

THE PHYSICIAN LABOR MARKET IN A MANAGED CARE-DOMINATED ENVIRONMENT

RICHARD A. HIRTH and MICHAEL E. CHERNEW*

The belief that doctors respond to declining demand by treating patients more aggressively has created skepticism about relying on market forces to restructure physician supply. We argue that even if the physician labor market is dysfunctional under fee-for-service incentives, it can perform better as managed care becomes dominant. Our model implies a nonlinear effect of managed care penetration on incomes. Physicians can offset most or all of initial declines in demand, but cannot insulate themselves indefinitely. This may explain the observation that, until recently, the growth of managed care has not been accompanied by large physician income changes. (JEL I11, J31)

I. INTRODUCTION

Health maintenance organizations (HMOs) and other managed care arrangements are rapidly growing alternatives to fee-for-service (FFS) insurance in the United States. HMO enrollment has grown from under 3 million in 1970 (InterStudy [1986]) to over 46 million by early 1995 (InterStudy [1995]). Little is known about how these changes in health care financing will ultimately affect physicians' incomes, but managed care is undoubtedly altering the economic environment of medical practice. The shift from FFS to salaried or capitated payment reduces physicians' incentives to overtreat.¹ Further, managed care

plans control physicians' access to their enrollees by monitoring practice patterns and removing overly aggressive doctors from their panels. HMOs also employ a different physician mix (with fewer specialists) than the mix that characterizes the FFS environment. Thus, a hypothesized consequence of rising HMO market share is reduced demand for specialists. Wennberg et al. [1993] estimate potential surpluses ranging from 10% for emergency medicine to 210% in pathology.

Surpluses of these magnitudes would be expected to reduce earnings. However, until recently, physicians have enjoyed stable or rising incomes (Simon and Born [1996]). This may be explained by asymmetric information between patients and physicians: if patients are poorly informed about the effectiveness of treatment options, doctors may be able to protect their incomes from the loss of patients to managed care by prescribing more aggressive treatment to their remaining FFS patients. Despite mixed evidence on the extent of supplier-induced demand (SID) (e.g., Fuchs [1978], Newhouse et al. [1982], Dranove and Wehner [1994], Rizzo and Blumenthal [1996]), this hypothesis remains popular among policymakers:

* We thank Tom McGuire, Dean Smith, Kevin Frick, James Baumgardner, and seminar participants at Northwestern University, the University of Florida, and the American Economic Association meetings for helpful comments and suggestions without implying any responsibility on their parts for remaining errors.

Hirth: Assistant Professor, University of Michigan, Ann Arbor, Phone 1-734-936-1306, Fax 1-734-764-4338 E-mail rhirth@umich.edu

Chernew: Assistant Professor, University of Michigan, Ann Arbor, Phone 1-734-936-1193, Fax 1-734-764-4338 E-mail ggbz@umich.edu

1. Physicians in group and staff model HMOs are usually salaried. Physicians in other types of managed care plans, such as independent practice associations or preferred provider organizations, are paid either at a discounted FFS rate or by capitation (a fixed payment per patient per month independent of services actually performed). Although the effects of managed care on the demand for physician services may depend on the mix of managed care plans, this paper uses the generic terms "HMO" and "managed care" for ease of exposition.

ABBREVIATIONS

FFS: Fee-For-Service
HMO: Health Maintenance Organization
MP: McGuire and Pauly
SID: Supplier-Induced Demand
TIH: Target Income Hypothesis

The proponents of quotas for the size and mix of the health workforce argue that the supply side of the health sector is simply impervious to signals from the demand side... [I]t is widely believed that ... physicians enjoy almost total discretion in dictating the volume and composition of health care they render ... allow[ing] physicians to extract very high "target incomes" virtually regardless of the local physician-to-population ratio. (Reinhardt [1994])

Interest in physician incomes arises in part from the central role that incomes play in sending appropriate signals about changing supply and demand conditions. If FFS physicians can maintain their incomes as the FFS patient pool shrinks, decisions about specialty, location, and HMOs' staffing patterns would have little reason to change. Alternatively, if FFS incomes are sensitive to managed care penetration, we would expect doctors to increasingly practice in areas and specialties in which employment opportunities have become relatively favorable, pursue non-clinical positions, or retire. Likewise, HMOs would reduce substitution towards non-physician providers and away from specialist physicians.

The consensus among health workforce experts is that the U.S. has too many physicians and too many specialists relative to primary care physicians and that market adjustments will be insufficient to restructure the physician supply. Thus, many health workforce experts and medical organizations have called for centralized workforce planning (e.g., Mitka [1997]), Committee on the U.S. Physician Supply [1996], Pew [1995], Cohen and Todd [1994], Schroeder [1994], Wennberg et al. [1993]). Proposals include limits on the number and specialty composition of residency positions, restrictions on medical school class sizes and immigration by international medical graduates, and incentive payments for early retirement, retraining, or diversion of physicians into non-clinical positions.

Before advocating interventions in the medical labor market, it is worth exploring the potential for market forces to generate the desired physician supply changes. Many supporters of workforce policies presume that the failures of a FFS-dominated physician services market would persist even in a radically

different context. Ironically, the resurgence of interest in physician workforce planning coincides with changes in the medical marketplace that may force the physician labor market far closer to a competitive outcome even without intervention. The purpose of this paper is not to assess whether or not this market has failed in the past. Rather, we will argue that *even if the physician labor market is quite dysfunctional under FFS incentives*, it may perform better if managed care becomes dominant.

We posit that the primary mechanism through which managed care affects FFS physicians is a declining FFS patient per physician ratio. Thus, the model developed here explores how FFS physicians respond to the loss of patients. The main implication is that the effect of managed care penetration on FFS physician incomes becomes stronger as market share rises. By increasing service intensity, physicians may offset most or all of the effects of initial declines in practice size, but they cannot indefinitely insulate themselves from market forces.

This non-linear effect of HMO penetration on FFS earnings provides a rationale for the puzzling stylized fact that for years the growth of managed care did not reduce physician incomes or narrow the earnings gap between primary care and specialty medicine. Further, such a non-linear effect suggests that the physician labor market will respond strongly to the continued growth of managed care, reducing the need for costly workforce planning measures. Physicians' ability to induce demand for their services has remained a contentious issue in health economics for decades (e.g., Rizzo and Blumenthal [1996] and accompanying commentaries). In a small way, our model helps bridge this dichotomy of opinion by showing how declines in FFS practice size eventually affect incomes regardless of attempts to induce demand.

II. SUPPLIER-INDUCED DEMAND

In standard economic theory, firms move down their supply curves if the exogenously set price of a service falls. Likewise, an expansion of supply causes equilibrium price to fall. These predictions have not always been verified in the physician services market. For example, Fuchs [1978] and Cromwell and

Mitchell [1986] find that an increase in the supply of surgeons leads to both higher utilization of surgical procedures and higher prices. Likewise, Rizzo and Blumenthal [1996] show that physicians earning less than their "target" incomes charge higher prices.

One explanation for such paradoxical results is the target income hypothesis (TIH). Under the TIH, physician-firms pursue a "target income" defined by historical income levels or incomes earned by other professionals rather than maximizing profits. Physicians facing a fee decrease or greater competition attempt to attain their target incomes by shifting out patients' demand for their services. Although the TIH has sometimes been considered to be at odds with utility maximization (Evans [1974]), it can be reconciled with utility maximization by making several strong assumptions about the utility function (Formby and Millner [1985]).

McGuire and Pauly [1991] (MP) analyze a physician's response to a fee reduction in a model that encompasses both profit maximization and income targeting behavior. The change in demand inducement in response to a fee reduction depends on the physician's marginal utility of income. Consider two extreme cases. First, with no income effects (constant marginal utility of income with additively separable utility), physicians maximize profits net of the value of foregone leisure and any disutility of inducing demand. Quantity of services and incomes unambiguously decline as physicians move down their supply curves. Although physicians may be inducing demand, they do not respond to a fee reduction by raising service intensity further. At the other extreme is the target income hypothesis, which arises when the marginal utility of income is very high below the target and very low above the target. This compels physicians to offset a fee reduction by prescribing enough additional services to maintain their incomes.

In intermediate cases, quantity could either rise or fall in response to a fee cut. If income effects are strong enough, physicians will induce more demand when fees fall. However, increased service intensity offsets only a portion of the effect of lower fees on physician incomes. Inducement is limited by the increasing marginal disutilities of labor and inducement.

Dranove [1988] models demand inducement differently, showing that if physicians develop reputations for the "aggressiveness" of their practice styles, demand inducement can be limited by patients' refusal to consent to treatment. Rochaix [1989] derives similar results using second opinions as the mechanism limiting physicians' ability to over-prescribe.

III. A MODEL OF THE EFFECTS OF DECLINING FFS PRACTICE SIZE

Our model builds upon the McGuire and Pauly [1991] model because it encompasses profit maximization and the TIH. The TIH case is useful because it demonstrates how economic pressures affect FFS physicians even if the physician labor market is very dysfunctional. We employ Dranove's [1988] and Rochaix's [1989] insights about patient behavior to analyze the effect of a decline in fee-for-service practice sizes. Since HMOs generally use fewer physicians per enrollee than the FFS sector, we hypothesize that decreasing the FFS patient to physician ratio is the primary mechanism through which the growth of managed care affects FFS physicians.²

Model of Physician Behavior

Physician utility is a function of profit from medical practice (π), leisure (L) and inducement effort (I): $U(\pi, L, I)$.³ Assumptions are analogous to those in MP. Utility increases in income ($U_\pi > 0$) and leisure ($U_L > 0$) and decreases in inducement ($U_I < 0$; physicians incur a psychic cost when violating the agency relationship by recommending services a fully informed patient would not demand). Marginal utilities of income and leisure are decreasing and marginal disutility of inducing demand is increasing ($U_{\pi\pi} < 0$, $U_{LL} < 0$, $U_{II} < 0$). To keep the model tractable, the cross partials are assumed to be zero ($U_{\pi L} = 0$, $U_{\pi I} = 0$, $U_{LI} = 0$). We also consider

2. The model describes the effects of any change that alters the physician to patient ratio, but we believe it is natural to use it to analyze the impact of managed care. Section IV sketches a model with endogenous HMO market share.

3. In this section, we are only modeling physicians who exclusively treat fee-for-service patients. Below, we will also model the choice of whether or not to see HMO patients.

the special case in which physicians experience no disutility from inducing demand ($U_I = 0$).

As physicians raise inducement, the quantity of services per patient [$X(I)$] rises [$X_I(I) > 0$]. Define inducement effort such that $X_{II}(I) \leq 0$ (as inducement effort rises, the quantity of services to which a patient consents rises at a non-increasing rate).⁴

McGuire and Pauly [1991] assume a fixed practice size. That is, high inducement effort does not drive away patients and physicians cannot steal patients away from aggressive inducers. In the context of the growth of managed care, the dwindling number of patients per fee-for-service physician is a crucial consideration. Thus, we employ the idea developed by Dranove and Rochaix that patients leave physicians whom they believe to be overly aggressive as the basis for relaxing this assumption. We hypothesize that the number of patients is a decreasing function of inducement effort. As I increases, more patients seek care elsewhere because they believe that their physician is acting as a poor agent. Denote the number of patients in a physician's practice by $n(I;N)$, where N represents a patient population shift parameter which is not under the control of any individual doctor. N can be thought of as the average FFS practice size. As this ratio declines, each physician loses some patients ($n_N > 0$).⁵ The number of patients in a physician's practice decreases at an increasing rate in inducement effort [$n_{II}(I;N) < 0$ and $n_{IN}(I;N) < 0$]. That is, patients become increasingly resistant as physicians recommend more aggressive treatment.

The cross partial derivative, n_{IN} , is ambiguous in sign and reflects the selection of patients remaining in a physician's practice as N falls. If patients leave the practice at the same rate at any N , n_{IN} would be negative (larger practice size implies that a larger absolute loss of patients in response higher in-

ducement). This would be plausible if the set of patients leaving a physician's practice is random, as might be expected if the source of a decline in N is an increase in the supply of otherwise identical physicians. If the growth of managed care is driving the decline in N , patients remaining in the practice after the decline in N are likely to be different in their responsiveness to a change in I than those who left the practice. If, as it seems likely, patients who are first to leave for managed care are those most sensitive to inducement or most averse to aggressive practice styles, the remaining patients are less likely to leave the physician's practice if I is increased, making n_{IN} even more negative than if departures were random.⁶

The quantity of services sold by a physician choosing inducement level I is $n(I;N)X(I)$. Given our assumptions on patients' demand for a physician's services, $n(I;N)X(I)$ will have an inverted U-shape in inducement effort if two boundary conditions are satisfied. First, a physician engaging in no inducement can increase quantity by recommending more aggressive care [$X_I(0) > -n_I(0;N)$]. Second, beyond some point, overtreatment becomes so blatant that the loss of patients more than offsets the greater quantity of services rendered to the remaining patients [\exists an I above which $X_{II}(I) < -n_{II}(I;N)$]. Defining I^* as the inducement effort that maximizes $n(I;N)X(I)$, attempts to induce beyond this point are counterproductive. The remainder of the paper analyzes physicians' responses to the loss of patients when facing such a demand constraint under various assumptions on physician utility.

We maintain MP's assumption that physicians are price-takers, with fees set by third party payers. Thus, physicians attempting to maintain a target income must raise the quantity of services per patient rather than their fees. If the growth of managed care creates downward pressures on fee levels, physicians would have to rely even more heavily on raising quantity in order to maintain their in-

4. Define $I=0$ as the inducement level at which the physician acts as a perfect agent for the patient (i.e., prescribes the amount of care the patient would choose if she were as well informed as the doctor).

5. Since HMOs use fewer doctors per enrollee than the FFS sector, growth in the HMO market share decreases N as a transfer of 1% of patients to HMOs is accompanied by a transfer of fewer than one percent of physicians.

6. Such patient selection effects could also affect X_I . If patients remaining in FFS medicine are relatively easy to induce (or have a relative preference for aggressive medicine), X_I may rise as N falls. For notational convenience, we posit that selection effects operate only through n_{IN} (i.e., $X_{IN} = 0$).

comes. The physician chooses a level of inducement to solve:⁷

- (1)
$$\max_I U(\pi, L, I)$$
- (2) subject to $\pi = mn(I; N)X(I)$
- (3) and $L = 24 - tn(I; N)X(I)$

Equation (2) represents net practice income, with m denoting the margin between price and the constant costs of inputs other than physician time and inducement. Equation (3) characterizes the physician's time constraint, with t representing work per unit of service.⁸

Substituting for π and L in the utility function simplifies the physician's problem to:

$$(4) \max_I U[mn(I; N)X(I), 24 - tn(I; N)X(I), I]$$

Dropping the functions' arguments, the first order condition is:

$$(5) (nX_I + n_I X)(U_{\pi}m - U_L t) + U_I = 0$$

The first order condition requires that the marginal utility of income generated by the last unit of service equals the marginal disutility of inducement and foregone leisure. The terms inside the first set of parentheses equal zero when the quantity of services, nX , is maximized.⁹ If physicians induce the income maximizing quantity of services ($I = I^*$), the first order condition cannot hold if $U_I < 0$.

7. Physician disutility is based on inducement per patient rather than on the total amount of inducement (number of patients times inducement per patient). Thus, disutility depends on practice style. The qualitative results would be similar if disutility were a function of the total amount of induced services.

8. If the physician must spend time convincing patients to consent to aggressive treatment, inducement could be modeled as a direct cost, with I appearing in the time budget rather than as an argument in the utility function.

9. nX_I is the number of patients times the rate of increase in services per patient as inducement rises. $n_I X$ is the rate at which patients leave the inducing physician times the quantity of services per patient.

Thus, it must be the case that $nX_I + n_I X > 0$ which implies $I < I^*$. When inducing demand is distasteful, the physician will not induce the income maximizing level of demand because as $I \rightarrow I^*$, the extra income generated goes to zero and, hence, does not generate sufficient additional utility to offset the marginal disutility of raising inducement. If $U_I = 0$, physicians can be at a corner solution with $I = I^*$. We will return to this case below.

The comparative static of interest is the effect of a change in the number of FFS patients per physician, a likely result of the increased presence of managed care, on inducement. The comparative static can be found by differentiating equation (5) and applying the implicit function theorem:

$$(6) I_N = \frac{-[(nX_I + n_I X)(U_{\pi\pi}m^2 n_N X + U_{LL}t^2 n_N X) + (n_N X_I + n_{IN} X)(U_{\pi}m - U_L t)]}{(nX_I + n_I X)^2 (U_{\pi\pi}m^2 + U_{LL}t^2) + U_{II} + (nX_{II} + 2n_I X_I + n_{II} X)(U_{\pi}m - U_L t)}$$

When $I < I^*$, $nX_I + n_I X$ must be positive. Thus, first order condition (5) can hold only if $(U_{\pi}m - U_L t) > 0$. Given the sign assumptions, the denominator of equation (6) is negative, satisfying the second order condition for maximization.

The numerator's sign is ambiguous. The first set of terms inside the brackets is negative when $I < I^*$, but the second set takes on the ambiguous sign of $(n_N X_I + n_{IN} X)$, which represents the change in the slope of the relationship between the total quantity of services (nX) and inducement (I) with respect to a change in the patient to physician ratio $[\partial(\partial nX / \partial I) / \partial N]$. If the relationship between nX and I simply shifts up or down with N (i.e., slope is independent of N), then $(n_N X_I + n_{IN} X) = 0$ and the second set of terms vanishes. In this case, I_N is negative implying that a decline in practice size leads physicians to induce more demand from their remaining patients.

I_N will only be positive if $(n_N X_I + n_{IN} X)(U_{\pi}m - U_L t)$ is positive and large enough to offset the negative terms. $n_N X_I$ is indeed positive reflecting the greater increase in services generated by raising inducement effort in

larger practices. However, $n_{IN}X$ is a patient selection term that, as noted earlier, is expected to be negative under managed care growth. Even if $(n_N X_I + n_{IN} X)$ is positive, the comparative static can be unambiguously signed if the disutility of inducing demand is sufficiently low because the first order condition (5) can hold as the marginal disutility of inducing demand approaches zero only if the ambiguous term $(n_N X_I + n_{IN} X)(U_{\pi} m - U_I t)$ also approaches zero. Thus, physicians induce more aggressively when facing exogenous declines in practice size under most scenarios.

We first summarize the intuition of the comparative static (6) in the general case and then consider three special cases (income targeting, profit-maximization, and physicians who do not dislike inducing demand). As the patient to physician ratio declines, income and workload fall. This raises marginal utility of income and decreases marginal utility of leisure, leading physicians to want to induce demand more aggressively. The fall in N also affects physicians' ability to generate demand. $nX_I + n_I X$ represents marginal returns from inducement (slope of total services with respect to I). For example, if this slope rises as N declines, any given increase in inducement generates more additional services than it would have when practice sizes were larger, enhancing the desirability of inducement. For physicians whose $U_I < 0$, there is a further effect of a change in this slope. If marginal productivity of inducement rises, a given amount of extra dollars can be generated with less extra inducement and, hence, less loss of utility.

This illustrates the distinction between the effects of fee reductions and decreases in practice size. MP show that if the payer lowers the fee, inducement rises if income effects are strong and falls when they are weak. In comparison, equation (6) demonstrates that unless the disutility of inducing demand is very high or the opportunities for inducement substantially worsen, a reduction in the number of patients per physician leads to more aggressive care *regardless of the strength of the income effect*. This implies that neither the target income hypothesis nor the existence of strong income effects is necessary to show that doctors respond to certain market conditions (such as the loss of patients due to rising

managed care penetration) by providing more services per patient.

Special Cases

Income Targeting Physicians. Under the target income hypothesis, marginal utility of income is very high below the target and very low above it ($U_{\pi} \rightarrow -\infty$ as income approaches the target). Thus, only terms multiplied by U_{π} matter and equation (6) simplifies immensely to:

$$(7) \quad I_N = \frac{-n_N X}{nX_I + n_I X}$$

To confirm that equation (7) corresponds to the intuition behind the target income hypothesis, consider the supply elasticity of the total quantity of services (nX) with respect to a change in average FFS practice size (N). This elasticity can be expressed as:

$$(8) \quad \frac{\partial nX}{\partial I} \frac{\partial I}{\partial N} \frac{N}{nX} = - \frac{(n_I X + n_N X)n_N X N}{(n_I X + n_N X)n x} = - \frac{n_N N}{n}$$

In a symmetric equilibrium, $n = N$ (each physician has the average practice size) and $n_N = 1$ (shocks to average practice size affect each physician equally). Thus, the elasticity of services supplied by physicians following the TIH is -1 . The physician exactly offsets the impact of a decline in the FFS patient to physician ratio as long as $I < I^*$.

Profit Maximizing Physicians. With no income effects ($U_{\pi} = 0$), physicians maximize profits net of all costs, including the opportunity cost of time and the "cost" of inducing demand. Equation (6) simplifies slightly to:

$$(9) \quad I_N = \frac{-[(nX_I + n_I X)U_{LL} t^2 n_N X + (n_N X_I + n_{IN} X)(U_{\pi} m - U_I t)]}{(nX_I + n_I X)^2 U_{LL} t^2 + U_{II} + (nX_{II} + 2n_I X_I + n_{II} X)(U_{\pi} m - U_I t)}$$

This expression will also be negative unless $(n_N X_I + n_{IN} X)$ is positive and large. Comparing

equations (7) and (9), inducement generally responds more strongly to a change in N under the TIH [e.g., if $(n_N X_I + n_{IN} X) = 0$]. Profit maximizing doctors attenuate, but do not eliminate, the effects of a decline in patient volume on their incomes. TIH physicians also experience disutility from inducing demand, but their behavior is determined solely by their devotion to achieving their income targets. It is possible, but unlikely, that profit maximizing physicians raise inducement even more than TIH physicians in response to a decline in N . This can occur if patient selection effects are so strong that inducement effort becomes much more productive as N declines. In this scenario, TIH physicians stop once their targets are achieved, but profit maximizers might find it worthwhile to induce demand beyond this point.

No Disutility of Inducing Demand. If the physician experiences no disutility from inducing demand ($U_I = 0$), equation (5) becomes:

$$(10) \quad (nX_I + n_I X)(U_{\pi} m - U_L t) = 0$$

This first order condition could be satisfied in two ways. If $U_{\pi} m - U_L$ remains positive when $nX_I + n_I X$ falls to zero (that is, when the quantity of services, nX , is maximized by setting $I = I^*$), the physician induces as much demand as possible. The physician would like to provide even more services at the existing fee (marginal utility of income exceeds that of foregone leisure). Alternatively, the physician will not maximize workload ($I < I^*$) if $U_{\pi} m - U_L t$ falls to zero while $nX_I + n_I X$ remains positive. In this case, the marginal utility of income generated by the last unit of service equals the marginal utility of foregone leisure and the physician does not fully exploit her ability to induce demand.

To assess the effects of a change in N on inducement, first consider the $I < I^*$ case. The comparative static becomes:

$$(11) \quad I_N = \frac{-(nX_I + n_I X)(U_{\pi} m^2 n_N X + U_{LL} t^2 n_N X)}{(nX_I + n_I X)^2 (U_{\pi} m^2 + U_{LL} t^2)} \\ + \frac{(n_N X_I + n_{IN} X)(U_{\pi} m - U_L t)}{(nX_{II} + 2n_I X_I + n_{II} X)(U_{\pi} m - U_L t)}$$

Unlike equation (6), there is no U_{II} term in the denominator and both the numerator and the denominator contain terms multiplied by $(U_{\pi} m - U_L t)$, which must equal zero for the first order condition to hold in the $I < I^*$ regime. Once these terms are eliminated, several others cancel, leaving:

$$(12) \quad I_N = \frac{-n_N X}{nX_I + n_I X}$$

Since this expression is identical to that derived for TIH physicians (7), there is no sign ambiguity. Not only is the comparative static always negative, but the strength of response is independent of whether physicians attempt to maximize profits or attain a target income. Physicians attempt to fully offset declines in patient volume by providing more services per patient regardless of whether or not the target income hypothesis is true.

This striking result is actually quite intuitive. The physician's initial choice incorporates her optimal income/leisure tradeoff. At a given fee and without disutility of inducing demand, the physician can remain at this optimal income/leisure bundle by varying inducement as long as $I < I^*$. Thus, the behavior of a non-TIH physician facing a shrinking practice is indistinguishable from income targeting. The physician is effectively insulated from competition as utility in the fee-for-service sector is independent of increases in managed care penetration or physician supply. When $I < I^*$, increases in managed care penetration will not affect the attractiveness of FFS medicine if physicians do not dislike performing unnecessary services.

If N continues to decline, the physician's ability to maintain income will eventually be threatened. This occurs regardless of whether the original income level arises from an explicit target or simply from the utility maximizing income/leisure tradeoff. Once the physician exerts the inducement effort I^* , the ability to offset shocks to practice size by practicing more aggressive medicine has been exhausted. The FFS physician can do no better than to remain at I^* despite her declining income and utility, allowing HMOs to attract physicians on more favorable terms and affecting physicians' incentives for specialty and location choice.

Graphical Interpretation of TIH and Profit Maximization Cases. We have shown that physicians generally respond to a decrease in practice size by recommending more services per patient. This process is illustrated in Figure 1. The decline in N from N_0 to N_1 shifts down the inducement/income curve.¹⁰ An income targeting physician with target Y_T (or a profit-maximizing physician with no disutility of inducement and an initial income of Y_T) increases inducement from I_0 to I_1 , insulating income and workload from the decline in patients per FFS physician. If N falls to N_2 , the desired income can be achieved by selecting inducement level I^* , which maximizes the quantity of services, nX . Any further declines in N cannot be offset by more intensive treatment. At N_3 the physician can do no better than to keep inducement at I^* . The first order condition (6) no longer holds as an equality as the physician is at a corner solution. Thus, declines in N beyond some level cause FFS incomes to drop even if physicians follow the TIH. Market forces must eventually affect incomes.

For profit maximizing physicians who dislike inducement, there is no threshold below which FFS incomes do not fall with N (unless patient selection effects are so strong that inducement becomes much more productive as N declines). The physician partially offsets the loss of patients by increasing service intensity to remaining patients. In Figure 1, a profit-maximizing physician with initial income Y^T when $N = N_0$ will not raise inducement all the way to I_1 if N falls to N_1 . As in the TIH case, N eventually declines enough that the optimal inducement rises almost to I^* and income falls at a more rapid rate as N declines further. That is, the effects of managed care on incomes accelerate beyond some level of HMO penetration.

Allowing FFS Physicians to See HMO Patients

While physicians employed by group or staff model HMOs exclusively treat HMO pa-

tients, doctors participating in other types of managed care plans usually serve both managed care and FFS patients. Thus, physicians must decide whether or not to participate in managed care plans and how many managed care patients to see. The model can be extended to cover these decisions. Redefine profits as:

$$(13) \quad \pi = mn(I;N)X(I) + dh$$

and leisure as:

$$(14) \quad L = 24 - t[n(I;N)X(I) + h]$$

The physician sees h managed care patients, receiving net income of d per patient. It is assumed that $d < m$ (HMO patients are less lucrative than FFS patients) and the quantity of services provided per HMO patient is normalized to one. d could be interpreted as either net income from a capitation payment or as a discounted fee-for-service rate paid by the HMO. Under capitation, the HMO must ensure that the physician does not "disincentive" demand to such an extent that enrollees are unhappy with their care. In the discounted FFS interpretation, the HMO must monitor utilization to prevent inducement. Since the HMO controls access to its enrollees, it can penalize a physician for deviations. In both cases, the one unit of service per patient normalization can be interpreted as the outcome of the HMO's monitoring of practice patterns.

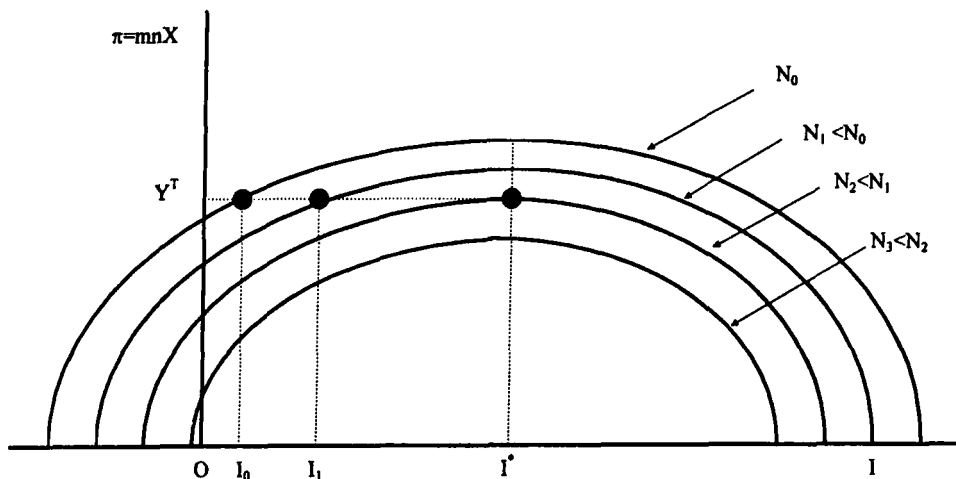
The physician's problem is to maximize utility with respect to both inducement of FFS patients (I) and the number of HMO patients seen (h). Except for the definitions of profit and leisure given by equations (13) and (14), the first order condition with respect to I is identical to equation (5). The first order condition with respect to h is:

$$(15) \quad U_{\pi}d - U_Lt = 0$$

This condition sets the marginal utility of net income generated by serving an additional HMO patient equal to the marginal utility of foregone leisure. Thus, equation (15) holds as an equality only if the physician is willing to treat HMO patients.

10. For ease of exposition, a change in N shifts the curve up or down without changing its slope ($n_N X_I + n_N X = 0$).

FIGURE 1



Starting with the no disutility of inducement case, the physician sees only the more lucrative, FFS patients until $I = I^*$. If utility is maximized at an $I < I^*$, the physician can attain her preferred workload without seeing any HMO patients. If FFS volume is maximized ($I = I^*$), the physician must decide whether or not to serve HMO patients. Since $d < m$, the marginal utility of the income generated by serving the first HMO patient might be less than the marginal utility of foregone leisure. If this occurs, the physician is at a "notch": she would willingly serve another FFS patient, but does not find it worthwhile to serve any HMO patients at the discounted rate. Alternatively, if marginal utility of managed care income exceeds the marginal disutility of foregone leisure, the physician accepts managed care patients until equation (15) is satisfied.

Physicians who dislike inducing demand are more likely to see HMO patients. While the FFS sector offers a higher fee, the managed care incentive and monitoring structure reduces the financial rewards for inducing demand. If d is close enough to m , the physician

may agree to see HMO patients before $I = I^*$ (i.e., despite unexploited FFS inducement opportunities) because of lower marginal inducement costs under managed care. At any reimbursement level, managed care will be more attractive to physicians the greater their aversion to overprescribing care.

IV. ENDOGENIZING MANAGED CARE MARKET PENETRATION

To this point, we have treated changes in managed care penetration and the associated changes in the physician/population ratio in the fee-for-service sector, as exogenous. Clearly, this is a strong assumption. It is beyond the scope of this paper to characterize an equilibrium including endogenously determined FFS premiums, managed care premiums and managed care market share. However, a simple, descriptive model of managed care market penetration reveals how changes in the exogenous determinants of managed care market share can be amplified (or dampened) by adjustments in the relative premiums in the two sectors.

Managed care market share (s) is a function of local patient and physician preferences for managed care (j), other area specific factors such as income, unionization, or firm size (x), FFS premiums relative to managed care premiums (r), and exogenous factors that vary over time (k).

$$(16) \quad s = g(j, x, r, k)$$

Cross sectional variation in managed care penetration rates can largely be explained by variation in the x 's or j 's. Growth in managed care penetration over time can be related primarily to changes in k . Because this paper focuses on how the physician labor market is affected by changes in managed care market share, we focus on k . What variables might be included in k ? Baumgardner [1991] suggests that certain types of technical change may lead to growth in demand for managed care. Other possibilities include changes in employer practices regarding subsidization of employee health care costs or changes in regulatory environment.

Whatever the cause behind growing managed care market share, this trend should reduce the patient/physician ratio in the FFS sector. As demonstrated earlier, a decline in FFS practice size would be expected to alter service intensity and physician incomes. These responses are likely to change the FFS premium relative to the managed care premium, leading to a further change in managed care penetration.

FFS premiums are a function of prices and quantities of services rendered to FFS patients. Quantities depend on the degree to which physicians induce demand. We refer to this as the inducement effect and expect it to raise FFS premiums relative to managed care premiums. On the price side, we expect FFS incomes to fall in response to higher managed care penetration, at least after some threshold penetration rate is reached. This would allow managed care plans to secure physician services at a lower cost which would further raise the relative cost of FFS coverage. Of course, if greater competition for patients causes FFS fees to fall, at least part of this effect would be offset. We denote these forces as price effects and cannot determine, *a priori* how they would impact relative premiums, though we

anticipate they would work to the advantage of managed care plans able to exercise monopsony power in physician markets.

FFS premiums also depend on the mix of patients and physicians left in the FFS sector. Premiums are influenced by the average health status of enrollees, the average susceptibility of enrollees to inducement, the average propensity of physicians to induce demand and physicians' and patients' preferences for aggressiveness of care. We call these phenomenon selection effects. Relative FFS premiums would rise if certain selection effects dominate (e.g., if managed care plans attract physicians who are averse to aggressive care, which could raise aggressiveness of care in the FFS sector even if the inducement patterns of individual physicians do not change). However, relative FFS premiums would fall if other selection effects dominate (e.g., if new managed care enrollees are sickly compared to prior enrollees or if those patients most vulnerable to inducement leave the FFS sector). Thus, the net impact of selection on relative FFS premiums remains ambiguous.

Given the variety of inducement, price and selection effects, reasonable arguments can support the positions that the expansion of managed care will increase or decrease relative FFS premiums. This ambiguity indicates that an increase in managed care penetration arising from a change in some exogenous factor may be amplified or offset by changes in the relative FFS premiums. The degree of amplification or offset is determined by the distribution of enrollee preferences. If a portion of the population has a strong aversion to managed care, increases in relative FFS premiums spawned by increased in managed care penetration may not translate into sizable further increases in managed care penetration. Alternatively, if most consumers are very sensitive to premium differentials, exogenous changes in managed care penetration rates could lead to a spiral in which FFS medicine virtually disappears. Given the increasing proportion of employers who are requiring employees to pay at the margin when choosing more expensive insurance plans, sensitivity to premium differences is likely to be increasing.

Despite the uncertain net result of these various forces, we believe that the fundamental conclusions of our model remain valid.

Under any scenario in which the growth of managed care decreases the average FFS practice size, physician incomes will come under increasing pressure, raising the likelihood that ability to induce demand becomes exhausted.

V. DISCUSSION

The non-linear effect of managed care penetration on physician incomes identified in the model has several implications for empirical research and policy. The apparently delayed response of specialty and location patterns to changes in the marketplace is neither surprising nor does it imply that physicians will not face financial strains in the future. Physicians would have to possess tremendous unexploited capacity to induce demand if they are to remain insulated from market forces. If half of the U.S. population were enrolled in HMOs, payments to FFS physicians would have to double maintain current income levels (Reinhardt [1994]).

Our model suggests that there can be two fundamentally different labor market regimes.¹¹ In a fee-for-service dominated regime, physicians partially or fully offset small shocks to practice size by inducing demand. In this regime, historical relative and absolute specialty incomes can largely be maintained, and the labor market does indeed seem dysfunctional as physician behavior prevents the market from sending the "proper" signals. Conversely, in a managed care dominated environment, FFS physician earnings fall in oversupplied specialties and locations as demand inducement opportunities are exhausted.¹²

The acceleration of the impact of managed care on FFS physician income appears robust to changes in the model's assumptions. The

basic result that doctors' ability to offset the effects of a declining patient base becomes exhausted after some point holds regardless of whether physicians act as profit maximizing firms or follow the target income hypothesis.

These findings highlight the difference between empirical tests of the TIH based on fee changes and those based on changes in the patient to physician ratio. In the MP model, the response of quantity to a fee change provides a clean test of the TIH vs. profit maximization since the comparative static has opposing signs under the two hypotheses. For a change in FFS practice size, distinguishing the hypotheses is more difficult. Physicians are expected to engage in demand creation under either hypothesis. Further, if doctors have no aversion to inducing demand, the magnitude as well as the sign of the change in inducement is independent of whether or not physicians follow the TIH.

Many health workforce experts do not anticipate vigorous, market-driven changes in the size and composition of the physician workforce. These beliefs have led to a variety of regulatory proposals. It is ironic that interest in intervening in the physician labor market has been revived at a time when it is becoming likely that market pressures against specialization and location in over-served areas will rise dramatically. As insurers acquire responsibility for organizing and managing the delivery of health care rather than merely financing it, conditions will change substantially in input markets such as the physician labor market. As physician earnings change, the desired reallocations may occur without policy actions.

Various market-driven adjustments are likely to occur. HMOs would have an incentive to substitute toward physician labor relative to nurse practitioners and physician's assistants. Further, HMOs can only expand if they attract patients who did not join managed care plans in the past. Since these patients are likely to be more concerned than early HMO enrollees with access to specialists, HMOs hiring more specialists than current norms will be better positioned to gain enrollment. The net result is that current staffing patterns cannot be extrapolated to a managed care-dominated market with different absolute and relative wages.

11. Interestingly, physicians have expressed views consistent with this dual regime notion: "A sponge will absorb liquid without a change in the rate of absorption until very near capacity, at which time there is a rapid falloff and cessation. All the additional fluid is excess. The market for physician services demonstrates a similar saturation process" (Michael J. Scotti, M.D., as quoted in Mitka [1997]).

12. Fee setting ability may have contributed to physicians' ability to maintain their incomes despite rising competition. Using data from the 1970s (before aggressive price setting by third party payers), Pauly and Satterthwaite [1981] found that prices rose with the physician to population ratio. In the current environment, even if physicians retain some fee-setting ability, upward flexibility on fees may further erode the competitive position of FFS insurance.

Likewise, existing and prospective physicians will respond to changing practice environments. Specialty and location decisions of new physicians are likely to be particularly fluid. For older physicians, lower earnings and distaste for the loss of autonomy associated with managed care would encourage retirement. Other physicians would relocate to underserved areas, change specialties, or accept non-clinical positions. Likewise, many internal medicine or pediatric subspecialists have appropriate training to provide primary care.

Even if physicians and HMOs respond to changing demand conditions, the labor market would adjust with a lag due to fixed costs of changing specialties and locations. Therefore, some policy initiatives may still be justified.¹³ Programs to encourage location in rural or low income areas may remain worthwhile. Unless health care purchasing power rises in these areas, the market may continue to provide fewer health care resources than society desires. Also, federal funding of graduate medical education rewards the creation of residency positions, providing hospitals with a pool of subsidized labor. Reinhardt [1994] suggests that teaching hospitals should be financed on a "break even" basis to eliminate these incentives.

Recent anecdotal evidence is consistent with our model (Borzo [1994]):

At the end of 1992, 83% of Aetna's enrollees working for small and midsize employers in the New York City area were covered by indemnity plans. By September 1993,

84% were in managed care plans. Just a year ago, insurers in New York were complaining that physicians would not join their managed care networks. Today, physicians—particularly specialists—are complaining that they can't sign up because many panels are filled.

The resistance of New York physicians to managed care may indicate that utility in the FFS sector was insulated from small levels of managed care penetration. However, as rising health care costs made managed care more attractive to consumers and employers, the loss of patients appears to have become increasingly troublesome to FFS doctors.

Recent data are beginning to confirm anecdotal reports of declining incomes in many specialties and a narrowing of the earnings gap between specialists and primary care physicians (Simon and Born [1996]). Likewise, medical students' shift from several formerly desirable specialties to primary care was remarkably rapid. Between 1991 and 1996, the number of medical students applying for family practice residencies rose by 73% while applications in anesthesiology and diagnostic radiology fell by 82% and 33%, respectively (National Resident Matching Program [1991–1996]). If such forces continue to operate on a system-wide basis, the desired changes in the physician labor market will arise without government intervention. Tracking the effects of managed care on the physician labor market is an important area for future research. The possibility that these effects become stronger as managed care penetration rises should be explicitly considered.

13. Regulatory approaches, which often focus on entry restrictions, face analogous criticisms about delayed effects. Further, to the extent that such restrictions lead physicians to expect less competition in the future, they may even discourage existing physicians from changing specialties or locations.

REFERENCES

- Baumgardner, J. R. "The Interaction Between Forms of Insurance Contract and Types of Technical Change in Medical Care." *RAND Journal of Economics* 22, 1991, 36-53.
- Borzo, G. "Insurers Race to Bring HMOs to New York." *American Medical News*, 24/31 January 1994, p. 1 & 38.
- Cohen, J. J., and J. S. Todd. "Association of American Medical Colleges and American Medical Association Joint Statement on Physician Workforce Planning and Graduate Medical Education Reform Policies." *Journal of the American Medical Association*, 272, 1994, 712.
- Committee on the U.S. Physician Supply, Institute of Medicine. *The Nation's Physician Workforce: Options for Balancing Supply and Requirement*. Washington: National Academy Press, 1996.
- Cromwell, J., and J. B. Mitchell. "Physician-Induced Demand for Surgery." *Journal of Health Economics* 5, 1986, 293-313.
- Dranove, D. "Demand Inducement and the Physician/Patient Relationship." *Economic Inquiry* 26, 1988, 281-98.
- Dranove, D., and P. Wehner. "Physician-Induced Demand for Childbirths." *Journal of Health Economics* 13, 1994, 61-73.
- Evans, R. G. "Supplier Induced Demand: Some Empirical Evidence and Implications," in *The Economics of Health and Medical Care*, edited by M. Perlman, Halstead Press: New York, 1974, 162-73.
- Formby, J. P., and E. L. Millner. "The Convergence of Utility and Profit Maximization." *Southern Economic Journal*, 1985, 1, 174-85.
- Fuchs, V. R. "The Supply of Surgeons and the Demand for Operations." *Journal of Human Resources* 13, 1978, 35-56.
- InterStudy. *National HMO Census 1985*. InterStudy: Excelsior, Minnesota, 1986.
- _____. *The InterStudy Competitive Edge 5.2*. InterStudy: Minneapolis, Minn., 1995.
- McGuire, T. G., and M. V. Pauly. "Physician Response to Fee Changes with Multiple Payers." *Journal of Health Economics* 10, 1991, 385-410.
- Mitka, M. "Consensus Panel Offers Response to Oversupply." *American Medical News* 40, 24 February, 1997, p. 1 & 39.
- National Resident Matching Program. *NRMP Data*. Association of American Medical Colleges: Washington, D.C., 1991-1996.
- Newhouse, J. P., A. P. Williams, B. W. Bennett, and W. B. Schwartz. "Does the Geographic Distribution of Physicians Reflect Market Failure?" *Bell Journal of Economics* 13, 1982, 365-90.
- Pauly, M. V., and M. A. Satterthwaite. "The Pricing of Primary Care Physicians' Services: A Test of the Role of Consumer Information." *Bell Journal of Economics* 12, 1981, 53-84.
- Pew Health Professions Commission. *Reforming Health Care Workforce Regulation: Policy Considerations for the 21st Century*. University of California at San Francisco: San Francisco, 1995.
- Reinhardt, U. "Planning the Nation's Health Workforce: Let the Market In." *Inquiry* 31, 1994, 250-63.
- Rizzo, J. A., and D. Blumenthal. "Is the Target Income Hypothesis an Economic Heresy?" *Medical Care Research and Review* 53, 1996, 243-66.
- Rochaix, L. "Information Asymmetry and Search in the Market for Physicians' Services." *Journal of Health Economics* 8, 1989, 53-84.
- Schroeder, S. A. "Managing the U.S. Health Care Workforce: Creating Policy Amidst Uncertainty." *Inquiry* 31, 1994, 266-75.
- Simon, C. J., and P. H. Born. "Physician Earnings in a Changing Managed Care Environment." *Health Affairs* 15, 1996, 124-33.
- Wennberg, J. E., D. C. Goodman, R. F. Nease, and R. B. Keller. "Finding Equilibrium in U.S. Physician Supply." *Health Affairs* 12, 1993, 89-103.