assumed that the patient incurs no time cost in receiving medical care. Under capitation user fees are typically small so, for simplicity, the user fee is set equal to zero. This means that the enrollee will want to drive the marginal benefit of care to zero.

Care is strictly speaking an intermediate good and the patient actually seeks to consume health. However, modeling the production of health is problematic. Most authors assume that the output of the medical industry is medical care not health. The relationship between medical care provided to a patient and the level of health enjoyed by the patient is discussed further in appendix B.

The Medical Care Production Function

The production function is assumed to be of the Cobb-Douglas form:

$$C = M_1^{\phi_1} M_2^{\phi_2}$$

ϕ_1 and ϕ_2 are coefficients on the medical care production function, ϕ_1, $\phi_2 > 0$. C is concave and is increasing in M_1 and M_2.

The Constraints

Income Constraint

$$Y = P_x X$$

Where P_x is the price of the composite commodity X

Minimum Level of Care Constraint

$$C \geq \bar{C}$$

Where $\bar{C}$ is the minimum level of care. $\bar{C}$ is determined by the HMO and it constitutes the profit maximizing level of care for the HMO. It is not necessarily the amount of care that is actually provided to the patient since it is the physician not the hospital who determines the level of care. The minimum level of care constraint is discussed further in appendix A.

Rationed Level of Nonphysician Inputs

$$M_2 \leq \bar{M}_2$$

In the wage earning and salaried cases the hospital imposes a cap on the level of M_2 that the physician can order for a treatment episode. This can take the form of explicit guidelines or be simply the provision of a rationed amount of M_2 so that the physician must compete with other doctors in the system for scarce resources. The hospital can observe the amount of M_2 ordered by a physician. The hospital will choose $\bar{M}_2$ so that it is the cost minimizing level of M_2 associated with the level of care $\bar{C}$.

Constraint on Salaried Physician Time

$$M_1 \geq \bar{M}_1$$

The HMO will require that the physician work a minimum number of hours, or alternatively, given the number of patients, work sufficient hours so that $\bar{C}$ is supplied to a representative patient. Effort is treated as unobservable here and is assumed to be constant.

Ethical Behavior

For a standard optimizing physician $\tau = 0$, where τ is the coefficient on patient utility in the physician's utility function. For the ethical physician $\tau > 0$. It is assumed that an ethical physician, except in the case of the profit sharing physician who is not subject to the constraint on the usage of M_2, will always choose to use the constrained level of M_2 ($\bar{M}_2$) since this increases patient utility and is available at no cost to himself.

Determination of the Level of Care and the Input Mix

The level of care provided and the input mix chosen is different for each of the six cases outlined below. Except for the salaried case where the results are intuitively obvious the method used is to substitute into U_p for the constraints, differentiate with respect to M_1, and solve for M_1. The level of M_2 used in the profit sharing ethical case is determined theoretically, for the other cases it is assumed that $\bar{M}_2$ is provided for reasons explained for each case below. Information about M_1 and M_2 usage in each case is used to determine the level of C that the physician will select. Comparisons between the levels of M_1, M_2 and C that arise in the six cases can then be made.

Salaried Physician (S)

For the salaried physician, income = Y and $\tau = 0$ (which means that patient utility drops out of the physician's utility function). Substituting into the utility function for the income constraint we get:

$$U_p = (T-M_1)^{\alpha}(Y/P_x)^{\beta}$$

Income is independent of the amount of M_1 supplied and therefore the physician will choose to supply $\bar{C}$ units of care. ($\bar{C}$ is the malpractice constrained minimum level of care.) Since the physician values leisure he will produce $\bar{C}$ using $\bar{M}_2$, the maximum possible amount of M_2. Since $\bar{M}_2$ is selected by the employing hospital with consideration to the relative cost of M_1 and M_2 the hospital implicitly sets the level of M_1 at $\bar{M}_1$. From the production function:

$$\bar{M}_1 = [\bar{C}/\bar{M}_2^{\phi_2}]^{1/\phi_1}$$

If the hospital sets $\bar{M}_2$ as the efficient level of M_2 (efficient in terms of the relative price of the two inputs) based on the assumption that $\bar{C}$ is the level of care provided then $\bar{M}_1$ will be the efficient level of M_1. So $\bar{C}$ will be produced using the cost minimizing combination of M_1 and M_2. This can be regarded as the benchmark case since it represents the level of care that a profit maximizing HMO would wish to provide, and it is provided using the cost minimizing combination of M_1 and M_2.

Salaried Ethical Physician (SE)

For the salaried ethical physician income is again equal to Y. $\tau > 0$ (the physician attaches positive value to additional care provided since it raises patient utility). Since the physician is ethical he always supply the patient with $\bar{M}_2$ since it raises patient utility and is available at no cost to himself. The physician maximizes the following function:

$$U_p = (T-M_1)^{\alpha}(Y/P_x)^{\beta}(M_1^{\phi_1}\bar{M}_2^{\phi_2})^{\tau}$$

The first order condition is:

$$\delta U_p/\delta M_1 = -\alpha(T-M_1)^{\alpha-1}(Y/P_x)^{\beta}(M_1^{\phi_1}\bar{M}_2^{\phi_2})^{\tau}$$

$$+ (T-M_1)^{\alpha}(Y/P_x)^{\beta}(M_1^{\phi_1}\bar{M}_2^{\phi_2\tau-1}\tau\phi_1 M_1^{\phi_1-1}\bar{M}_2^{\phi_2} = 0$$

Solve for M_1 by rearranging the first order condition:

$$M_1 = \left[\frac{\tau\phi_1}{\alpha+\tau\phi_1}\right]T$$

Therefore:

$$C = \left[\frac{\tau\phi_1 T}{\alpha+\tau\phi_1}\right]^{\phi_1}\bar{M}_2^{\phi_2}$$

M_1 and C are functions of input productivity (ϕ_1 and ϕ_2), the degree of ethical feeling on the part of the physician (τ) and the value that the physician attaches to leisure (α). The partial derivatives of M_1 and C with respect to α, τ, and ϕ_1 are as follows:

$$\delta M_1/\delta\alpha = \frac{-\tau\phi_1 T}{(\alpha+\tau\phi_1)^2} \qquad < 0$$

$$\delta M_1/\delta\tau = \frac{\alpha\phi_1 T}{(\alpha+\tau\phi_1)^2} \qquad > 0$$

$$\delta M_1/\delta\phi_1 = \frac{\alpha\tau T}{(\alpha+\tau\phi_1)^2} \qquad > 0$$

$$\delta C/\delta\alpha = \bar{M}_2^{\phi_2\phi_1}\left[\frac{\tau\phi_1 T}{\alpha+\tau\phi_1}\right]^{\phi_1-1}(\delta M_1/\delta\alpha) \qquad < 0$$

$$\delta C/\delta\tau = \bar{M}_2^{\phi_2\phi_1}\left[\frac{\tau\phi_1 T}{\alpha+\tau\phi_1}\right]^{\phi_1-1}(\delta M_1/\delta\tau) \qquad > 0$$

$$\delta C/\delta\phi_1 = \bar{M}_2^{\phi_2}\left[\frac{\tau\phi_1 T}{\alpha+\tau\phi_1}\right]^{\phi_1}\left[\ln\left[\frac{\tau\phi_1 T}{\alpha+\tau\phi_1}\right]\right](\delta M_1/\delta\phi_1) \qquad > 0$$

M_1 and C are increasing in ϕ_1 and τ and decreasing in α. ϕ_1 is the parameter on the productivity of M_1 in the production of medical care. This implies that a greater level of M_1 and C will be provided the higher is the productivity of the physician in the production of medical care. The physician in this case is salaried and does not receive additional income as a result of being more productive, ϕ_1 raises physician utility via its effect on patient utility. For the profit sharing and wage earning cases ϕ_1 also raises physician utility via its effect on income. τ represents the degree of ethical feeling on the part of the physician, so the greater the degree of ethical feeling that the physician displays the more M_1, and hence patient care, he will supply. Since α measures the value that the physician

places on additional leisure the higher is his valuation of leisure the lower the amount of M_1 he will supply.

Wage Earning Physician (W)

For the wage earning physician $\tau = 0$ and income is equal to w_1M_1, therefore:

$$Up = (T-M_1)^{\alpha}(w_1M_1/P_x)^{\beta}$$

The first order condition is:

$$\delta Up/\delta M_1 = -\alpha(T-M_1)^{\alpha-1}(w_1M_1/P_x)^{\beta}$$

$$+ \beta(T-M_1)^{\alpha}(w_1M_1/P_x)^{\beta-1}(w_1/P_x) = 0$$

Solve for M_1 by rearranging the first order condition:

$$M_1 = \frac{\beta T}{\alpha+\beta}$$

The amount of care provided also depends on the level of M_2 the physician chooses to use. Since the physician in this case is not "ethical" we cannot say categorically that he will choose to supply $\bar{M}_2$. However, since M_2 is available at no cost to the physician we also cannot say a priori that he will not supply $\bar{M}_2$ units of nonphysician inputs. The range for C will be as follows:

$$\bar{C} \leq C \leq \left[\frac{\beta T}{\alpha+\beta}\right]^{\phi_1}\bar{M}_2^{\phi_2}$$

The upper limit on C implies use of $\bar{M}_2$, the constraint on M_2.

The partial derivatives of M_1 and C with respect to α, β, and ϕ_1 are as follows:

$$\delta M_1/\delta\alpha = \frac{-\beta T}{(\alpha+\beta)^2} \qquad < 0$$

$$\delta M_1/\delta\beta = \frac{\alpha T}{(\alpha+\beta)^2} \qquad > 0$$

$$\delta M_1/\delta\phi_1 \qquad = 0$$

$$\delta C/\delta\alpha = M_2^{\phi_2}\phi_1\left[\frac{\beta T}{\alpha+\beta}\right]^{\phi_1-1}(\delta M_1/\delta\alpha) \qquad < 0$$

$$\delta C/\delta\beta = M_2^{\phi_2}\phi_1\left[\frac{\beta T}{\alpha+\beta}\right]^{\phi_1-1}\left[\frac{T(\alpha+\beta)-\beta^2 T}{(\alpha+\beta)^2}\right] > 0$$

$$\delta C/\delta\phi_1 = M_2^{\phi_2}\left[\frac{\beta T}{\alpha+\beta}\right]^{\phi_1}\ln\left[\frac{\beta T}{\alpha+\beta}\right](\delta M_1/\delta\phi_1) = 0$$

β measures the value that the physician attaches to additional income. M_1 and C are increasing in β (the higher valuation that the physician attaches to additional income the larger the amount of M_1 and hence C he will supply). Again, M_1 and C are decreasing in α, the coefficient on the physician's leisure time. The partial derivatives of M_1 and C with respect to ϕ_1 are equal to zero because the standard optimizing physician is concerned solely with the tradeoff between income and leisure in making the decision about how much M_1 to provide.

Wage Earning Ethical Physician (WE)

The physician receives an income equal to w_1M_1 and, due to the assumption of ethical behavior, $\tau > 0$. M_2 is available at no cost to the physician and so an ethical physician will always choose $\bar{M}_2$ since it increases patient utility. The physician maximizes the following function:

$$U_p = (T-M_1)^{\alpha}(w_1M_1/P_x)^{\beta}(M_1^{\phi 1}\bar{M}_2^{\phi 2})^{\tau}$$

The first order condition is:

$$\delta U_p/\delta M_1 = -\alpha(T-M_1)^{\alpha-1}(w_1M_1/P_x)^{\beta}(M_1^{\phi 1}\bar{M}_2^{\phi 2})^{\tau}$$

$$+ \beta(T-M_1)^{\alpha}(w_1M_1/P_x)^{\beta-1}(w_1/P_x)(M_1^{\phi 1}\bar{M}_2^{\phi 2})^{\tau}$$

$$+ \tau(T-M_1)^{\alpha}(w_1M_1/P_x)^{\beta}(M_1^{\phi 1}\bar{M}_2^{\phi 2})^{\tau-1}\phi_1M_1^{\phi_1-1}\bar{M}_2^{\phi 2} = 0$$

Solve for M_1 by rearranging the first order condition:

$$M_1 = \left[\frac{\beta+\tau\phi_1}{\alpha+\beta+\tau\phi_1}\right]T$$

Therefore:

$$C = \left[\frac{\beta+\tau\phi_1 T}{\alpha+\beta+\tau\phi_1}\right]^{\phi 1}\bar{M}_2^{\phi 2}$$

The partial derivatives of M_1 and C with respect to α, β, τ, and ϕ_1 are as follows:

$$\delta M_1/\delta\alpha = \frac{-(\beta+\tau\phi_1)T}{(\alpha+\beta+\tau\phi_1)^2} \qquad < 0$$

$$\delta M_1/\delta \beta = \frac{\alpha T}{(\alpha+\beta+\tau\phi_1)^2} > 0$$

$$\delta M_1/\delta \tau = \frac{\alpha\phi_1 T}{(\alpha+\beta+\tau\phi_1)^2} > 0$$

$$\delta M_1/\delta \phi_1 = \frac{\alpha\tau T}{(\alpha+\beta+\tau\phi_1)^2} > 0$$

$$\delta C/\delta \alpha = \bar{M}_2^{\phi_2}\phi_1\left[\frac{\beta+\tau\phi_1 T}{\alpha+\beta+\tau\phi_1}\right]^{\phi_1^{-1}}(\delta M_1/\delta\alpha) < 0$$

$$\delta C/\delta \beta = \bar{M}_2^{\phi_2}\phi_1\left[\frac{\beta+\tau\phi_1 T}{\alpha+\beta+\tau\phi_1}\right]^{\phi_1^{-1}}(\delta M_1/\delta\beta) > 0$$

$$\delta C/\delta \tau = \bar{M}_2^{\phi_2}\phi_1\left[\frac{\beta+\tau\phi_1 T}{\alpha+\beta+\tau\phi_1}\right]^{\phi_1^{-1}}(\delta M_1/\delta\tau) > 0$$

$$\delta C/\delta \phi_1 = \bar{M}_2^{\phi_2}\left[\frac{\beta+\tau\phi_1 T}{\alpha+\beta+\tau\phi_1}\right]^{\phi_1}\left[\ln\left[\frac{\beta+\tau\phi_1 T}{\alpha+\beta+\tau\phi_1}\right]\right](\delta M_1/\delta\phi_1) > 0$$

M_1 and C are increasing in β, τ, and ϕ_1, the coefficients on the utility function for physician income and patient utility, and the coefficient on the production function for physician time respectively. The physician will supply a greater amount of his own input to the medical production function the higher the value he attaches to additional income, the more ethical he is, or the more productive he is in the production of medical care. M_1 and C are decreasing in α, the coefficient on physician leisure time in the physician's utility function. Less M_1 and hence C is supplied the greater the value the physician attaches to additional leisure time.

Profit Sharing Physician (P)

The profit sharing physician receives an income per treatment episode equal to $(P_c - P_2M_2)/NP_x$. The standard optimizing physician chooses to supply $\bar{C}$. The hospital selects $\bar{M}_2$ to be the efficient level of M_2 if $\bar{C}$ is the level of care provided, consistent with the relative cost of M_1 and M_2. This means that the profit sharing physician will select the same level of M_2 as the salaried physician and, with $\bar{C}$ chosen in each case the level of M_1 provided will again be:

$$\bar{M}_1 = [\bar{C}/\bar{M}_2^{\phi_2}]^{1/\phi_1}$$

Since the physician is a residual claimant and therefore incurs the full cost of M_2 he selects the efficient level of M_1 and M_2.

Profit Sharing Ethical (C Constrained to $\bar{C}$)

The profit sharing ethical case is first considered by assuming that the physician chooses the level of care $\bar{C}$. This assumption is then dropped and the behavior of the physician when $\bar{C}$ is nonbinding is examined. Income is equal to $(P_c - P_2M_2/NP_x)$ and $\tau > 0$.

The profit sharing physician will always choose to produce whatever level of care is selected as cheaply as possible. Therefore $\bar{C}$ will be produced using $\bar{M}_2$ and $\bar{M}_1$. This arises because the physician must pay the cost of M_2 so he always chooses to produce care using the cost minimizing combination of inputs, whether he is ethical or not. In the salaried ethical and wage earning ethical cases the amount of nonphysician inputs was constrained to $\bar{M}_2$ and an ethical physician

might "overuse" M_1 in order to supply a particular level of care. In a profit-sharing HMO the level of care may be inadequate if the physician is not ethical--but the choice of input mix will be efficient.

Note that this case also implies the same level of care and input mix that would prevail in the standard optimizing salaried case. Again, if $\bar{C}$ were market determined and represented the market clearing level of health care as opposed to the malpractice constrained level of care then the institutional arrangements of either salaried standard optimizing or profit sharing standard optimizing would be preferred--and ethical behavior leads to an overprovision of care. However, it seems more likely that $\bar{C}$ constitutes an inadequate level of care and if these arrangements are in effect we rely on physician ethics to ensure that the appropriate level of care is received by patients.

Profit Sharing Ethical ($\bar{C}$ Nonbinding)

Income is equal to $(P_c - P_2M_2/NP_x)$ and $\tau > 0$. The physician maximizes the following function:

$$U_p = (T-M_1)^{\alpha}\left[\frac{P_c - P_2M_2}{NP_x}\right]^{\beta}(M_1^{\phi_1}M_2^{\phi_2})^{\tau}$$

The first order conditions are:

$$\delta U_p/\delta M_1 = -\alpha(T-M_1)^{\alpha-1}\left[\frac{P_c - P_2M_2}{NP_x}\right]^{\beta}(M_1^{\phi_1}M_2^{\phi_2})^{\tau}$$

$$+ (T-M_1)^{\alpha} \left[\frac{P_c-P_2M_2}{NP_x}\right]^{\beta\tau} (M_1^{\phi_1}M_2^{\phi_2})^{\tau-1} M_2^{\phi_2}\phi_1 M_1^{\phi_1-1} = 0$$

$$\delta U_p/\delta M_2 = -\beta(T-M_1)^{\alpha}\left[\frac{P_c-P_2M_2}{NP_x}\right]^{\beta-1}(P_2/NP_x)\,(M_1^{\phi_1}M_2^{\phi_2})^{\tau}$$

$$+ (T-M_1)^{\alpha} \left[\frac{P_c-P_2M_2}{NP_x}\right]^{\beta\tau} (M_1^{\phi_1}M_2^{\phi_2})^{\tau-1} M_1^{\phi_1}\phi_2 M_2^{\phi_2-1} = 0$$

Solve for M_1 by rearranging the first order condition with respect to M_1:

$$M_1 = \left[\frac{\tau\phi_1}{\alpha+\tau\phi_1}\right]T$$

Solve for M_2 by rearranging the first order condition with respect to M_2:

$$M_2 = \frac{P_c\tau\phi_2}{P_2(\beta+\tau\phi_2)}$$

Therefore:

$$C = \left[\frac{\tau\phi_1 T}{\alpha+\tau\phi_1}\right]^{\phi_1}\left[\frac{P_c\tau\phi_2}{P_2(\beta+\tau\phi_2)}\right]^{\phi_2}$$

Note that this is the only case in which M_2 is explicitly determined by the optimizing physician. This combination of M_1 and M_2 constitutes the efficient input mix for any level of care that the profit sharing ethical physician selects. This is the same level of M_1 that is provided in the salaried ethical case. The salaried ethical physician combines this amount of M_1 with the constrained level of M_2 ($\bar{M}_2$) set by the employing hospital. In the discussion of the salaried ethical

physician we observed that $\bar{M}_2$ is the efficient level of M_2 based on an assumption that $\bar{C}$ is supplied. We show later that $C^{PE} > C^{P}$ so this necessarily means that $\bar{M}_2$ cannot be the efficient level of M_2 to supply an amount of care greater than $\bar{C}$. Therefore, this case implies a greater usage of M_2 than the salaried ethical case. Since more M_2 is combined with the same level of M_1 the profit sharing ethical physician supplies more care than the salaried ethical physician and, furthermore, the input mix is efficient.

The partial derivatives of M_1, M_2 and C with respect to the parameters are as follows:

$$\delta M_1/\delta\alpha = \frac{-\tau\phi_1 T}{(\alpha+\tau\phi_1)^2} \qquad < 0$$

$$\delta M_1/\delta\tau = \frac{\alpha\phi_1 T}{(\alpha+\tau\phi_1)^2} \qquad > 0$$

$$\delta M_1/\delta\phi_1 = \frac{\alpha\tau T}{(\alpha+\tau\phi_1)^2} \qquad > 0$$

$$\delta M_2/\delta\tau = \frac{P_c\phi_2 P_2\beta}{[P_2(\beta+\tau\phi_2)]^2} \qquad > 0$$

$$\delta M_2/\delta\phi_2 = \frac{P_c\tau P_2\beta}{[P_2(\beta+\tau\phi_2)]^2} \qquad > 0$$

$$\delta M_2/\delta\beta = \frac{-P_2P_c\tau\phi_2}{[P_2(\beta+\tau\phi_2)]^2} \qquad < 0$$

$$\delta M_2/\delta P_c = \frac{\tau\phi_2[P_2(\beta+\tau\phi_2)]}{[P_2(\beta+\tau\phi_2)]^2} \qquad > 0$$

$$\delta M_2/\delta P_2 = \frac{-P_c\tau\phi_2(\beta+\tau\phi_2)}{[P_2(\beta+\tau\phi_2)]^2} \qquad < 0$$

$$\delta C/\delta\alpha = \delta M_1/\delta\alpha + \delta M_2/\delta\alpha \qquad < 0$$

$$\delta C/\delta\beta = \delta M_1/\delta\beta + \delta M_2/\delta\beta \qquad < 0$$

$$\delta C/\delta\tau = \delta M_1/\delta\tau + \delta M_2/\delta\tau \qquad > 0$$

$$\delta C/\delta\phi_1 = \delta M_1/\delta\phi_1 + \delta M_2/\delta\phi_1 \qquad > 0$$

$$\delta C/\delta\phi_2 = \delta M_1/\delta\phi_2 + \delta M_2/\delta\phi_2 \qquad > 0$$

$$\delta C/\delta P_c = \delta M_1/\delta P_c + \delta M_2/\delta P_c \qquad > 0$$

$$\delta C/\delta P_2 = \delta M_1/\delta P_2 + \delta M_2/\delta P_2 \qquad < 0$$

The level of care provided is decreasing in α and β, the coefficients on physician income and leisure respectively. It is also decreasing in P_2, the price of a unit of M_2. Care is increasing in P_c, the revenue on a unit of care. As might be expected care is increasing in τ, ϕ_1, and ϕ_2, the measure of physician ethics, and the coefficients on M_1 and M_2 in the medical care production function respectively.

This last case is interesting because, like the standard optimizing salaried case and the other profit sharing cases, care is again produced efficiently. Efficient input usage arises because the physician bears the cost of supplying M_2 to the patient. All other things, equal the greater the degree of ethical feeling on the part of the physician the more care will be provided. If the physician was in

some sense "perfectly" ethical then we might regard the level of care provided in this case as being optimal since the physician has no incentive to overprovide care, and he would provide this level of care efficiently. However, this seems unlikely. The physician faces a tradeoff between the provision of care and the other arguments in his utility function. Care is increasing in P_c and it is possible to imagine fixing P_c, the revenue on a unit of care, high enough that the physician would supply the "appropriate" level of care.

Comparisons Between the Cases

The levels of M_1, M_2 and C provided are reproduced for all cases in table I. In the discussion that follows the level of M_1 provided in each case is first ranked and then these rankings are combined with information about M_2 to rank the amount of care provided in each case. It will be seen that the amount of M_1 provided can be ranked unambiguously in most cases but that ranking the level of care provided is rather more complicated.

Determination of the Level of M_1

Salaried and Profit Sharing Physician

The levels of M_1, M_2 and C are the same for these two cases. This result arises because in both cases the physician has no incentive to provide more than $\bar{C}$ level of care. Having selected this level of care the salaried physician will set M_2 usage to $\bar{M}_2$ because this will minimize the amount of M_1 he has to provide. $\bar{M}_2$, by

assumption, is the efficient level of M_2 if $\bar{C}$ is produced--so the profit sharing physician who is a residual claimant will set $\bar{M}_2$ usage equal to $\bar{M}_2$, which implies provision of the same level of M_1.

Table 1--Summarized Values for M_1, M_2 and C

Case	M_1	M_2	C
S	$(\bar{C}/\bar{M}_2^{\phi 2})^{1/\phi 1}$	$\bar{M}_2$	$\bar{C}$
SE	$\tau\phi_1 T/\alpha+\tau\phi_1$	$\bar{M}_2$	$\left[\frac{\tau\phi_1 T}{\alpha+\tau\phi_1}\right]^{\phi 1}\bar{M}_2^{\phi 2}$
W	$\beta T/\alpha+\beta$	$\bar{M}_2^{*}$	$\left[\frac{\beta T}{\alpha+\beta}\right]^{\phi 1}\bar{M}_2^{\phi 2}$
WE	$\frac{(\beta+\tau\phi_1)T}{\alpha+\beta+\tau\phi_1}$	$\bar{M}_2$	$\left[\frac{\beta+\tau\phi_1 T}{\alpha+\beta+\tau\phi_1}\right]^{\phi 1}\bar{M}_2^{\phi 2}$
P	$(\bar{C}/\bar{M}_2^{\phi 2})^{1/\phi 1}$	$\bar{M}_2$	$\bar{C}$
PE	$\tau\phi_1 T/\alpha+\tau\phi_1$	$\frac{P_c\tau\phi_2}{P_2(\beta+\tau\phi_2)}$	$\left[\frac{\tau\phi_1 T}{\alpha+\tau\phi_1}\right]^{\phi 1}\left[\frac{P_c\tau\phi_2}{P_2\beta+\tau\phi_2}\right]^{\phi 2}$

*M_2 is here assumed equal to $\bar{M}_2$. For further discussion please see text.

Salaried, Profit Sharing and Salaried Ethical Physician

Since $M_1^S = M_1^P$ the discussion will refer to M_1^S alone with the proviso that all comments apply equally to M_1^P. It cannot be the case that $M_1^{SE} < M_1^S$ since this would mean that an ethical physician would wish to supply less care than a standard optimizing physician working

under the same incentive structure. If $M_1^{SE} = M_1^S$ then this would mean that $\bar{C}$ was the level of care that the ethical physician would wish to supply. There is no particular reason why C would be this level, unless $\bar{C}$ represented a market determined level of care. However, M_1^{SE} is increasing in τ and ϕ_1. Therefore there exists some τ and ϕ_1 such that $M_1^{SE} > M_1^S$.

This result is interesting because it shows that in the case of a standard optimizing physician in an HMO setting the possibility exists that the level of care provided will be inadequate if $\bar{C}$ is a malpractice constraint on the level of care (remembering that $M_2 = \bar{M}_2$ in both cases). This can be contrasted to the fee-for-service case where unethical behavior on the part of the physician will tend to lead to the provision of excessive amounts of care.

The salaried and salaried ethical case are illustrated in figure I where C^* represents some unspecified "appropriate" amount of care. The hospital constrains M_2 use to $\bar{M}_2$ and the standard optimizing physician will work $\bar{M}_1$ hours in order to supply the patient with $\bar{C}$ units of care. An ethical physician who wishes to provide the patient with C^* units of care may do so by "overproviding" his own time, working M_1' instead of $\bar{M}_1$. (For example, the physician may arrive earlier or work into breaks.) In this sense $M_1'-\bar{M}_1$ might be thought of as the charitable provision of care by an HMO physician.

Note that the hospital will select $\bar{M}_2$ and $\bar{M}_1$ to reflect the relative unit cost of each of these inputs (w_2 and w_1, respectively, in figure I)—so $\bar{C}$ would be produced efficiently. If C^* is selected by the physician it will be produced using an inefficiently large amount

of M_1. Cost minimization would imply producing C^* using M_1^* and M_2^* levels of the two inputs.

Wage Earning and Wage Earning Ethical Physician

In making the comparison between a wage earning and wage earning ethical physician note that if $\tau = 0$ then the amount of M_1 supplied will be the same in each case. Since $\tau > 0$ it is necessarily true that $M1^{WE} > M_1^{W}$. Furthermore, the larger is τ (the greater the degree of ethical feeling that the physician displays) the larger will M_1^{WE} be in relation to M_1^{W}. Assuming that $\bar{M}_2$ is selected in both cases this means that the wage earning ethical physician will supply more care.

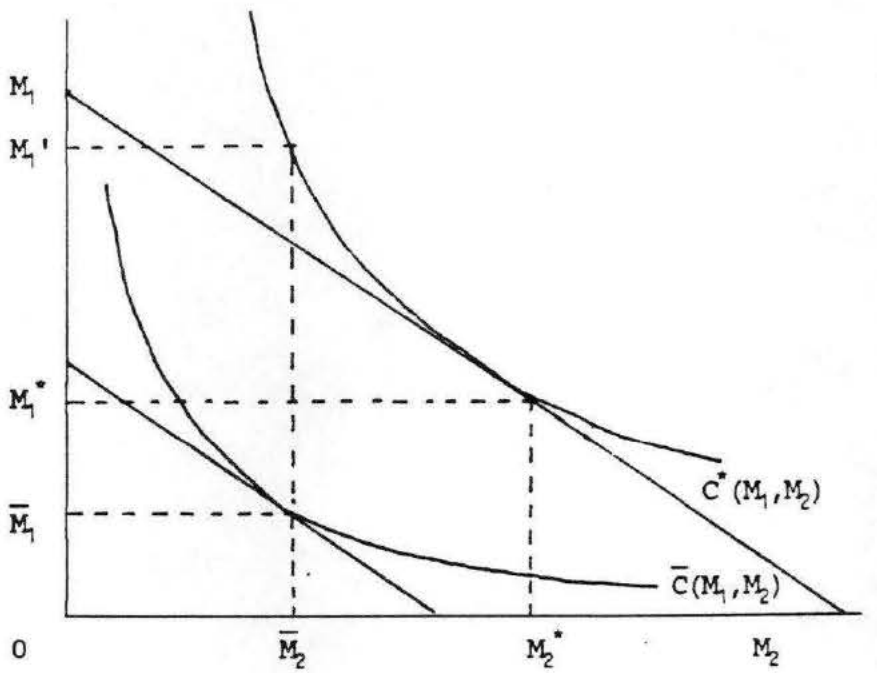

Figure 1. Determination of Input Usage when $\bar{C}$ and C^* are Selected

Wage Earning Ethical and Salaried Ethical Physician

If $\beta = 0$ this would imply that the wage earning physician attaches no utility to additional income and in this case M_1^{WE} would equal M_1^{SE}. If $\beta > 0$ then $M_1^{WE} > M_1^{SE}$. It is not clear that one or other of these levels of M_1 is associated with the "appropriate" amount of care ($\bar{M}_2$ is selected in both cases). However, note that by paying a physician an hourly wage some of the adverse incentives associated with the fee-for-service case may reemerge. The physician is ethical in both cases but he will supply more care when he is a wage earner as opposed to a salaried employee. This means that the assumption of ethical behavior is not in itself sufficient to ensure that enrollees in an HMO receive the "appropriate" amount of care.

Salaried Ethical and Profit Sharing Ethical Physician

It can be seen from table 1 that $M_1^{SE} = M_1^{PE}$. Intuitively, if the physician is equally ethical (τ is the same in both cases) each physician is concerned only with the cost of M_1 in terms of leisure (α) and the productivity of M_1 in the medical care production function (ϕ_1) in making the decision about how much M_1 to provide.

Salaried Ethical and Wage Earning Physician

The relative size of M_1^{SE} and M_1^{W} depends on the relative size of $\tau\phi_1$ and β. It is not clear cut.

$$M_1^{SE} \gtrless M_1^{W} \text{ as } \tau\phi_1 \gtrless \beta$$

The relationship between $\tau\phi_1$ and β, for a given value of ϕ_1, depends on the relative size of τ and β. The possibilities are as follows:

1. If $\tau = \beta$ then $\tau\phi_1 < \beta \rightarrow M_1^{SE} < M_1^{W}$

2. If $\tau < \beta$ then $\tau\phi_1 < \beta \rightarrow M_1^{SE} < M_1^{W}$

3. If $\tau > \beta$ then $\tau\phi_1 \gtrless \beta \rightarrow M_1^{SE} \gtrless M_1^{W}$

The third case, that $\tau > \beta$, may seem unrealistic since it implies that the physician values patient utility more than income. However, it obviously cannot be ruled out. For most values of ϕ_1 that can be postulated τ would have to be a fairly large multiple of β for it to be true that $M_1^{SE} > M_1^{W}$. This can be seen from table 2. The left hand column give values of ϕ_1 in small increments over the relevant range. The right hand column gives the size of τ necessary, in terms of a multiple of β, if M_1^{SE} is to be larger than M_1^{W}. For example, if $\phi_1 = 0.3$ then unless τ is 3 1/3 times the size of β it will be true that $\tau\phi_1 < \beta$ which means that $M_1^{SE} < M_1^{W}$.

Simulated values of M_1 are provided in appendix C. In comparing salaried and wage earning physicians in the empirical section I implicitly assume that the two groups have the same level of ethics. Therefore, for all $\beta > 0$ the wage earner supplies more care than the salaried physician. However, it is further assumed that the labor supply function is first increasing and then decreasing in β (i.e., the labor supply curve is backward bending) which means that this result may not hold at "high" wages.

Table 2--Critical Values of ß for Determining the Relationship Between M_1^{SE} and M_1^{W}

ϕ_1	$M_1^{SE} < M_1^{W}$ if $\tau <$
.1	10.0ß
.2	5.0ß
.3	3.3ß
.4	2.5ß
.5	2.0ß
.6	1.7ß
.7	1.4ß
.8	1.3ß
.9	1.1ß

The Determination of the Level of Care

The level of M_1 provided will partially determine the level of care provided. The profit sharing and salaried physicians both supply $\bar{C}$, the malpractice constrained level of care and they combine M_1 and M_2 efficiently to provide that level of care.

The salaried ethical and the profit sharing ethical physician both supply the same amount of M_1 but they use different amounts of M_2 and therefore the level of care provided will not be the same. The amount of M_1 selected by the salaried ethical physician is, as discussed above, an inefficiently large level of M_1. The salaried ethical physician supplies an inefficiently large level of M_1 because he wishes to provide more than the malpractice constrained level of care and is prevented by the hospital from ordering more than $\bar{M}_2$ of nonphysician inputs. The salaried ethical physician will use $\bar{M}_2$ and the profit sharing ethical physician will use the efficient level of

M_2 for any particular level of M_1 that he selects (this level of M_2 is derived from his optimization problem). If $\bar{M}_2$ is selected by an employing hospital to be the efficient input of M_2 to produce $\bar{C}$ then it must be less than the amount of M_2 that an efficient profit sharing ethical physician will select to provide a level of care greater than $\bar{C}$.

In general the amount of care provided by a wage earning ethical physician will be greater than that provided by a standard optimizing wage earning physician. This must be true since $M_1^{WE} > M_1^{W}$ and the ethical physician will supply $\bar{M}_2$. The standard optimizing physician cannot provide any more than $\bar{M}_2$ amount of nonphysician inputs so it must be true that $C^{WE} > C^{W}$.

Ranking the amount of care provided becomes problematic when considering the relationship between C^{PE} and C^{W}. We know that $M_1^{PE} < M_1^{W} < M_1^{WE}$. It is also true that $M_2^{W} \leq M_2^{WE} = \bar{M}_2$ and that $M_2^{PE} > \bar{M}_2$. It is impossible to know whether the wage earning physician driven by the desire to earn more income will provide more care than a profit sharing ethical physician. For the wage earning case the possibility of excess care <u>does</u> exist if the physician combines a very excessive amount of M_1 with the constrained level of M_2, and in addition, care is provided inefficiently.

In ranking the level of care provided the following is certainly true:

$$C^{P} = C^{S} < C^{SE} < C^{PE},\ C^{W},\ C^{WE}$$

We also know that $C^{W} < C^{WE}$.

Conclusion

The rankings of the level of care provided and comparisons of the input mix used are a function of the parameter values. For example the relationship between M_1^{SE} and M_1^{W} is a function of the physician's valuation of additional income compared to his concern for patient welfare (i.e., the relative size of τ and β). If τ is large relative to β, or ϕ_1 is large relative to ϕ_2, then it is possible that a salaried ethical physician would supply more care than a wage earning physician. In comparing the wage earning physician with the profit sharing ethical physician it is not theoretically clear whether the profit sharing physician's ethics or the wage earner's interest in additional income dominates.

One of the most important implications of these rankings is that ethical physicians will supply differing amounts of care depending on how they are paid. This means that the assumption of ethics alone is not sufficient to ensure that patients receive appropriate care within HMOs.

A further implication is that if a physician is assumed to be a standard optimizer either on a fixed salary or a residual claimant in an HMO he will tend to supply the level of care $\overline{C}$, which, given the information asymmetry discussed earlier, constitutes an inadequate level of care. This can be contrasted with the fee-for-service case where a lack of ethics tends to lead to the overprovision of care. Wage earning physicians will certainly supply more care than salaried physicians. However, it is impossible within the context of this

model to say whether which level of care is dominant in any welfare sense. This inability to make welfare comparisons derives from the fact that a unique "appropriate" level of care cannot be specified. We do know for certain that a wage earning physician probably use an inefficient combination on inputs ("oversupplying" his own time). On the other hand the level of care $\bar{C}$, while inadequate, will be supplied efficiently by both the salaried and the profit sharing physicians.

In the empirical section ß is not a constant, rather it declines as hours worked increased. As mentioned before, this assumption allows the possibility that at "high" wages a wage earner could supply less labor than a salaried physician, assuming equal degrees of ethical behavior.

CHAPTER IV

DIFFERENCES IN PERSONAL CHARACTERISTICS AND LABOR MARKET EXPERIENCES OF EMPLOYED VERSUS SELF-EMPLOYED PHYSICIANS

One implication of the theoretical model developed in chapter III is that physician labor supply is sensitive to the payment scheme under which the physician is paid. The effect of physician employment status on labor supply will be pursued empirically. Salaried physicians are shown to have an incentive to minimize the input of their own time. This result is derived theoretically in the context of a prepaid group practice but it is, of course, independent of the payment method under which the physician's patients are billed. Physicians who are employees in a prepaid or fee-for-service setting can be expected to supply less labor than physicians working under incentive based reimbursement. In chapter III a salaried physician is described as a physician who receives a fixed income for a specified number of hours worked. In this chapter and in chapter V the group of interest will be referred to as "employees" where employee status is self-reported and implies employment by an HMO, a clinic, or another physician. This chapter introduces some of the empirical issues associated with physician labor supply, describes the data set to be used for the formal tests performed in chapter V, and discusses some

of the apparent differences in employed versus self-employed physicians.

Description of Data Set

Data for the analysis are drawn primarily from the 1984-1985 Physician's Practice Costs and Income Survey (PPCIS).[1] Similar surveys were conducted by the National Opinion Research Center (NORC) in previous years but the 1984-85 survey is by far the most comprehensive. A total of 4729 responses from office-based and hospital-based physicians in all specialties were collected. Physicians were asked to respond to questions about practice characteristics, practice methods, fees, output mix and employment of other personnel. They were also asked a number of biographical questions. The survey excludes physicians who worked less than twenty hours a week, who were engaged primarily in teaching and research, or who were in residency. Therefore, the focus is on physicians who deliver patient care on a full time basis. The exclusion of physicians working fewer than twenty hours a week might be problematic if one were interested in female physicians with young children or physicians nearing retirement. However, these groups are not specifically a focus of my analysis and so this is not likely to be a great cause for concern.

In addition to the primary data set the following secondary sources are used:

1. 1984 Physician Master File of the American Medical Association (AMA) for data on gender, year of licensure and board certification.

2. May 1984 Area Resource File for county level data on physician supply, per capita income and vital statistics.

3. 1982 American Hospital Association (AHA) Annual Survey of Hospitals for information about the hospital to which the physician is principally affiliated.

4. The 1981-1982 Center for Health Economics Research (CHER) Hospital Data Tape for information about the hospital to which the physician is principally affiliated.

5. 1984 Health Care Financing Administration (HCFA)'s Interns and Residents tape for information about the number of interns and residents in a particular hospital.[2]

To reduce the problems anticipated from pooling physicians with widely different practice styles only five specialties are examined, namely family and general practice, internal medicine, obstetrics and gynecology, and pediatrics. The data set used for this particular study contains responses from 1693 physicians in these specialties which constitutes 36% of the entire sample. Due to missing values, the number of observations used in the regression analysis was 1247, or 26% of the entire sample and 74% of the specialties listed above.

Table 3 contains descriptions of all variables used in the analysis. Variables used in this study fall into four broad categories: (1) personal characteristics such as age, gender, race, board certification and specialty; (2) practice characteristics such as memberships in AHPs and whether or not the physician is an employee; (3) environmental influences such as location, the number of physicians per capita, county per capita income and the size of the elderly population and (4) "economic" variables such as hours of work, wages and the average time taken to complete a visit. For most of the variables the brief description in table 3 should suffice.

Table 3--Variable Descriptions

Variable Name	Variable Description
AGE	Age of physician at survey date
EXP	Years of experience (age minus year at licensure)
SINGLE	1 = no spouse present, 0 = spouse present
NWHITE	1 = nonwhite, 0 = white
FEMALE	1 = female, 0 = male
OTHERINC	1 = physician has nonmedical income (including spousal income) in excess of $10,000), 0 = otherwise
BOARD	1 = physician is board certified, 0 = otherwise
HMO50	1 = member of at least one HMO and receives more than 50% of income from prepaid programs, 0 = otherwise
HMO^3	1 = member of at least one HMO, 0 = otherwise
IPA	1 = member of at least one IPA, 0 = otherwise
PPO	1 = member of at least one PPO, 0 = otherwise
EMPLOYEE	1 = physician is an employee of a hospital, a clinic, an HMO, another physician, or a corporation, 0 = otherwise
GRPSZ	Number of physicians associated with physician's principle practice (including self)
MULTSPEC	1 = member of a group containing physicians in more than one specialty, 0 = otherwise
FAMILY	1 = family practitioner, 0 = other specialty
GENERAL	1 = general practitioner, 0 = other specialty
INTERNAL	1 = internal medicine, 0 = other specialty
OBGYN	1 = obstetrics-gynecology, 0 = other specialty
PED	1 = pediatrician, 0 = other specialty

Table 3--Continued

Variable Name	Variable Description
TOWN	1 = resident of Standard Metropolitan Statistical Area (SMSA), 0 = otherwise
NORTH	1 = resident of the Northeast (Connecticut, Delaware, District of Columbia, Maine, Massachusetts, New Hampshire, New Jersey, New York, Pennsylvania, Rhode Island, Vermont), 0 = otherwise
CENTRAL	1 = resident of the Midwest (Illinois, Iowa, Indiana, Kansas, Michigan, Minnesota, Missouri, Nebraska, North Dakota, Ohio, South Dakota, Wisconsin), 0 = otherwise
SOUTH	1 = resident of the South (Alabama, Arkansas, Florida, Georgia, Kentucky, Louisiana, Maryland, Mississippi, North Carolina, Oklahoma, South Carolina, Tennessee, Texas, Virginia, West Virginia), 0 = otherwise
WEST	1 = resident of the West (Alaska, Arizona, California, Colorado, Hawaii, Idaho, Montana, Nevada, New Mexico, Oregon, Utah, Washington, Wyoming), 0 = otherwise
VISITS	Number of visits completed during the reference week (total of office visits, emergency room and out-patient consultations, operations and assists participated in, in-patient visits on hospital rounds, visits to nursing homes, convalescent homes and extended care facilities)
HOURS	Number of hours spent in medical activities and administrative activities during the reference week
PROD	HOURS/VISITS (average time spent per visit)
WEEK	Number of weeks worked in 1983
YHOURS	HOURS X WEEK (yearly hours)
PEROFF	Percentage of total hours spent conducting office visits

Table 3--Continued

Variable Name	Variable Description
PERADMIN	Percentage of total hours spent in administrative activities
INCOME	Net income in 1983 from all medical practices after practice deductions, but before taxes
WAGE	Hourly wage (INCOME/YHOURS)
MDPC	Number of patient care physicians per 100,000 of the population in the county in which the physician practices
PCINC	Per capita income of the county or SMSA in which the physician practices
ELDERLY	Number of people aged 65 or older per 100,000 of the population in the county in which the physician practices

HMO50 is designed to capture physicians who are committed to an HMO in the sense of receiving a significant portion of their income from an HMO (50% or more). Since the survey does not distinguish between income received from HMOs, IPAs, and PPOs the variable HMO50 is coded 1 if the physician does not belong to any IPAs or PPOs. This has the effect of excluding some physicians who are committed to an HMO but it is unavoidable given the way in which the survey is set up. The variables IPA and PPO were not utilized in the regression analysis because, given the great diversity among IPAs and PPOs, there was no a priori expectation about their influence.

It is possible that physicians in multispecialty groups will generate business for each other and so work longer hours--hence the

inclusion of MULTSPEC in the regression analysis. Physicians who prefer to work longer hours may select multispecialty groups for this reason.

TOWN is designed to capture any urban-rural disparities that may arise, such as competitive pressure and patient density. The regional dummies may capture price differentials, differences in physician preferences for leisure, institutional and regulatory differences (although the division of the nation into four sectors is probably too crude to capture this effect) and possibly differences in the population the physician serves not picked up by the variables ELDERLY and PCINC.

PEROFF and PERADMIN may proxy differences in output mix across physicians. My a priori expectation is that high values of PEROFF and PERADMIN indicate a relatively "simple" output mix since the other activities included in the measurement of HOURS (surgical assists, in-patient visits, etc.) seem to indicate a more complex case mix.

Income was reported exactly for only a small number of observations. Most physicians were asked to report their income in ranges of $10,000 per year at lower incomes and $20,000 per year at higher incomes. For the purposes of estimation the measure of income used was the median income in each range. This is an obvious limitation of this data set. Furthermore, it would be better to have an independent measure of WAGE but since this is seldom the case most researchers resort, as I do, to imputing the wage.

The Descriptive Statistics

Table 4 details the employment status of physicians surveyed. 29% of all physicians report being an employee. Specialty is an important determinant of employee status, 34% of pediatricians but only 15% of general practitioners report employee status. Participation in AHPs is also sensitive to specialty with general practitioners having rather low involvement and pediatricians and obstetrician-gynecologists more likely to participate in AHPs.

Table 4--Employment Status of Physicians

	Specialty					
Status	All # (%)	Family # (%)	General # (%)	Internal # (%)	Obgyn # (%)	Ped # (%)
EMPLOYEE	1357 (29.0)	97 (20.8)*	35 (15.0)*	107 (22.7)*	46 (19.0)*	94 (33.6)*
HMO	738 (16.01)	72 (15.5)	15 (6.4)*	77 (16.3)	46 (19.0)	59 (21.1)*
HMO50	132 (2.8)	19 (4.1)	2 (0.9)	19 (4.0)	14 (5.8)*	20 (7.1)*
IPA	543 (11.0)	46 (9.9)	24 (10.3)	57 (12.1)	34 (14.0)	31 (11.1)
PPO	518 (11.0)	46 (9.9)	10 (4.3)*	34 (7.2)*	36 (14.9)*	27 (9.6)
Total	4729 (100.0)	466 (100.0)	233 (100.0)	472 (100.0)	242 (100.0)	280 (100.0)

*Indicates that the mean for a specialty is significantly different from the mean for all other physicians at the 0.05 level or better.

Therefore, in order to isolate the independent effect of employment status on labor supply we must control for specialty since this influences employment status in addition to labor supply.

Table 5 reports means of variables as well as differences in means of variables between employed and self-employed physicians. Table 6 shows differences in means of variables between employed and self-employed physicians by specialty. For the sake of brevity the means of variables by specialty are not reported. A negative sign in the difference columns in tables 5 and 6 indicates that the mean value for employed physicians is less than that for self-employed physicians. The PROC TTEST program in SAS was used to test for significant differences in the means. In table 6 fewer differences in means are found to be significant due to the smaller number of observations in each category.

Employed physicians are significantly younger and have less experience than their self-employed colleagues. They are more likely to be single (perhaps because they are younger) and female. These differences in personal characteristics suggest that employee status is chosen over traditional practice arrangements by physicians with a stronger preference for nonmarket time. There is no discernible difference in the tendency for physicians to be board certified at the aggregate level. However, employees specializing in family practice and internal medicine are more likely to be board certified than their self-employed colleagues. This suggests, at least at a superficial level, that the employed cannot be distinguished from the self-employed in terms of quality.

Table 5--Descriptive Statistics: Mean Values of Variables

Variable Name	All Physicians Mean Value (SD)	Employed Physicians Mean Value (SD)	Self-employed Physicians Mean Value (SD)	Difference[a]
AGE	46.68	43.63	47.89	-4.26*
	(11.14)	(10.53)	(11.15)	
EXP	20.25	17.32	21.43	-4.11*
	(11.11)	(10.53)	(11.13)	
SINGLE	0.12	0.16	0.10	0.06*
	(0.32)	(0.37)	(0.31)	
NWHITE	0.17	0.18	0.16	0.01
	(0.37)	(0.39)	(0.37)	
FEMALE	0.10	0.18	0.06	0.12*
	(0.29)	(0.38)	(0.24)	
OTHERINC	0.35	0.36	0.34	0.02
	(0.48)	(0.48)	(0.48)	
BOARD	0.68	0.70	0.69	0.02
	(0.46)	(0.46)	(0.46)	
HMO	0.16	0.20	0.14	0.12*
	(0.36)	(0.40)	(0.34)	
HMO50	0.03	0.09	0.00	0.09*
	(0.16)	(0.28)	(0.06)	
IPA	0.11	0.07	0.13	-0.16*
	(0.32)	(0.26)	(0.34)	
PPO	0.11	0.06	0.13	-0.18*
	(0.31)	(0.24)	(0.34)	
GRPSZ	9.16	30.93	3.82	27.11*
	(36.87)	(75.85)	(11.74)	
MULTSPEC	0.12	0.23	0.08	0.13*
	(0.33)	(0.42)	(0.27)	
FAMILY	0.10	0.07	0.11	-0.04*
	(0.30)	(0.26)	(0.31)	
GENERAL	0.05	0.03	0.06	-0.08*
	(0.22)	(0.16)	(0.24)	
INTERNAL	0.10	0.09	0.11	-0.03*
	(0.30)	(0.27)	(0.31)	
OBGYN	0.05	0.03	0.06	-0.05*
	(0.22)	(0.18)	(0.23)	
PED	0.06	0.07	0.06	0.02
	(0.24)	(0.25)	(0.23)	
TOWN	0.82	0.86	0.80	0.06*
	(0.39)	(0.35)	(0.40)	
NORTH	0.24	0.28	0.23	0.03*
	(0.43)	(0.45)	(0.42)	

Table 5--Continued

Variable Name	All Physicians Mean Value (SD)	Employed Physicians Mean Value (SD)	Self-employed Physicians Mean Value (SD)	Difference[a]
CENTRAL	0.23	0.26	0.22	0.04*
	(0.42)	(0.44)	(0.41)	
SOUTH	0.31	0.26	0.33	-0.03*
	(0.46)	(0.44)	(0.47)	
WEST	0.22	0.20	0.22	-0.02
	(0.41)	(0.40)	(0.42)	
VISITS	146.48	135.15	150.26	-15.11*
	(175.62)	(167.15)	(178.22)	
HOURS	49.01	43.42	51.15	-7.72*
	(24.73)	(25.12)	(24.24)	
PROD	0.71	0.79	0.68	0.11*
	(0.91)	(1.30)	(0.73)	
WEEK	47.12	46.16	47.50	-1.34*
	(5.14)	(6.07)	(4.65)	
YHOURS	2317.69	2000.79	2439.59	-438.80*
	(1218.60)	(1206.29)	(1201.42)	
PEROFF	0.52	0.52	0.52	0.00
	(0.23)	(0.26)	(0.22)	
PERADMIN	0.29	0.38	0.25	0.12*
	(0.36)	(0.41)	(0.34)	
INCOME	100475.28	93364.98	103344.71	-9979.73*
	(67827.10)	(57493.31)	(71385.26)	
WAGE	163.36	178.32	157.56	20.76
	(473.19)	(435.53)	(486.94)	
MDPC	2299.27	2563.54	2192.92	370.63*
	(3835.05)	(3883.35)	(3810.87)	
PCINC	10030.89	10188.89	9967.31	221.58*
	(2078.70)	(2048.77)	(2087.55)	
ELDERLY	1050.88	1166.84	1004.21	162.63*
	(1677.08)	(1718.01)	(1658.29)	

[a]Negative (positive) sign indicates that the mean value of the variable for employed physicians is smaller (larger) than the mean value for self-employed physicians.

*Indicates difference in means is significant at the 0.05 level or better.

Table 6--Differences in Mean Values of Variables Between Employed and Self-Employed Physicians by Specialty

Variable Name	Difference[a] Family	General	Internal	Obgyn	Ped
AGE	-6.88*	-9.41*	-6.89*	-3.13	-7.16
EXP	-6.60*	-8.56*	-6.96*	-2.31	-6.77*
SINGLE	0.12*	0.02	0.06	0.07	0.08*
NWHITE	-0.01	0.30*	-0.07	-0.08	0.09
FEMALE	0.12*	0.11*	0.21*	0.07	0.29*
OTHERINC	0.01	-0.03	0.08	-0.13	0.10
BOARD	0.10*	-0.03	0.15*	0.05	-0.05
HMO	0.01	-0.04	0.12	0.18	-0.02
HMO50	0.17*	0.06	0.15*	0.27*	0.19*
IPA	-0.10	-0.34*	-0.17*	-0.29*	-0.29*
PPO	-0.24*	-0.30	-0.03	-0.34*	-0.27*
GRPSZ	32.00*	5.34*	24.45*	30.31*	32.47*
MULTSPEC	0.32*	0.26*	0.29*	0.28*	0.27*
TOWN	0.15*	0.11	0.06	0.13*	0.03
NORTH	-0.01	0.08	0.02	0.01	-0.01
CENTRAL	0.00	0.03	0.07	0.12	0.03
SOUTH	-0.05	-0.06	-0.09	-0.18*	-0.06
WEST	0.07	0.04	0.00	0.05	0.05
VISITS	-23.09	4.90	-50.80*	-42.95*	-10.35
HOURS	-4.24*	-0.34	-4.14*	-9.58*	-3.04
PROD	0.10	0.18	0.45*	-0.03	0.11

TABLE 6--Continued

Variable Name	Difference[a] Family	General	Internal	Obgyn	Ped
WEEK	-1.79*	-2.29*	-3.08*	-2.00*	-2.10*
YHOURS	-324.07*	-136.05	-383.07*	-578.12*	-265.46*
PEROFF	0.03	0.04	-0.02	0.02	0.01
PERADMIN	-0.01	0.01	0.01	0.00	0.00
INCOME	-5505.27	-11658.19	-11854.86	-38152.74*	-6514.52
WAGE	-5.76	-3.08	-0.15	-2.48	-1.44
MDPC	363.92	423.16	893.36*	1628.00*	1,957.47*
PCINC	636.15*	717.65	185.02	463.23	474.93
ELDERLY	148.77	103.72	374.13*	775.87*	836.01*

[a]Negative (positive) sign indicates that the mean value of the variable for employed physicians is smaller (larger) than the mean value for self-employed physicians in that particular specialty.

*Indicates that the mean value for employed physicians in that particular specialty is significantly different from the mean for self-employed physicians at the 0.05 level or better.

Turning to variables that capture practice characteristics there are some clear differences between the employed and the self-employed. The differences in means for HMO50 are all positive and most are significant indicating that employed physicians are more likely to be committed to HMOs. This suggests that innovative billing procedures are associated with less traditional employment arrangements. Differences in means for IPA and PPO are mostly

significant and negative which highlights how different the organizational structures of IPAs and PPOs are compared to HMOs. The average group size reported by all employed physicians is larger by 27 physicians (a fact that makes intuitive sense) and employed physicians are also more likely to be found in multispecialty group. At the specialty level there would appear to be no significant difference in either PEROFF or PERADMIN between the employed and the self-employed. This implies that within a specialty the output mix is fairly constant across the employed and the self-employed.

The mean difference in INCOME is significant and in favor of the self-employed who earn on average nearly $10,000 more per year than employees. The biggest difference is found among obstetrician-gynecologists with the self-employed earning about $38,000 more per year than the employed. Differences in WAGE are not significant, although employed physicians earn on average between $0.15 less per hour in internal medicine and $5.76 less per hour in family practice. This may imply a compensating differential in favor of the self-employed due to unobservable advantages to employee status.

Physician employees conduct about 15 fewer visits per week and work 7.72 fewer hours per week--a finding consistent with the notion that physician employees may select this practice style due to the opportunities afforded for increased leisure. In addition, employees work on average 1.34 fewer weeks per year. The difference in means for the variable PROD is significant and positive (employed physicians spend longer on each visit than the self-employed). This finding would also appear to be borne out at the specialty level except in the

case of obstetrician-gynecologists. Longer visits may result because the physician does not gain financially from seeing additional patients.

Differences in means for the geographical variables are significant. As might be expected employed physicians are more common in urban areas and in the North and Central parts of the country. The South sees significantly fewer employed physicians: Several authors note that the South, for whatever reason, seems to have maintained a more traditional institutional structure (see, for example, Hurdle and Pope 1989).

The difference in mean MDPC is positive and significant. Employed physicians are found more frequently in areas of greater physician density. This would be consistent with the hypothesis that physicians resort to taking salaried positions when demand conditions are less favorable. However, since PCINC and ELDERLY are also significant and positive (proxies for the demand for physician services) greater physician density may be the result of higher demand for physician services and therefore not necessarily imply that employees are concentrated in markets characterized by physician surplus.

Examination of the differences between employed and self-employed physicians raises some interesting questions. It would seem clear that employed physicians work fewer hours and receive lower wages. However, examination of means does not enable one to decide if that is a result of employment status per se or the result of other factors also shown to be associated with employment status, such as

location, age, and gender. The next chapter seeks to determine the independent impact of employee status on the hours worked and remuneration of physicians.

Notes

[1]The PPCIS is conducted by the National Opinion Research Center (NORC) for Health Care Financing Administration (HCFA).

[2]The additional data sets were merged by the Center for Health Economic Research (CHER). The purpose of the merger of secondary data into the PPCIS was to incorporate "data reflecting the characteristics of the community in which the physician practices and the primary hospital with which the physician is affiliated" (Rosenbach 1985, 1).

[3]The majority of physicians reporting membership in an HMO belong to 1 HMO (78%). For physicians belonging to more than 1 HMO the maximum number of HMO affiliations reported was 9. A similar pattern holds true for IPA and PPO membership.

CHAPTER V

ESTIMATION OF A LABOR SUPPLY FUNCTION FOR PRIMARY CARE PHYSICIANS

In this chapter I estimate simultaneous labor demand and supply functions to assess the effect of employee status on the labor supply and hourly returns of physicians. Another important issues is addressed in this chapter: is the increase in the number of physician employees motivated primarily by institutions or by physicians? Medical institutions are concerned about the rising price of medical care--and the move to salaried physicians may help cut costs. On the other hand, physicians are increasingly heterogeneous, and some may prefer nontraditional contractual arrangements. In addition, the threat of malpractice suits and escalating insurance costs may make practice modes that reduce risk to the individual physician more appealing.

Evidence suggests that we are currently experiencing a tendency towards an excess supply of physicians.[4] Excess supply may permit institutions to introduce cost saving contractual arrangements that would otherwise be rejected by physicians.[5] If the move towards a greater number of physician-employees is motivated on the "demand side" (by the institutions' interest in cutting costs) then one might anticipate that among physician-employees would be found the least favored segment of the physician population. This suggestion is

refuted by the finding of equal rates of board certification among the employed and self-employed, more evidence on this point might be provided by examining the effect of employee status on labor supply and wages. If physicians become employees because they lack other opportunities there is no reason to expect physician-employees to work reduced hours, but they might have lower wages. If the growth in physician-employees is motivated mainly on the "supply side" (by physicians' preferences for these arrangements) we would not necessarily expect employee status to be associated with lower wages but it might reduce work effort.[6]

The casual evidence from chapter IV is that employee status may be associated with lower wages for the specialties considered here, although differences in means of wages are not significant. Being an employee seems to be associated with reduced work effort. However, multiple regression analysis is necessary to isolate the impact of employee status on wages and work effort. Before proceeding to the empirical estimation a brief summary of some other work in this area is presented.

Review of the Literature

There is a fairly substantial body of literature that examines the determination of wages and labor supply for health care professionals. Vahovich (1977) estimates labor supply functions for physicians to find out if the typical physician is on the backward bending portion of his labor supply curve. Up until the 1980s a physician shortage was threatened, if the labor supply function is

negatively sloped at current wages then efforts to induce more work effort from existing physicians by increasing their hourly "wage" would be futile. Vahovich estimates labor supply functions for general practice, internal medicine and surgery using two-stage least squares (2SLS) and finds critical wages (beyond which the supply curve is negatively sloped) that are "less than one standard deviation above the mean wage for the respective specialties, implying that for a substantial number of physicians in these specialties, the income effect dominates the substitution effect" (Vahovich, 58).

Boulier (1979) estimates an hours equation for self-employed dentists as part of a system of equations where the focus is the supply of dental care to patients. Equations in the model are estimated individually using Ordinary Least Squares (OLS). In place of a wage and wage squared term in the hours equation Boulier uses the cost of an extraction and its square on the grounds that "the relation between hours worked and net income per hour is complex and not analogous to that between hours worked and wage rates for employed persons" (Boulier, 892). If the cost of an extraction is a suitable instrument for the wage, his use of OLS can be justified as an instrumental variables (IV) procedure.[7] Boulier finds evidence of a backward bending supply curve of labor for dentists.

Scheffler and Rossiter (1983) examine the effect of different incentive schemes on the labor supply decisions of dentists. Dummy variables are included in the labor supply equation for the following incentive schemes: (1) net income only, (2) fixed salary only, (3) salary plus unequal share of net income and (4) unequal share of net

income; the reference group being solo practitioners. They report that "dentists on fixed salary worked more hours per week (approximately 6 more hours per week) and worked more weeks per year (approximately 2 more weeks per year) than solo dentists who receive net practice income only, ceteris paribus" (Schleffer and Rossiter, 37). These results are rather surprising but the authors do not account for them.

Kehrer (1976) estimates weekly hours worked and wage equations for male and female physicians separately using AMA data from the 1973 Periodic Survey of Physicians. The method used is 2SLS although she actually reports OLS estimates of the wage equation. One possible problem with her specification is that hours and its square are not treated as endogenous in the wage equation and only the wage but not its square appears in the hours equation. The smaller number of observations on females (288 versus 2,962 for males) results in fewer significant independent variables in the female equations. Kehrer finds that female physicians show a negative supply response to increases in the hourly wage, whereas for males the effect is positive.

One of the included explanatory variables is EMPLOYEE defined as "a binary variable set equal to one if 50 percent or more of a physician's 1972 net income was in the form of salary" (Kehrer 1976, 543). This variable is not comparable to the variable EMPLOYEE used in my study which is coded 1 if the physician reports employee status. The coefficient on EMPLOYEE is negative and insignificant in Kehrer's OLS estimate of the wage equation for women, but it is negative and

highly significant in the wage equation for men. She concludes "employee status is associated with a 12 percent negative differential in hourly net income for men physicians, but not with any significant loss for women, ceteris paribus" (Kehrer, 535). In her 2SLS estimates of weekly hours worked equations the coefficient of EMPLOYEE is negative and significant in the female equation whereas it is positive and completely insignificant in the male equation:

> Employee status is associated with significantly fewer--11 percent--hours worked per week in the female equation, but not in the male equation. Thus, employee status seems to operate on annual net income indirectly for women, through its association with fewer hours worked, but directly for men, through its association with lower net income per hour. (Kehrer, 541)

Kehrer reports a finding by Phelps (1968) that female physician-employees work fewer hours than other female physicians and that the likelihood of employee status for female physicians is increased by the presence of young children. This suggests that employee status is chosen by female physicians with a high demand for nonmarket time. For employed men, however, lower wages than self-employed men coupled with approximately the same hours as the self-employed is consistent with the hypothesis that male physicians will "resort" to salaried positions if their labor market opportunities are poorer.

In a 1984 article Mitchell examines the reasons why female physicians work fewer hours than their male counterparts. She reestimates Kehrer's weekly hours equations using NORC data for 1978-1979 and includes wage and its square as explanatory variables. Mitchell's results suggest that male physicians are on the negatively

sloped portion of their labor supply curve whereas female physicians appear to be on the vertical portion of their labor supply curve.

Mitchell includes an explanatory variable, EMPLOYEE, described as a salaried physician. Her definition of EMPLOYEE is more restrictive than that used in my study, mean values are 0.02 and 0.04 for men and women respectively, while the mean value of EMPLOYEE in my study is 0.29. In the Mitchell study male physician-employees appear to work almost four hours less per week and more than a week less per year than their self-employed counterparts. On the other hand, salaried female physicians appear to work the same number of hours and weeks as self-employed female physicians. This is in direct contrast to Kehrer's findings that male employees work less than self-employed male physicians and that employed females work less hours than self-employed females. Mitchell finds that being married--but not the presence of children--is the most significant deterrent to female physicians working long hours. Married female physicians frequently are married to physicians[8] and joint labor supply decisions within wealthy families appear to encourage female physicians to deemphasize work despite their high earnings potential.

Hurdle and Pope (1989) employ the data set used in this study, as well as earlier NORC data sets to examine trends in physician productivity and the determinants of physician productivity. Their model of physician productivity is specified as follows (Hurdle and Pope, 102):

$$V/Y = (V/H)*(H/Y)$$

Where: V/Y = annual productivity (patient visits per year)

V/H = hourly productivity

H/Y = work effort

Data limitations prevent them from estimating a production function for employed physicians and so the production function is estimated for the self-employed alone. However, they do include observations on employed physicians when estimating the work effort function. This gives an accurate picture of physician productivity only if employed physicians are identical to the self-employed in their usage of nonphysician inputs. My theoretical work showed that employed and self-employed physicians probably differ in their usage of nonphysician inputs and therefore this approach may be unwise.

Hurdle and Pope estimate reduced form work effort equations (not labor supply functions) with (1) annual patient care office hours and (2) annual patient care total hours as dependent variables. These two measures specifically exclude administrative hours which are included in my measure of the dependent variable. However, since Hurdle and Pope focus on productivity, which they define as visits produced per year, their definition of the dependent variable is appropriate here. The authors include an independent variable, EMPLOYEE, which is not described but which has a mean of 0.17. The coefficient on EMPLOYEE is negative and significant in the work effort equation.

The evidence presented above is mixed. Employee status may have a negative effect on labor supply: however, Kehrer disputes that for male physicians, Mitchell disputes that for female physicians, and

Schleffer and Rossiter find that employed dentists work harder than dentists who are residual claimants. The effect of employee status on the wage is ambiguous. It is difficult to make direct comparisons between the articles cited above and my work due to differences in specification and focus, and differences in the definition of EMPLOYEE.

Estimation of Labor Demand and Supply functions

Expressions for M_1 (labor supply) under several different incentive schemes were derived in chapter III from a simple Cobb-Douglas utility function maximized subject to time and income constraints. The estimated labor supply function differs from the expressions for M_1 from chapter III in several ways. At an empirical level no satisfactory proxy for "ethical feelings" could be found. Therefore utility is redefined as a function of leisure time and consumption of market goods only.[9] The Cobb-Douglas functional form does not permit a backward bending supply curve of labor and, for empirical purposes, it was felt that such a specification was appropriate. Therefore, wage and its square appear in the estimating equation. The testable hypothesis from chapter III, that salaried physicians supply less labor than physicians paid under incentive based reimbursement, depends upon the utility parameter on income (β) being relatively "small". A backward bending labor supply curve implies that in a more flexible utility function an equivalent measure of β may decline as the wage increases. Thus, at very high wages employees may not supply less labor than wage earners. The estimating

equation includes dummy variables reflecting individual characteristics that might be expected to influence a physician's "taste" for leisure versus market goods. An intercept term is incorporated in the estimating equation.

Equilibrium wage and labor supply result from the interaction of demand and supply. Equilibrium in the labor market for an individual physician is defined by the following system of equations:

$$D_1 = D(w_1)$$

$$S_1 = S(w_1, U)$$

$$D_1 = S_1$$

Where D_1 and S_1 are the demand for, and supply of, individual labor, respectively. U is unearned income. Typically, this market is specified recursively, so that individuals are wage takers regardless of hours supplied: the wage offer does not depend upon hours of work. Alternatively, when the wage does depend upon hours, estimates of one equation in the system using OLS will lead to estimates that are biased and inconsistent.[10]

Wages and hours may well be determined simultaneously. According to Barzel (1973) and Killingsworth (1981) marginal productivity may not be constant across the length of a working day and therefore the marginal and average wage will likely differ. There is a "warm up" period at the beginning of a working day during which workers settle into the day's tasks. Beyond a certain critical number of hours per day marginal productivity will fall due to "fatigue effects". Killingsworth also cites the existence of fixed and quasi-

fixed costs of employment to employers such as roll calls and handing out assignments which "may make marginal productivity net of such costs a function of hours worked even if gross marginal productivity is constant" (Killingsworth, 21).[11]

Some of the other variables may be endogenous. It would be extremely complicated to estimate a model in which all the possible interactions between variables are recognized but it is necessary to examine the possibility for the variables with which one is most concerned. In my work these variables are EMPLOYEE and HMO50. Therefore, I test for the endogeneity of HMO50 and EMPLOYEE in addition to WAGE, $WAGE^2$, HOURS, and $HOURS^2$.

The Hausman Test (Hausman 1978) is a general test of the specification of an estimating equation that may provide information as to whether an independent variable is correlated with the error term. It requires running the following regression:

$$Y = X\beta + \hat{X}\alpha + V$$

Where β and α are coefficients, V is an error term and $\hat{X}$ is "a suitably transformed version of X" (Hausman, 1252). Specifically, the Hausman test is a test of the significance of the coefficient of $\hat{X}$ (α) where H_a, the alternative hypothesis that a = 0, implies the orthogonality assumption fails (i.e., X is correlated with the error term). An alternative estimator, $\hat{X}$, must be found that is consistent under both the null and alternative hypotheses. Predicted values of X are suitable since they are correlated with X but uncorrelated with

the error term--so they are consistent whether X is correlated with the error term or not.

Predicted values of the suspect variables are included in the estimating equation along with the actual value of the variables. t-statistics for the coefficients of the predicted variables are reported in table 7. The wage appears to be endogenous in the hours equation: the t-statistic for the coefficient of predicted wage (LWAGEHAT) is 5.8. Similarly, hours appears to be endogenous in the wage equation: the t-statistic for the coefficient of predicted hours (LHRSHAT) is 1.8. The variable HMO50 does not appear to be endogenous

Table 7--Results of the Hausman Tests

Variable Name	Wage Equation t-statistic	Hours Equation t-statistic
LHRSHAT	1.8	-
LWAGEHAT	-	5.8
EMPHAT	0.5	1.4
HMOHAT	0.3	-

in the wage equation (it is not included in the hours equation): the t-statistic for the coefficient of predicted HMO50 (HMOHAT) is 0.3. EMPLOYEE does not appear to be endogenous in the wage equation: the t-statistic for the coefficient of EMPHAT being 0.5.

Several problems may emerge in conducting the Hausman test. It may be sensitive to: (1) the order in which the variables are tested, (2) which additional variables are included in the estimating equation

as independent variables and (3) the exact choice of instruments used to form the predictions. Problems arise when testing for the endogeneity of EMPLOYEE in the hours equation. EMPLOYEE does not appear to be endogenous in the hours equation but this result is sensitive to the choice of instruments used to predict LWAGE, $LWAGE^2$, and EMPLOYEE. Based on the most plausible set of instruments the t-statistic for the coefficient of predicted EMPLOYEE is 1.4. Since most attempts at the Hausman test show that it is appropriate to treat EMPLOYEE as an exogenous variable in both the hours and wage equations this is the approach taken.[12]

OLS estimators are generally unsuitable for estimating a simultaneous equation system. In this study labor demand and supply equations are estimated using the IV technique which replaces the OLS estimator, $\beta = \Sigma xy/\Sigma x^2$, with the estimator $\beta' = \Sigma zy/\Sigma zx$, where z is an instrument chosen to replace x. In large samples, assuming that z is uncorrelated with the error term, β' is a consistent estimator (in small samples OLS may still be preferred). In this study predictions of the endogenous variables are obtained from "first stage" estimations. The predicted values of the endogenous variables serve as instruments in the estimating equations. In tables 8 and 9 OLS estimates are provided alongside IV estimates for comparison.

The labor supply function to be estimated is specified as follows:

$$LHOURS = INTERCEPT + LWAGEHAT + LWAGEHAT^2 + OTHERINC + AGE + AGE^2 + EMPLOYEE + FEMALE + SINGLE + NORTH + CENTRAL + SOUTH + PED + OBGYN + GENERAL + INTERNAL + GRPSZ2 + GRPSZ35 + GRPSZ610 + GRPSZG10 + MULTSPEC$$

The instruments used to predict LWAGEHAT and $LWAGEHAT^2$ are: EXP, EXP^2, BOARD, HMO, EMPLOYEE, MDPC, PEROFF, PERADMIN, ELDERLY, NWHITE, FEMALE, NORTH, CENTRAL, SOUTH, TOWN, PED, OBGYN, GENERAL, INTERNAL, PCINC, GRPSZ2, GRPSZ35, GRPSZ610, GRPSZG10 and MULTSPEC. An "L" in front of a variable name indicates that it is entered in log form. The use of the double log form means that the coefficients of the right hand side variables can be interpreted as elasticities and the slopes are functions of the levels of the variables.[13] All the variables are described in table 3 except GRPSZ2, GRPSZ35, GRPSZ610, AND GRPSZG10 which are dummy variables for group sizes 2, 3-5, 6-10, and greater than 10 respectively--the omitted category being solo practitioners. A finding of a positive sign on LWAGEHAT and a negative sign on $LWAGEHAT^2$ is evidence in support of the backward bending supply curve of labor. The inclusion of AGE^2 in the hours equation recognizes that peak work effort for most workers is concentrated in the middle years.

The weeks equation is formulated similarly except that LWEEK replaces LHOURS as the dependent variable and the predicted values of weekly wage, LWWAGEHAT, and weekly wage squared, $LWWAGEHAT^2$, replace LWAGEHAT and $LWAGEHAT^2$ as independent variables. WWAGE, the weekly wage, is constructed by multiplying WAGE by HOURS. LWWAGEHAT and

LWWAGEHAT^2 are predicted by regressing LWWAGE and LWWAGE^2 against the same set of instruments used to predict LWAGEHAT and LWAGEHAT^2.

The wage equation to be estimated is specified as follows:

LWAGE = INTERCEPT + LHRSHAT + LHRSHAT^2 + EMPLOYEE + HMO50 + EXP + EXP^2 + BOARD + LMDPC + LPEROFF + LPERADMIN + LELDERLY + NWHITE + FEMALE + NORTH + CENTRAL + SOUTH + TOWN + PED + OBGYN + GENERAL + INTERNAL + LPCINC + GRPSZ2 + GRPSZ35 + GRPSZ610 + GRPSZG10 + MULTSPEC

The instruments used to predict LHRSHAT and LHRSHAT^2 are: AGE, AGE^2, FEMALE, SINGLE, EMPLOYEE, OTHERINC, NORTH, CENTRAL, SOUTH, PED, OBGYN, GENERAL, INTERNAL, GRPSZ2, GRPSZ35, GRPSZ610, GRPSZG10 and MULTSPEC. Again "L" denotes the variable is entered in log form.

The hours equation is identified by the exclusion of EXP, EXP^2, HMO50, BOARD, LMDPC, LPEROFF, LPERADMIN, LELDERLY, NWHITE, and LPCINC. It seems reasonable that a physician's age will influence his labor supply and experience will exert more influence on his wage. HMO50 does not appear in the hours equation because being committed to an HMO should not have any independent effect on labor supply. BOARD is not expected to exert any direct influence on a physician's choice of hours--although as an indicator of quality it may well affect the wage. LPEROFF and LPERADMIN, as proxies for output complexity, may influence the wage but should not directly affect hours. LMDPC, LELDERLY, and LPCINC are demand side variables and no apparent case can be made for including them in the labor supply equation. NWHITE may influence demand due to discrimination but there is no expectation that race will influence the taste for leisure. The wage equation is

identified by the exclusion of AGE, AGE^2, SINGLE, and OTHERINC which are not expected to exert any direct influence on the wage.

Selectivity bias is a potential problem in any estimates of labor supply functions but in this study, given the high earnings potential of physicians which imply high participation rates, the bias associated with the exclusion of nonworkers is likely to be small.

The Empirical Results

Estimated coefficients and t-statistics for the hours, weeks, and wage equations are reported in tables 8 and 9. Both IV estimates and OLS estimates are given for purposes of comparison. The adjusted R^2 tends to be rather low (0.15) for the IV equations, in the OLS equations it is inflated by the spurious correlation between HOURS and WAGE since WAGE = INCOME/HOURS.[14]

In the hours equation (table 8) many of the explanatory variables are significant at the 0.05 level or better. In the IV equation LWAGEHAT and $LWAGEHAT^2$ are significant and have the expected signs as are LWAGE and $LWAGE^2$ in the OLS equation, although the coefficients are very different in magnitude. The IV estimates imply a backward bending supply curve of labor at log of the hourly wage of 2.48 whereas the OLS estimates show the critical wage to be 1.69.

Increasing age appears to be associated with increases in hours worked up to 51.5 years of age. EMPLOYEE and FEMALE have the expected negative signs and both are highly significant. SINGLE was not signed a priori since married physicians might be expected to work more hours due to their greater financial responsibilities--or less hours due to

Table 8--Estimates of Labor Supply Functions

Independent Variable	Hours per Week		Weeks per Year	
	IV	OLS	IV	OLS
LWAGEHAT (+)	2.723**		0.240	
	(5.012)		(0.368)	
$LWAGEHAT^2$ (-)	-0.550**		-0.014	
	(6.347)		(0.294)	
LWAGE (+)		0.253**		0.417**
		(3.969)		(5.136)
$LWAGE^2$		-0.047**		-0.032**
		(7.244)		(5.495)
AGE (+)	0.034**	0.021**	0.012**	0.016**
	(6.052)	(4.674)	(3.982)	(3.001)
AGE^2 (-)	-0.000**[a]	-0.000**[b]	-0.000**[c]	-0.000**[d]
	(4.808)	(4.611)	(2.931)	(5.841)
EMPLOYEE (-)	-0.109**	-0.069**	-0.043**	-0.052**
	(4.566)	(3.138)	(3.079)	(4.812)
FEMALE (-)	-0.171**	-0.151**	-0.040*	-0.053**
	(6.027)	(5.697)	(2.485)	(4.070)
SINGLE	0.022	0.010	0.014	0.011
	(0.781)	(0.397)	(1.051)	(0.871)
OTHERINC (-)	-0.031	-0.026	0.010	0.009
	(1.770)	(1.588)	(1.163)	(1.149)
NORTH	0.036	0.013	-0.017	-0.014
	(1.401)	(0.558)	(1.421)	(1.171)
CENTRAL	0.106	0.044	-0.006	0.000[e]
	(4.143)	(1.917)	(0.452)	(0.053)
SOUTH	0.076**	0.063**	0.015	0.024
	(3.043)	(2.918)	(1.163)	(2.195)
PED	-0.056*	-0.064**	0.016	0.014
	(2.032)	(2.603)	(1.229)	(1.179)
OBGYN	0.389	0.119	-0.012	0.023
	(9.046)	(4.594)	(0.393)	(1.767)
GENERAL	-0.042	-0.043	0.017	0.014
	(1.551)	(1.657)	(1.261)	(1.077)
INTERNAL	0.108**	0.030	-0.008	0.003
	(4.468)	(1.434)	(0.586)	(0.250)
GRPSZ2 (-)	0.124**	0.067**	-0.010	0.002
	(4.544)	(2.794)	(0.683)	(0.202)
GRPSZ35 (-)	0.124**	0.050*	-0.019	-0.002
	(4.104)	(2.270)	(1.176)	(0.144)
GRPSZ610 (-)	0.183**	0.091*	-0.049*	-0.028
	(3.893)	(2.245)	(1.974)	(1.396)
GRPSZG10 (-)	0.330**	0.119**	0.004	0.038
	(6.233)	(2.953)	(0.113)	(1.909)

Table 8--Continued

Independent Variable	Hours per Week		Weeks per Year	
	IV	OLS	IV	OLS
MULTSPEC (+)	-0.084**	-0.050	-0.005	-0.011
	(2.624)	(1.691)	(0.297)	(0.724)
Adj. R^2	0.15	0.25	0.07	0.11
# of obs.	1247	1247	1247	1247

*Significant at the 0.05 level

**Significant at the 0.01 level

[a]Coefficient without rounding is -0.000336841

[b]Coefficient without rounding is -0.000242394

[c]Coefficient without rounding is -0.000110877

[d]Coefficient without rounding is -0.000152529

[e]Coefficient without rounding is -0.0005996686

Note: The t-statistics for the coefficients are in parentheses

their higher demand for nonmarket time, particularly in the presence of children. The coefficient of SINGLE is not significantly different from 0. OTHERINC has the expected negative sign and it is approaching significance.

Physicians in the NORTH, CENTRAL, and SOUTH regions work more hours than physicians in the WEST region. This may reflect unobservable differences in preferences that cause physicians who place a higher value on leisure time to select western locations. Pediatricians work less than family practitioners, general practitioners would appear to work similar hours to family

practitioners and obstetrician-gynecologists work significantly more hours than family practitioners.

Contrary to the hypothesis that larger groups provide better opportunities for leisure, all the group size dummies came in positive and significant. This result can probably be attributed to unobservable differences in institutional arrangements (such as better monitoring of physicians) that result in physicians in larger groups working more hours. It was hypothesized that physicians in multispecialty groups would work more hours since they may receive more referrals than physicians working in single specialty groups. Since the coefficient of MULTSPEC is negative and significant this is either not the case, or some other unaccounted for difference in practice in multispecialty groups is causing this result.

The IV weeks equation (table 8) has an adjusted R^2 of 0.07 and few of the independent variables are significant. Lack of explanatory power and insignificant coefficients for weeks-worked equations are common in the literature and appear to be the result of the lack of variability in weeks worked per year. The coefficients of LWWAGEHAT and $LWWAGEHAT^2$ are completely insignificant in the IV equation, although they have the correct sign and are highly significant in the OLS estimation. This points to one of the problems with using OLS since "correct" results may lead the researcher into assuming that the estimating equation is correctly specified--when tests for endogeneity clearly show OLS estimation to be inappropriate.

The coefficients on LWAGEHAT AND $LWAGEHAT^2$ in the IV equation imply that the supply of weeks bends backwards at a log of the weekly

wage of 8.57. The coefficients of AGE and AGE^2 are significant and have the expected sign. They imply a critical age of 54 years which is very similar to the result from the hours equation. The coefficient of EMPLOYEE is significant in the weeks equation and has the expected negative sign. Being an employee appears to be associated with fewer weeks worked per year in addition to fewer hours worked per week. Specialty choice has no discernible effect on weeks worked per year, at least for these five specialties. Differences in hours worked per year by specialty appear to manifest themselves in differences in hours worked per week not in weeks worked per year. In terms of weeks worked per year being in a group appears to have the expected negative effect on labor supply, except for groups greater than 10, although it is significant only in the case of physicians working in groups containing 6-10 physicians. This suggests that the increased opportunities for leisure that occur when physicians work in groups may be experienced in terms of more vacation weeks per year, while hours worked per week appear to be higher for groups.

The adjusted R^2 for the IV estimate of the wage equation (table 9) is 0.13 which means this equation explains only a small portion of the variation in wages between physicians. In the IV estimation LHRSHAT and $LHRSHAT^2$ are significant and have the expected signs (positive and negative respectively). In the OLS estimates LHOURS and $LHOURS^2$ both have negative signs although they are insignificant. The coefficients on LHRSHAT and $LHRSHAT^2$ in the IV equation imply that the hourly wage declines beyond a log of the weekly work effort of 3.97.

Table 9--Estimates of Wage Equation (Dependent Variable Hourly Wage)

Independent Variable	IV	OLS
LHRSHAT (+)	31.997**	
	(3.690)	
$LHRSHAT^2$ (-)	-4.025**	
	(3.585)	
LHOURS (+)		-0.598
		(1.228)
$LHOURS^2$ (-)		-0.029
		(0.464)
EMPLOYEE	-0.038	-0.093*
	(0.549)	(1.980)
HMO50	0.186	0.113
	(1.793)	(1.167)
EXP (+)	(0.013)	0.033**
	(1.699)	(6.120)
EXP^2 (-)	0.000[a]	-0.001**
	(0.533)	(5.535)
BOARD (+)	0.097*	0.102**
	(2.441)	(2.783)
LMDPC (-)	-0.076*	-0.053
	(2.065)	(1.563)
LPEROFF (-)	0.171**	-0.080
	(3.370)	(1.579)
LPERADMIN (-)	-0.015	-0.034
	(0.635)	(1.585)
LELDERLY (+)	0.076	0.046
	(1.716)	(1.130)
NWHITE (-)	-0.074	-0.058
	(1.450)	(1.227)
FEMALE (-)	0.4049	-0.138*
	(0.436)	(2.524)
NORTH	0.093	0.031
	(1.684)	(0.620)
CENTRAL	0.069	0.037
	(1.177)	(0.748)
SOUTH	0.154*	0.121**
	(2.303)	(2.585)
TOWN (+)	0.076	0.071
	(1.373)	(1.383)
PED	0.029	0.020
	(0.403)	(0.373)
OBGYN	0.641**	0.409**
	(7.465)	(7.364)
GENERAL	-0.129	0.003
	(1.501)	(0.052)

Table 9--Continued

Independent Variable	IV	OLS
INTERNAL	0.193**	0.129**
	(3.871)	(2.860)
LPCINC (+)	0.091	0.019
	(1.012)	(0.223)
GRPSZ2	0.070	0.155**
	(1.221)	(3.147)
GRPSZ35	-0.019	0.205**
	(0.261)	(4.417)
GRPSZ610	0.026	0.230**
	(0.264)	(2.716)
GRPSZG10	0.084	0.363**
	(0.779)	(4.197)
MULTSPEC	0.064	-0.042
	(0.817)	(0.675)
Adj. R^2	0.131	0.25
# of obs.	1247	1247

*Significant at the 0.05 level

**Significant at the 0.01 level

[a]Coefficient on EXP^2 (without rounding) is -0.0001088465

Note: The t-statistics for the coefficients are in parentheses

(The average weekly hours reported by all physicians is 49 hours). EMPLOYEE has a negative sign but it is completely insignificant while HMO50 is positive and on the verge of being significant. Contractual arrangements do not appear to have any obvious effect on wages.

Pediatricians, internists and general practitioners do not earn significantly different wages from family practitioners. Obstetrician-gynecologists, on the other hand, have a positive wage differential that is highly significant. EXP and EXP^2 are not significant in the wage equation--but the coefficients on EXP and EXP^2

imply that the wage begins to decline beyond 60 years of experience--a figure that is outside the range for all physicians. BOARD is positive and significant as would be expected since it proxies physician quality.

Surprisingly, LPEROFF is positive and significant. It was hypothesized that a high percentage of time spent seeing patients in the office was an indicator of a less complicated output mix. This result may be attributable to unobservables in output mix. Alternatively, physicians spending a large percentage of their time in the office may have higher remuneration per hour of medical practice because of the efficiency gains associated with specialization. LPERADMIN had the expected negative sign although it is not significant.

The coefficient of LMDPC is negative and significant--higher physician density has a depressing effect on physician remuneration.

LELDERLY is positive although not significant. However, its positive sign is consistent with the view that a high elderly population in the county in which the physician practices is an indicator of high demand for physician services.

The coefficient of NWHITE is negative and insignificant. The predicted negative sign arises from the possibility that racial discrimination on the part of employers, other physicians, or patients may depress the returns to nonwhite physicians. The coefficient of FEMALE is positive although not significant. There is certainly no evidence to suggest that being female is associated with lower hourly wages. Female physicians clearly earn less than male physicians, but

the difference would appear to be attributable to differences in hours worked, specialty choice, and employment status. However, all discussion of female physicians, either in terms of their hours or wages must be tentative, since the PPCIS along with most surveys specifically excludes part time workers--and they are likely to be disproportionately female.

Physicians working in the NORTH, CENTRAL and SOUTH regions of the country appear to earn more than those located in the WEST region although only in the case of the SOUTH is the coefficient significant. TOWN was included to reflect the higher cost of living in urban areas. As expected TOWN has a positive sign although it is not significant.

Conclusion

This chapter detailed the estimation of simultaneous labor demand and supply functions using two measures of labor supply. The principal interest was the impact of employee status on the labor supply and the wage of physicians. It is found that employee status has a negative effect on the hours supplied per week by physicians and no effect on the number of weeks supplied per year. On the wage side employee status does not appear to be associated with any reduction in compensation per time period. This is consistent with a "supply-side" explanation for the existence of physician-employees (i.e., some physicians express a preference for the lifestyle afforded by a salaried position).

Notes

[1]There is some controversy about whether such a surplus actually exists. Hurdle and Pope provide evidence of a decline in physician productivity between 1975 and 1984 that partially offsets the effects of increasing numbers of physicians.

[2]In support of such a contention it is noted in chapter IV that physician-employees are more common in areas of high physician density--although these would also seem to be areas of high physician demand.

[3]The alternative explanation for any observed reduction in work effort on the part of employed physicians is that since employed physicians are not directly compensated for marginal work effort they have an incentive to shirk.

[4]In an earlier specification of my model I examined the possibility of using the physician's customary fee as a proxy for the wage. This idea was abandoned partly because the simple correlation coefficient between the wage and the customary fee was approximately equal to zero.

[5]Kehrer (1976, footnote; 541-542) reports that, of female physicians responding to a question concerning the occupation of their husband, 49.6% reported being married to another physician.

[6]To the extent that ethical feelings influence labor supply we add to the unexplained portion of the regression equation by this omission. It may also lead to bias in the coefficients.

[7]See, for example, Intrilligator (1978, 384).

[8]The other main explanation advanced for the relationship between hours and wages is not applicable to doctors since it involves the notion of productive consumption. In subsistence economies higher wages may improve the physical condition of workers and enable them to work longer hours.

[9]With minor adjustments to the specification of the Hausman test EMPLOYEE can appear to be endogenous. In the instrumental variable estimation of the hours equation if EMPLOYEE is assumed to be endogenous it does not alter coefficients or significance for variables other than EMPLOYEE. For EMPLOYEE, however, the assumption of endogeneity in the hours equation results in a sign reversal from negative to positive and the loss of significance.

[10]Double log functions are used here for convenience in interpreting elasticities. Results with these functions are superior in terms of the general significance of the coefficients.

[11]The IV technique utilizes predicted wages that should be free of measurement error and therefore it avoids the problem of spurious correlation (Borjas 1980, 414).

CHAPTER VI

CONCLUSION

Contribution of Dissertation

This dissertation examines theoretical and empirical issues related to physician behavior. At a theoretical level, a model of an HMO was developed to evaluate the role of ethics and remuneration schemes in the provision of care within an HMO. The model predicts that physicians will supply differing levels of labor supply and care depending on the way in which they are compensated. An assumption of ethics will raise physician labor supply and care under any compensation scheme considered. However, ethical behavior only mitigates the effect of incentive schemes: there is no one "appropriate" level of care that an ethical physician will produce regardless of how he is compensated. This means that ethics alone are not sufficient to ensure that patients receive the correct amount of care. Salaried physicians have an incentive to overuse nonphysician inputs because they are a substitute for physician time. Ethical physicians will overuse nonphysician inputs because they raise physician utility via their effect on patient utility at no cost to the physician.

An empirical test of one implication of the theoretical work was conducted. Simultaneous labor demand and supply functions were

estimated for primary care physicians. The results suggest that employed physicians supply less labor than their self-employed colleagues, both in terms of hours per week and weeks per year but no apparent differences between the wages of the employed and self-employed emerge.

Future Research

The most important results of my work are that salaried physicians will tend to under-supply labor and may under-treat patients even if they are ethical. In addition, they will supply care using an inefficiently large amount of nonphysician inputs. I was able to confirm the negative effect on labor supply at the empirical level but many interesting questions remain to be answered.

It would be very desirable to obtain data on medical input usage by all physicians including employees. The PPCIS data set systematically lacks information on input usage by employed physicians. To the best of my knowledge this information is currently not available. Data on input usage would enable one to test the hypothesis that salaried physicians tend to overuse nonphysician inputs. It is also of interest to ask what mechanisms employing hospitals use to monitor physician performance and to prevent the overuse of other inputs.

Does it matter what type of organization employs physicians? This question was not considered in this dissertation but it is possible that for-profit hospitals and not-for-profit hospitals differ in their treatment of salaried physicians. Finally, and rather

esoterically, the question of physician ethics is very important although to date no satisfactory empirical proxy for ethics has been found.

Research on alternative institutional arrangements in the medical industry is of great interest at the practical level due to the recent expansion in enrollment in such plans, and the large expenditure on health care both privately and publicly. This dissertation provides some insight into one issue in this area.

APPENDIX A

THE MALPRACTICE CONSTRAINT ON THE LEVEL OF CARE PROVIDED

$\bar{C}$ may be thought of as a malpractice constraint on the level of care. Care will be provided up to the point where the marginal cost of providing the last unit of care is equal to the expected marginal cost saving associated with avoidance of a lawsuit by providing the last unit of care. Some authors make the point that lawsuits and care may be positively related as doctors have to cover themselves in a litigious climate. I feel that this incentive is outweighed by the desire to cost minimize--at least within the context of an HMO.

The HMO seeks to maximize profit (π), where:

$$\pi = R - Z$$

In an HMO R (total revenue) is fixed and Z (total cost) is:

$$Z = C_T + P_1 L_T$$

Where: C_T = total cost of providing patient care in period T, excluding malpractice costs.

L_T = number of lawsuits incurred in period T

P_1 = cost of one lawsuit--treated as a constant across lawsuits

The cost of one unit of care is μ: Therefore:

$$C_T = \mu C$$

Where: C = number of units of care provided.

The number of lawsuits is inversely related to the amount of care provided:

$$L_T = \Omega/C$$

Where: Ω = a constant.

As Ω increases the probability of a lawsuit rises which means that for any given level of care the number of lawsuits incurred is higher. This specification implies that once a lawsuit occurs there is a known cost associated with it. I am assuming that it is the number of suits that is sensitive to the level of care provided. Furthermore, the relationship between the number of lawsuits and care is known with certainty. Profit is therefore given by the following:

$$\pi = R - \mu C - P_l\Omega/C$$

Maximizing profit with respect to C (the level of care provided) the first order condition is:

$$\delta\pi/\delta C = -\mu + P_l\Omega/C^2 = 0$$

Solving for $\bar{C}$, where $\bar{C}$ is the profit maximizing level of care referred to in the text as the malpractice constrained level of care:

$$\bar{C} = (P_l\Omega/\mu)^{1/2}$$

The partial derivatives of $\bar{C}$ with respect to P_1, Ω, and μ respectively are as follows:

$$\delta\bar{C}/\delta P_1 = \Omega^{1/2}/2(P_1\mu)^{1/2} \qquad > 0$$

$$\delta\bar{C}/\delta\Omega = P_1^{1/2}/2(\mu\Omega)^{1/2} \qquad > 0$$

$$\delta\bar{C}/\delta\mu = -(P_1\Omega)^{1/2}/\mu^{1.5} \qquad < 0$$

The level of care provided is positively related both to the cost of a lawsuit (P_1) and Ω, and negatively related to the cost of a unit of care (μ). Ω can be regarded as a measure of the sensitivity of the number of lawsuits to the level of care provided.

Competitive Markets for Medical Care

Is it possible that $\bar{C}$ is the market clearing level of medical care? If this were true then $\bar{C}$ would be the optimal level of care from a societal point of view, ignoring distributional considerations and also possible externalities. Since the market for medical care is characterized by poor information this is unlikely to be so.

Furthermore, an enrollee may not be responsible for choosing and paying for his own health insurance. Typically a worker receives health insurance as part of the benefits package from the firm in which he is employed. The employer will face two competing objectives, to provide a compensation package that is attractive to its workers and to minimize labor costs. If we make the assumption that the labor market is competitive the firm will have to offer the

market total compensation package. However, to the extent that the employee is unable to evaluate the quality of a health care scheme the employer may not offer the competitive level of health benefits.

It does not seem reasonable to conceptualize $\bar{C}$ as being a competitive constraint given the particular features of the market for medical care. Throughout this paper $\bar{C}$ will be assumed to be set by a malpractice constraint and hence $\bar{C}$ will constitute an inadequate level of care.

APPENDIX B

THE RELATIONSHIP BETWEEN MEDICAL CARE AND HEALTH

It is clear that the patient is interested in obtaining health, not purchasing medical care. According to Pauly (1980) health is produced in the following way:

$$H = H(H_0, M_1, M_2)$$

where: H_0 = existing state of health

M_1 = input of physician time into the medical production function

M_2 = input of nonphysician time into the medical production function

H_0 is a very important determinant of the level of health actually achieved but H_0 is partially beyond the control of the patient since it represents his genetic endowement and random life experiences. However, as Fuchs and Zeckhauser (1987) point out, it is largely self-produced by the patient as a result of decisions the patient makes about lifestyle (diet, smoking, etc.) H_0 is also partially determined by previous decisions by the patient to consume medical services. However, the concept of health is hard to model. I will be following the practice, common in the literature, of assuming that the patient is primarily interested in the consumption of medical

services and that the cost of the services obtained is a satisfactory proxy for the utility derived from their use.

APPENDIX C

SIMULATION OF VALUES OF M_1

It is not possible to provide a value for M_1^S and M_1^P. However, we have shown theoretically that $M_1^S = M_1^P$ and that they are both less than M_1^{SE}. Tables 10-13 show values for M_1 in the other four cases (SE, W, WE, PE) based on differing assumptions about the parameter values. In each table one parameter is varied while the other parameters are set equal to 0.5. Total discretionary time available to the physician (T) is set equal to 12. M_1 in each case does not vary as ϕ_2 varies, for the salaried ethical case and profit sharing ethical case $M_1 = 4.00$, for the wage earning case $M_1 = 6.00$, for the wage earning ethical case $M_1 = 7.20$ for all values of ϕ_2.

Table 10. Variations in α

It is confirmed that M_1^{SE}, M_1^W, M_1^{WE}, and M_1^{PE} are all decreasing in α. Furthermore, for all values of α it can be seen that $M_1^{SE} < M_1^W < M_1^{WE} < M_1^{PE}$.

Table 11. Variations in ß

M_1^{SE} is invariant with respect to changes in ß. The ranking between M_1^{SE} and M_1^W is reversed at a value for ß of .3. If ß < .3 then

Table 10--Impact of Variations in the Level of α on the Amount of M_1 Supplied

α	$M_1^{SE}=M_1^{PE}$	M_1^{W}	M_1^{WE}
0.10	8.57	10.00	10.59
0.20	6.67	8.57	9.47
0.30	5.45	7.50	8.57
0.40	4.62	6.67	7.83
0.50	4.00	6.00	7.20
0.60	3.53	5.45	6.67
0.70	3.16	5.00	6.21
0.80	2.86	4.62	5.81
0.90	2.61	4.29	5.45
1.00	2.40	4.00	5.14

Note: $\beta = \tau = \phi_1 = \phi_2 = .5$, T = 12 hours

Table 11--Impact of Variations in the Level of β on the Amount of M_1 Supplied

β	$M_1^{SE}=M_1^{PE}$	M_1^{W}	M_1^{WE}
0.10	4.00	2.00	4.94
0.20	4.00	3.41	5.68
0.30	4.00	4.50	6.29
0.40	4.00	5.33	6.78
0.50	4.00	6.00	7.20
0.60	4.00	0.11	7.55
0.70	4.00	7.00	7.86
0.80	4.00	7.38	8.13
0.90	4.00	7.71	8.36
1.00	4.00	8.00	8.57

Note: $\alpha = \tau = \phi_1 = \phi_2 = .5$, T = 12 hours

Table 12--Impact of Variations in the Level of τ on the Amount of M_1 Supplied

τ	$M_1^{SE}=M_1^{PE}$	M_1^{W}	M_1^{WE}
0.10	1.09	6.00	6.29
0.20	2.00	6.00	6.55
0.30	2.77	6.00	6.78
0.40	3.43	6.00	7.00
0.50	4.00	6.00	7.20
0.60	4.50	6.00	7.38
0.70	4.94	6.00	7.56
0.80	5.33	6.00	7.71
0.90	5.68	6.00	7.86
1.00	6.00	6.00	8.00

Note: $\alpha = \beta = \phi_1 = \phi_2 = .5$, T = 12 hours

Table 13--Impact of Variations in the Level of ϕ_1 on the Amount of M_1 Supplied

ϕ_1	$M_1^{SE}=M_1^{PE}$	M_1^{W}	M_1^{WE}
0.10	1.09	6.00	6.29
0.20	2.00	6.00	6.55
0.30	2.77	6.00	6.78
0.40	3.43	6.00	7.00
0.50	4.00	6.00	7.20
0.60	4.50	6.00	7.38
0.70	4.94	6.00	7.56
0.80	5.33	6.00	7.71
0.90	5.68	6.00	7.86
1.00	6.00	6.00	8.00

Note: $\alpha = \beta = \tau = \phi_2 = .5$, T = 12 hours

$M_1^{W} < M_1^{SE}$. For values of β greater than .3 $M_1^{W} > M_1^{SE}$. When this was examined in chapter III it was found that what is important is the size of β relative to τ. M_1^{PE} is larger than M_1^{SE}, M_1^{W}, and M_1^{WE} for all values of β.

Table 12--Variations in τ

M_1^W is constant since it derives from a standard optimizing case. M_1^{SE} and M_1^{WE} and M_1^{PE} are increasing in τ. The ranking of the four cases is consistent across values of τ--$M_1^{SE} < M_1^W < M_1^{WE} < M_1^{PE}$.

Table 13. Variations in ϕ_1

M_1^W is constant since it derives from the standard optimizing case. $M_1^{SE} = M_1^{PE}$ and M_1^{WE} are increasing in ϕ_1. Ethical physicians supply more labor the higher the productivity of their own time. The wage earning physician also earns a higher income by supplying more care.

BIBLIOGRAPHY

Arrow, Kenneth J. "Uncertainty and the Welfare Economics of Medical Care," American Economic Review 53 (December 1963): 941-973.

Barzel, Yoram. "The Determination of Daily Hours and Wages." Quarterly Journal of Economics 87 (May 1973): 220-238.

Becker, Gary. "A Theory of the Allocation of Time." Economic Journal 75 (September 1965): 493-517.

Borjas, George J. "The Relationship Between Wages and Weekly Hours of Work: The Role of Division Bias." Journal of Human Resources 15 (Summer 1980): 409-423.

Boulier, Bryan L. "Supply Decisions of Self-Employed Professionals: The Case of Dentists." Southern Economic Journal 45 (January 1979): 892-902.

Dionne, Georges and Andre-Pierre Contandriopoulos. "Doctors and Their Workshops--A Review Article." Journal of Health Economics 4 (March 1985): 21-33.

Eastaugh, Steven R. Medical Economics and Health Finance. Boston, MA: Auburn House, 1981.

Ellis, Randall P., and Thomas G. McGuire. "Provider Behavior Under Prospective Reimbursement: Cost Sharing and Supply." Journal of Health Economics 5 (June 1986): 129-151.

Farley, Pamela J. "Theories of the Price and Quantity of Physician Services: A Synthesis and Critique." Journal of Health Economics 5 (December 1986): 315-333.

Feldstein, M. "Econometric Studies of Health Economics." Frontiers of Quantitative Economics II, Edited by M. Intriligator and D. Kendrick. NL: North Holland Amsterdam, 1974.

Ferguson, Brian S. "Physician Objectives and Resource Allocation." Journal of Health Economics 4 (March 1985): 35-42.

Fuchs, Victor R., and Richard Zeckhauser. "Valuing Health--A "Priceless" Commodity." AEA Papers and Proceedings 77 (May 1987): 263-268.

Hamermesh, Daniel S., and Albert Rees. The Economics of Work and Pay, 3rd ed., New York: Harper and Row, 1984.

Harris, Jeffrey E. "The Internal Organization of Hospitals: Some Economic Implications." Bell Journal of Economics 8 (Autumn 1977): 467-482.

Hart, Oliver, and Bengt Holmstrom. "The Theory of Contracts" in Advances in Economic Theory, Edited by T. Bewley. Cambridge U.K.: Cambridge University Press, 1985.

Hausman, J. A. "Specification Tests in Econometrics." Econometrica 46 (November 1978) 1251-1271.

Hurdle, Sylvia, and Gregory C. Pope. "Physician Productivity: Trends and Determinants." Inquiry 26 (Spring 1989) 100-115.

Intriligator, Michael D. Econometric Models, Techniques, and Applications Englewood Cliffs NJ: Prentice Hall (1978).

Kehrer, Barbara H. "Factors Affecting the Incomes of Men and Women Physicians: An Exploratory Analysis." Journal of Human Resources 11 (Fall 1976) 526-545.

Killingsworth, Mark R. Labor Supply Cambridge UK: Cambridge University (1983).

Kittrell, Alison. "HMO Enrollment Tops 30 Million: Interstudy." Business Insurance 22 (May-August 1988) 18, 35.

Mitchell, Janet B. "Why Do Women Physicians Work Fewer Hours Than Men Physicians?" Inquiry 21 (Winter 1984) 361-368.

Newhouse, Joseph P. "A Model of Physician Pricing." Southern Economic Journal 37 (October 1970) 174-183.

Osterman, Paul. "White-Collar Internal Labor Markets." in Internal Labor Markets, Edited by Paul Osterman. Cambridge MA: The MIT Press (1984).

Pauly, Mark V. Doctors and Their Workshops Chicago IL: University of Chicago Press (1980).

_____. "Reflections on Using Physician Agents to Minimize the Cost of Health." (editorial) Journal of Health Economics 4 (March 1985) 79-81.

Pauly, Mark V. and Redisch, Michael. "The Not-For Profit Hospital as a Physicians' Cooperative." American Economic Review 63 (March 1973) 87-99.

Reinhardt, Uwe E. "A Production Function for Physicians' Services." Review of Economics and Statistics 54 (February 1972) 55-66.

Reinhardt, Uwe E. "The Theory of Physician-Induced Demand--Reflections after a Decade." (editorial) Journal of Health Economics 4 (June 1985) 187-193.

Rosenbach, Margo L. "Merger of Secondary Data: Final Report August 1985." National Opinion Research Center, 1985 (mimeo).

Scheffler, Richard M., and Louis F. Rossiter. "Compensation Schemes and the Labor Supply of Dentists." Quarterly Review of Economics and Business 23 (Autumn 1983) 29-43.

Sloan, Frank A., and Bruce Steinwald. Hospital Labor Markets. Boston MA: Lexington (1980a).

_____. Insurance, Regulation, and Hospital Costs. Boston MA: Lexington (1980b).

Vahovich, Stephen G. "Physicians' Supply Decisions by Specialty: 2SLS Model." Industrial Relations 16 (February 1977) 51-60.

Woodward, Robert S., and Frederick Warren-Boulton, "Considering the Effects of Financial Incentives and Professional Ethics on 'Appropriate' Medical Care." Journal of Health Economics 3 (December 1984) 223-237.

Yellen, Janet L. "Efficiency Wage Models of Unemployment." American Economic Association Papers and Proceedings, 74 (May 1984) 200-205.