Medical Specialists I:
Coordination of Specialists and Patients

Suzette Baker
University of North Carolina at Chapel Hill
Department Of Economics

December 5, 2008

Abstract

Without a flow of "treatable" patients, a medical specialist won't be able to do his job. Many specialists rely on referring doctors to ensure that flow. I develop a model that demonstrates which doctors are likely to specialize, which doctors are likely to refer, and which doctors are likely to treat without a referral. I show that the introduction of more specialists – and the corresponding need for more referring doctors – can reduce the overall number of health care providers actually treating patients. Finally, I compare the socially optimal and joint profit maximizing (1) quantity of specialists, (2) price of speciality services, and (3) price of generalist services. I find that, when doctors collectively set prices for both specialist and generalist treatment, the welfare ramifications differ from the textbook monopoly model.

1 Introduction

Specialization by physicians has transformed the way medical care is provided. By focusing on a single illness or injury, a doctor obtains a deep knowledge about that ailment and is able to keep abreast of the latest research. And, not surprisingly, many studies show that doctors with more knowledge provide higher quality care (see, e.g., Clark et al., 1997; Carbona et al. 2006). The problem, however, is that in order for patients and doctors to get the benefits of specialization, they need to be correctly matched. The question I address is: What is the optimal amount of medical specialization, taking into account the cost of coordinating patients and physician-specialists.[1]

[1] Starting shortly after World War II, the United States began to see a rise in doctors who specialize in treating, say, one type of cancer in children or a specific problem related to ligament damage in the knee (Donini-Lenhoff and Hedrick 2000). Since then, the provision of medical services has become increasingly fragmented, with doctors treating a smaller sliver of the range of possible illnesses and injuries. In 2007, 86 percent of graduating medical students planned to become certified in a speciality or subspeciality (Association of American Medical Colleges 2007).

The existing literature on specialization focuses on costs and benefits of team production, where each team member has a different skill essential to production of a good (Jones 2008, Becker and Murphy 1992, Alchian and Demestz 1972). I study a different coordination problem here – the matching of patients with a specific illiness with a doctor who specializes in that illness. This problem shares some similarity with the economic literature analyzing the role of middlemen in facilitating consumer choice in other market contexts (see, e.g., Rubinstein and Wolinsky 1987). But the results from this literature do not translate well into doctor/patient matching. Unlike consumers in other contexts, patients do not know what they want. Their goal is to "become healthy," but they don't know how to make that happen. This difference is important and serves as the launching point for my contributions to the literature, which are fourfold.

First, in my model, the coordinating agent – the referring doctor – must have a special skill, the ability to diagnose illnesses. As I explain, this requirement eliminates the risk of mismatch but crowds out treatment. Second, I show that the degree of patient risk aversion determines the optimal amount of medical specialization; the more risk averse patients are, the more specialization. Third, I demonstrate that specialization drives up the wages of all doctors, not just the specialists. The reason is that workers with unequal talents, i.e., those who are excellent at treating one illness and bad at treating other illnesses, perform tasks for which they have a comparative advantage. On the other hand, workers who are equally good at treating all illnesses remain in the general market and provide treatment no matter the injury. This reduces the costs of generalized care and, as a result, patients will spend more for that service.

As prices increase, the patients with the highest cost of seeking care forgo treatment. The optimal amount of specialization balances the cost associated with this patient exit against the gains from each additional patient-doctor match. Finally, and counterintuitively, I show that doctors acting in concert do not necessarily set prices to restrict the amount of specialization services.[2] This result stands in contrast to the standard monopoly model and suggests care in the antitrust treatment of groups of doctors.[3]

Illustrative of referring-doctor-as-middleman is the "Best Doctor, Inc." agency [www.bestdoctors.com]. This agency recruits doctors to help match patients who are facing a very serious illness with the appropriate medical expert.[4] These doctors don't treat patients, but instead serve as an information conduit – a mediator between patients and other experts. A second illustration comes from the behavior of general practitioners. While some practitioners are willing to provide more than generalized care, many are quick to refer. When choosing a

[2] In health care markets, prices are occasionally set by agreement between groups of doctors and insurance carriers (Choudhry and Brennan 2001 p. 1143). In these agreements, individual doctors have little control over the price of services (McGuire 2000; p. 481). With no control over price, McGuire suggests that physicians exercise their market power by increasing the quantity of services above what the patient prefers. See also Ma and McGuire (2002).

[3] On the antitrust concerns unique to doctors, see Gaynor and Vogt (2000).

[4] According to their marketing materials, Best Doctors "gives members insight and information about their diagnosis, the latest advances and where they can turn for state-of-the-art care when faced with a serious medical problem." See www.bestdoctors.com.

primary care physician, patients decide whether they want a "true" generalist who treats most illnesses or a chief provider who oversees and coordinates, while specialists treat. Depending on the size of the local market and the frequency of malpractice claims, true generalists may be crowded out by chief providers and specialists.

The paper relates to the matching model literature (see, for example, Diamond 1982 and Mortensen 1982). In those models, workers match with firms. A successful match creates a surplus, which can be divided between the two parties. One question is what wage rate, if any, ensures a stable equilibrium. The equilibrium number of job vacancies and the level of unemployment is then compared with the ones a social planner would pick. Demange and Gale (1985) and Gale and Shapley (1962) study two-sided matching models. In these models, an outcome is stable if no two parties from opposite sides of the market can gain by deviating and forming a different partnership. The model here involves matching of patients with doctors. The equilibrium concept is similar. Given a set of prices, an equilibrium exists if no doctor and no patient want to change positions.

A second set of related literature concerns two-sided markets (Rochet and Tirole 2002; Rochet and Tirole 2006). In these markets, a platform provider must ensure that both consumers and producers use the service. With credit cards, for example, issuers must consider how the interchange transfer fee will affect the merchants' propensity to accept the card for purchases and the consumers propensity to use the card. In the leading article on the subject, Rochet and Tirole (2002) demonstrate that issuers might set a fee that results in the overuse of credit cards as compared to the social optimum. In their model, merchants too readily accept cards. The reason is that merchants make the decision whether to accept credit cards anticipating that they will service the average card user, not the marginal user. The average user will attach a higher benefit to card use and, as a result, be willing to pay more for the convenience of using his card. Praying on the merchant's eagerness to attract card-carrying customers, issuers set a higher than optimal transfer fee.

In the model developed here, there exists a similar potential for an overprovision of services, specifically medical specialist services. This problem occurs when doctors collective set prices. The expansion of specialist services decreases the price of specialist services, while increasing the price of generalist services. The price is based on the average benefit a patient receives from treatment, rather than the marginal benefit. As a result, under some conditions, the boost in the price of generalist services more than compensates for the decrease in the price of speciality services, making the expansion of speciality services attractive.

A third related strand is the wide-ranging work on labor specialization and investment in human capital. For example, Kim (1989) considers a situation where workers can invest in both the depth and breath of human capital. As market size increases, workers want to deepen their specific skill set, rather than increase the number of tasks they are capable of doing. The reason is that a large market contains more employers. With more employers, there is greater chance the worker will be matched with an employer who values – and therefore rewards – the deep specific skill set. Along similar lines, Baumgardner (1988) looks at the division of labor within service industries. He shows a trade-off between increasing returns to production in each activity and decreasing marginal revenue. A more narrowly-focused worker is better at a specific task, allowing him to charge higher prices. But specialization has a downside – fewer potential customers. The optimal degree of specialization trades off the gains from specialization against the losses from weaker demand. Finally, as Bolton and Dewatripoint (1994) and Becker and Murphy (1992) point out the amount of specialization depends heavily on the cost of coordination among specialists: the more specialists, the greater the coordination costs and the lower the net return from an additional specialist.

Part 2 develops the model. Part 3 explores the level of specialization that maximizes welfare. Part 3.1 considers the impact of patient risk aversion on the optimal amount of specialization. Part 3.2 considers corner solutions, where welfare is maximized by having either no specialists or as many specialists as feasibly possible. In part 4, doctors set prices to induce a certain number of referring doctors, specialists, and generalists in equilibrium. Depending on the parameter configurations, doctors set prices to induce too much or too little specialization as compared to the social optimum. Part 5 concludes.

2 The Model

Patients are denoted by j. Each patient represents a point on the continuum $[0, n]$. Patients are equally likely to have a type 1 illness or a type 2 illness. Over the entire continuum half the patients have a type 1 illness, half the patients have a type 2 illness. Patients do not know their illness type. Given this uncertainty, each patient decides whether to (1) seek care from a generalist doctor; (2) go to a referring doctor, have their illness identified, and be rerouted to a medical specialist, or (3) forgo treatment. Assume that patients cannot see a specialist without first getting a referral.

To understand the meaning of a "generalist" doctor, consider a patient with chest pain. The patient can go to a generalist, a primary care physician who has a reputation for solving problems immediately. This doctor has devoted his resources to treating patients with a broad range of illnesses. So, his initial action is to treat the patient as opposed to refer to a specialist. The doctor might run tests and prescribe a better diet, exercise, vitamins, and medicine. That treatment would be beneficial and provide some relief. However, this treatment might not be drastic enough. If instead the patient went to a primary care physician known for referring patients, that doctor would have sent him to a leading cardiologist. The cardiologist might have considered a more aggressive and newer treatment option.

Denote doctors by i. Like patients, doctors fall on the continuum $[0, n]$. Doctors can treat one patient per period. In the time it takes to do a treatment, a doctor can refer two patients to specialists. That is to say, referrals take half the time of an actual treatment.

The benefit of medical treatment depends on the doctor's skill level. If the patient has a type 1 illness, $\theta(i)$ is the patient's benefit from treatment by doctor i. If the patient has a type 2 illness, $\Psi(i)$ is the patient's benefit from treatment. Doctors have the same amount of human capital, spread differently between the two types of illnesses. Some doctors are equally good at treating both illnesses. Other doctors have a high ability in treating one illness and a low ability in treating the other illness. A doctor cannot have a high ability in treating both illnesses. To capture differences in ability, I assume that the benefit function for illness one, $\theta(i)$, is linear and decreasing in the doctor index. The benefit function for illness two, $\Psi(i)$, is linear and increasing in the doctor index. The benefit functions are symmetric ($\theta(1) = \Psi(n)$).

Figure 1 represents the patient's benefit associated with treatment by each doctor i. The vertical axis represents the patient's benefit from treatment for illness 1, $\theta(i)$, and for illness 2, $\Psi(i)$. The horizontal axis represents the doctors, indexed from 0 to n.

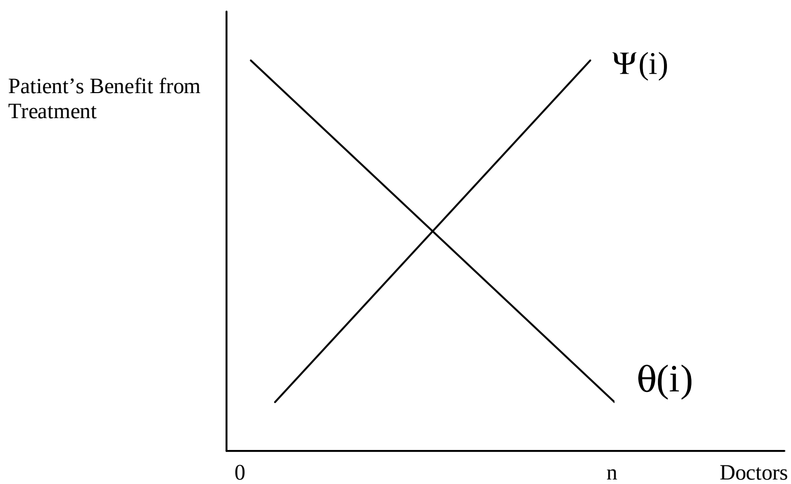

Figure 1: The Relationship Between Doctor Skill and Patient Benefit

2.1 Patient Payoffs

Patients pay an out-of pocket price to the treating doctor. t_e is the price paid after getting a referral and being rerouted to an expert doctor; t_g is the price paid for treatment by a generalist without the referring middleman. In addition to the out-of-pocket expenses, patients face a cost, $k(j)$, per doctor visit. Because of a less debilitating sickness, helpful family members, or a home located in an area with lots of medical services, some patients find it easier to go to a doctor. Patients who reside close to zero on the continuum have a lower cost per visit than patients who reside close to n. More specifically, I assume that the cost function, k, is increasing and linear in j. If a patient sees a generalist, he makes one trip to the doctor. A referral requires two doctor visits – one to the referring doctor and a second to the specialist. The patient's utility from treatment is $v(\cdot)$, where $v'(\cdot) > 0$ and $v''(\cdot) < 0$.

Patients are uncertain which specialist doctor they will be sent to after seeking a referral. Likewise, patients are uncertain which generalist will treat them if they select generalist treatment. They form expectations about these facts by looking at the pool of available specialists and generalists.

If, for example, there are six referring doctors and twelve specialists, the patient seeking a referral anticipates treatment by a doctor with the average skill level among the twelve specialists. Similarly, if there are thirty generalists in the market, a patient seeking care anticipates the care associated with the average generalist among these thirty physicians.

In short, each patient faces a lottery over possible outcomes, where each doctor in the generalist pool or specialist pool has an equal chance of being selected.[5] Generalist treatment involves a compound lottery. First, there is a lottery over which illness the patient has – type 1 or type 2. Second, there is a lottery over the possible outcomes from treatment given the generalist pool.

To sum up, patient $j's$ utility depends on (1) the patient's benefit from treatment, (2) the price paid for medical services; (3) whether she sees a generalist or, via a referral sees a specialist; (4) the skill level of the pool of doctors doing the treatment, and (5) the individualized cost per visit.

Given a pool of type 1 specialists, $[a, b]$, a patient who seeks a referral and is discovered to have a type 1 illness receives expected utility

$$\frac{1}{b-a}\int_a^b v(\theta(i))di - t_e - 2k(j) \tag{1}$$

Given a pool of type 2 specialists, $[c, d]$, a patient who seeks a referral and is discovered to have a type 2 illness receives expected utility

$$\frac{1}{d-c}\int_c^d v(\Psi(i))di - t_e - 2k(j) \tag{2}$$

Finally, given a pool of generalists, $[e, f]$, the patient's expected utility from seeking generalist treatment is

$$\frac{1}{2(f-e)}\int_e^f \left(v(\theta(i)) + v(\Psi(i))\right) di - t_g - k(j) \tag{3}$$

[5] Suppose that the doctor interval $[a, b]$ represents a specialist pool. The probablity of seeing a doctor of skill level x or less is distributed uniformly over that range. The density is the same for each doctor in the interval, reflecting that a patient has an equal chance of seeing each doctor in the pool.

The utilty function shows that patients differ in the "net" benefit from medical treatment. Because $k(n) > k(1)$, patients close to n have a lower net benefit from treatment than patients close to 1.

2.2 Doctor Payoffs

Individual doctors are price takers. Referring doctors diagnose the patient's illness, 1 or 2, and then send patients to specialists. Unlike patients, referring doctors know the skill level of each specialist and route patients to the available specialist with the highest skill level. The specialist cannot bypass the referring doctor and solve the matching problem by signalling their speciality through advertisements or other marketing materials. The patients don't know what illness they have. As a result, they don't know which specialist to see.

Let Φ_r, Φ_g, Φ_e be the probabilities that a doctor has patient demand for his services if he chooses to be a referring doctor, generalist, or specialist respectively. In modeling these probabilities, first consider markets for referring doctors and generalists. In these markets, the probability that a doctor actually has a patient to treat depends on the number of patients and the number of doctors. If the supply of doctors outstrips the demand for doctors, the chance an individual doctor actually sees a patient is less than one, but greater than zero.

To capture this easily, let $\Phi_g = \frac{\text{\# of patients seeking generalists}}{\text{\# of generalists doctors}}$ and $\Phi_r = \frac{\text{\# of patients seeking referrals}}{\text{\# of referring doctors}}$. If, for example, the number of patients in the market for generalist treatment is 80 and the number of doctors is 100, the probability that an individual doctor sees a patient is $\frac{8}{10}$, If the number of patients exceeds the number of doctors in a market, let $\Phi = 1$.

If a doctor enters the referral market, he might refer one patient, two patients, or no patients. Each draw from the pool of patients seeking referrals is independent. The probability a referring doctor sees two patients is $\Phi_r \Phi_r$; the probability a referring doctor sees one patient is $2\Phi_r(1-\Phi_r)$; and the probability he sees no patients is $(1 - \Phi_r)(1 - \Phi_r)$.

The specialist market is different. Because referring doctors know the skill level of the specialist doctors, a specialist doctor will only be routed a patient if his skill level exceeds the skill level of the weakest member of the specialist pool focusing on that illness. What this means is that if the supply of specialists exceeds the demand for specialists, the specialists closest to the middle of the distribution of doctors are referred no patients.[6]

Finally, let F be the fee that the specialist pays the referring doctor out of the payment he receives from the patient, t_e.

[6] Formally, we could denote the least-skilled specialist focusing on illness 1 as $\underline{i}$ and the least-skilled specialist focusing on illness 2 as $\underline{\underline{i}}$. Then, for illness type 1, $\Phi_e = 1$ if $i \leq \underline{i}$ and $\Phi_e = 0$ if $i \geq \underline{i}$ (recall that for type 1 illness doctor 0 is the most skilled and doctor n is the least skilled). For a type 2 illness, it's reversed, $\Phi_e = 1$ if $i \geq \underline{\underline{i}}$ and $\Phi_e = 0$ if $i \leq \underline{\underline{i}}$.

We can now define the doctor's expected payoff concisely as

$$EU = \left\{ \begin{array}{ll} \Phi_e[t_e - F] & \text{if specialist} \\ \Phi_r \Phi_r 2F + 2\Phi_r(1-\Phi_r)F & \text{if referring doctor} \\ \Phi_g t_g & \text{if generalist} \end{array} \right\}$$

2.3 Timing and Initial Results

The timing of the game follows: First, doctors announce prices. Second, each doctor decides what to do: become a specialist, become a referring doctor, or become a generalist. Third, patients decide whether to seek a referral, go to a generalist, or forgo treatment. Finally, all patients seeking care are treated and outcomes observed.

Solve by backward induction. Take prices as given for now. Let s be the number of specialists and $\frac{s}{2}$ the number of referring doctors needed to support those specialists.

Proposition 1 *For any value $s \in [0, \frac{2n}{3}]$ there exists a set of prices $\left\{t_e^*(s), t_g^*(s), F^*(s)\right\}$ such that the following is a Nash equilibrium: (1) Patients in the interval $[0, s]$ seek referrals and treatment by a specialist; (2) Patients in the interval $(s, n - \frac{s}{2}]$ seek treatment from generalists; (3) Patients in the interval $(n - \frac{s}{2}, n]$ forgo treatment; (4) Doctors in the intervals $[0, \frac{s}{2}]$ and $[n - \frac{s}{2}, n]$ specialize; (5) doctors in the interval $[\frac{3s}{4}, n - \frac{3s}{4}]$ provide generalist treatment; and (6) doctors in the intervals $(\frac{s}{2}, \frac{3s}{4})$ and $(n - \frac{3s}{4}, n - \frac{s}{2})$ refer patients.*

Proof:

Step One:

Find the value of t_g^* that makes the patient indexed by $n - \frac{s}{2}$ indifferent between seeing a generalist, who he anticipates to be the average generalist in the market, and forgoing treatment. This value is determined by

$$\frac{1}{n - \frac{3s}{2}} \int_{\frac{3s}{4}}^{n - \frac{3s}{4}} \frac{1}{2} \left[v(\theta(i)) + v(\Psi(i)) \right] di - k(n - \frac{s}{2}) - t_g^* = 0 \tag{4}$$

Step Two:

Assume the last patient seeking a specialist as a type 1 illness. Find the value t_e^* that makes this patient indexed by s just indifferent between seeing the average specialist in the market for type 1 specialists and seeing the average generalist in the market. This value is determined by

$$\frac{1}{\frac{s}{2}} \int_0^{\frac{s}{2}} v(\theta(i)) di - t_e^* - 2k(s) = \tag{5}$$
$$\frac{1}{n - \frac{3s}{2}} \int_{\frac{3s}{4}}^{n - \frac{3s}{4}} \frac{1}{2} \left[v(\theta(i)) + v(\Psi(i)) \right] di - k(s) - t_g^*$$

Step Three:

The cost per visit, $k(j)$, is smaller than $k(s)$ for all $j < s$. As a result, for these patients, the LHS of (5) is bigger than the RHS of (5). This means that, at the price, t_e^*, each patient between $[0, s]$ strictly prefers the average specialist in the pool to the average generalist in the pool. Conversely, at the price t_e^* every patient in the interval $(s, n]$ strictly prefers treatment by the average generalist to the average specialist available in the market.

The cost per visit, $k(j)$, is smaller than $k(n - \frac{s}{2})$ for $j \in (s, n - \frac{s}{2}]$. As a result, for these patients, the LHS of (4) is bigger than the RHS of (4). So these patients do not want to deviate and forgo treatment. And since, as noted above, these patients do not want to deviate and take the specialist treatment, they have no profitable deviation. For patients in the interval $(n - \frac{s}{2}, n]$, the RHS of (4) is bigger than the LHS of (4). As a result, at the price t_g^* , these patients can't deviate and take generalist treatment without being made worse off. And, as shown previously, these patients also do not want to deviate and take specialist treatment, leaving them no profitable deviation.

Step Four:

Moving to doctors, find the value of F that makes the following hold

$$t_g^* \geq 2\Phi_r\Phi_r F + 2\Phi_r(1 - \Phi_r)F \tag{6}$$

$$t_e^* - F \geq 2\Phi_r\Phi_r F + 2\Phi_r(1 - \Phi_r)F \tag{7}$$

$$2F \geq \Phi_g t_g^* \tag{8}$$

$$t_e^* - F \geq \Phi_g t_g^* \tag{9}$$

Equations (6) and (7) ensure that no generalist and no specialist want to deviate and become a referring doctor, given the number of other doctors doing referrals. Equations (8) and (9) ensure that no referring doctor and no specialist want to deviate and become a generalist, given the number of other generalists in the market. As defined earlier, $\Phi_r = \frac{\text{\# of patients seeking referrals}}{\text{\# of referring doctors}}$ and $\Phi_g = \frac{\text{\# of patients seeking generalists}}{\text{\# of generalists doctors}}$. In equilibrium, $\Phi_r = \Phi_g = 1$. In addition, the price taking assumption means that a doctor deviation won't change the market price. The previous four equations therefore reduce to

$$t_g^* \geq 2F \tag{10}$$

$$t_e^* - F \geq 2F \tag{11}$$

$$2F \geq t_g^* \tag{12}$$

$$t_e^* - F \geq t_g^* \tag{13}$$

Equations (10) and (12) can only be satisfied simulteneously if $F^* = \frac{t_g^*}{2}$. Given F^*, equations (11) and (13) hold if $t_e^* \geq \frac{3}{2}t_g^*$, which is true if the utility uptick from specialist treatment is sufficiently large.

A referring doctor or generalist doctor who switched and tried to snag a patient from the specialists would reap no patients. The referring doctor recognizes that the deviating doctor has a lower skill level than every specialist treating that illness and so routes them no patients, making this deviation unprofitable.

Finally, the weakest condition that ensures positive prices for medical services is

$$k(n) \leq \frac{1}{n}\int_0^n \frac{1}{2}\left[v(\theta(i)) + v(\Psi(i))\right] di$$

The lower bound on the price of specialist services is $\frac{3}{2}t_g^*$. The lowest possible value of t_g^* occurs when no doctor specializes. The above inequality ensures this price is positive.

For $\left\{t_e^*, t_g^*, F^*\right\}$ as defined above, no doctor and no patient has a profitable deviation, making this equilibrium with s specialists a Nash equilibrium ■

For any number of patients seeking treatment by medical experts, t_g^*, t_e^*, and F^* ensure that supply equals demand in every market: the number of referring doctors equals one half the number of specialists; the number of patients seeking referrals equals twice the number of referring doctors and the number of patients seeking care without a referral equals the number of generalist doctors. Figure 2 illustrates what the equilibrium looks like.

Within a given market, all doctors make the same amount. Despite differential skill levels among doctors, patients can't observe those differences and, as a result, pay for the "average" benefit associated with a doctor in that market.

Across markets, referring doctors and generalists make the same amount, t_g^*. That must happen to equilibrate the number of doctors in these two markets. If, say, the price of generalist services was greater than the price of referring services, each generalist would switch markets and become a referring doctor.

Specialists make more than generalists or referring doctors (the specialist's net payment, $t_e^* - \frac{t_g^*}{2}$ must be greater than t_g^*). Referring doctors can identify the skill level of specialists. This identification means a generalist or referring doctor who deviated to take advantage of the higher specialist wage would be routed no patients. The specialist, in other words, is compensated for his higher skill level, albeit at the level of the average specialist in his market.

Since referring doctors and generalists have the same payoff, they could switch places and it would still be an equilibrium. In other words, the lineup of doctors in proposition 1 is not the only possible Nash equilibrium. Notice, however, the prices (t_g^*, F^*, t_e^*) depend on the anticipated generalist pool. The price consumers are willing to pay for generalist treatment is highest when the doctors in the middle of the distribution fill the generalist role. Because consumers are risk averse, they are willing to pay more for generalist treatment by a doctor with a lower spread in outcomes.[7] Among the possible lineups of

[7] Formally, this can be seen by noting that the patient's expected utility if matched with doctor i for generalist treatment is

$$\frac{1}{2}v(\theta(i)) + \frac{1}{2}v(\Psi(i)) - t_g - k(j)$$

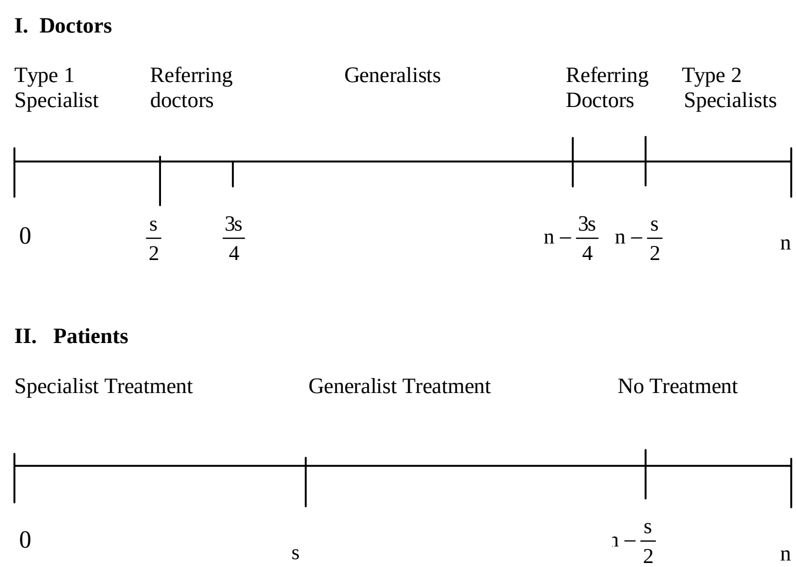

Figure 2: Equilibrium Allocation of Patients and Doctors

doctors with s specialists, the lineup in proposition 1 gives the highest prices for all doctors and hence is a natural one to focus on.

A couple of further points are worth mentioning. First, patients with the lowest cost per visit – those patients, say, near a cluster of doctors– are the most likely to seek referrals. Second, patients with the highest cost per visit – those patient in, say, rural area – are the most likely to forgo treatment. Third, doctors with the highest skill level specialize. Fourth, doctors who aren't very good at any one illness become generalists. Fourth, the doctors who choose to do referrals don't have the high skill level required to specialize, but aren't good enough at more than one illness to become generalists. These doctors do have an important task in the model; they correctly diagnose and refer to the best available specialist. Finally, because they are risk averse, patients will not gamble and go directly to a specialist, rather than see a referring doctor first for a diagnosis.

3 Welfare Analysis

To get a handle on the welfare effects of specialization, turn now to the social planner's problem. How many specialists would a planner want to have? The welfare associated with an equilibrium with s patients seeking referrals is

$$\begin{aligned} W(s) = & \int_{i=0}^{\frac{s}{2}} v(\theta(i))di + \int_{i=n-\frac{s}{2}}^{n} v(\Psi(i))di + ... \\ & \frac{1}{2}\int_{i=\frac{3s}{4}}^{n-\frac{3s}{4}} \{v(\theta(i)) + v(\Psi(i))\}di - ... \\ & \int_{j=0}^{s} 2k(j)dj - \int_{j=s}^{n-\frac{s}{2}} k(j)dj \end{aligned} \tag{14}$$

Because of the symmetry of the benefit functions for type 1 and type 2

The first and second derivatives with respect to i are

$$\frac{1}{2}[v'(\theta(i)]\theta'(i) + \frac{1}{2}[v'(\Psi(i)]\Psi'(i)$$

$$\frac{1}{2}[v''(\theta(i)]\theta'(i)^2 + \frac{1}{2}[v''(\Psi(i))]\Psi'(i)^2$$

Since $v'' < 0$, the second derivative is negative, making the expected utility concave in i. Setting the first derivative equal to zero and solving yields

$$\frac{1}{2}[v'(\theta(i)]\theta'(i) = -\frac{1}{2}[v'(\Psi(i))]\Psi'(i)$$

Since $\theta'(i) = -\Psi'(i)$, this equality holds when $\theta(i) = \Psi(i)$. This occurs for the doctor located at $\frac{n}{2}$. Since generalist treatment by this doctor results in the highest expected utility, the patient is willing to pay the most for it and for treatment by doctors close to it.

illnesses, welfare can be rewritten as

$$W(s) = 2\int_{i=0}^{\frac{s}{2}} v(\theta(i))di + \int_{i=\frac{3s}{4}}^{n-\frac{3s}{4}} \frac{1}{2}\{v(\theta(i)) + v(\Psi(i))\}di$$
$$- \int_{j=0}^{s} 2k(j)dj - \int_{j=s}^{n-\frac{s}{2}} k(j)dj \quad (15)$$

Equation (15) is concave. Setting the FOC equal to zero and solving yields

$$v(\theta(\frac{s}{2})) - \frac{1}{2}\left\{(v(\theta(\frac{3s}{4})) + v(\Psi(\frac{3s}{4})\right\} + \frac{1}{2}k(n-\frac{s}{2})$$
$$= \frac{1}{2}\left\{\frac{1}{2}[v(\theta(\frac{3s}{4})) + v(\Psi(\frac{3s}{4})]\right\} + k(s) \quad (16)$$

The solution to (16) provides the optimal number of medical specialists, s^W. To increase the number of specialists by one unit requires an additonal 1/2 unit of referral services, leading to the rationing of 1/2 unit of patient care. The left hand side of (16) represents the marginal benefit from adding a specialist. This benefit has two parts: (1) the surplus above what that matched patient would have received from a generalist treatment.and (2) the cost saving from having 1/2 a patient forgo treatment. The right hand side of (16) reflects the marginal cost of adding another specialist. The addition of a specialist means that 1/2 a unit of patients are no longer treated by generalists, resulting in a utility loss. And, the addition of a specialist comes at the cost of a unit of patients making an extra doctor visit.

3.1 Patient Risk Aversion

To explore the role of risk aversion, suppose that each patient's utility function takes the constant relative risk aversion form:

$$v(x) = \frac{x^{1-g}}{1-g} \quad (17)$$

g is the risk aversion parameter. As g goes up, the patient becomes more risk averse. A natural question is how patient risk aversion impacts the welfare maximizing amount of specialization. The next proposition answers this question.

Proposition 2 *As patients become more risk averse, the welfare maximizing amount of specialization increases (that is,* $\frac{\partial s^W}{\partial g} > 0$*).*

Proof:
Plugging in the CRRA function form, rewrite (16) as

$$\frac{\theta(\frac{s}{2})^{1-g}}{1-g} - \frac{3}{4}\left\{\frac{\theta(\frac{3s}{4})^{1-g}}{1-g} + \frac{\Psi(\frac{3s}{4})^{1-g}}{1-g}\right\} +$$
$$\frac{1}{2}k(n-\frac{s}{2}) - k(s) = 0 \quad (18)$$

The implicit function theorem tells us that

$$\frac{\partial s^*}{\partial g} = \frac{-\frac{\partial^2 W}{\partial s \partial g}}{\frac{\partial^2 W}{\partial s \partial s}} \tag{19}$$

Because $W(s)$ is concave, the SOC condition is negative. Accordingly, the sign of $\frac{\partial s^*}{\partial g}$ equals the sign of $\frac{\partial^2 W}{\partial^2 s \partial g}$. We know that

$$\frac{\partial^2 W}{\partial s \partial g} = -\theta(\frac{s}{2})^{-g} + \frac{3}{4}\left\{\theta(\frac{3s}{4})^{-g} + \Psi(\frac{3s}{4})^{-g}\right\} \tag{20}$$

Rewrite $\frac{\partial^2 W}{\partial s \partial g}$ as

$$-\theta(\frac{s}{2})^{-g} + \frac{3}{4}\left\{\theta(\frac{3s}{4})^{-g} + \theta(n - \frac{3s}{4})^{-g}\right\} \tag{21}$$

Because $\theta(\frac{s}{2}) > \theta(\frac{3s}{4})$, we know that

$$\frac{1}{\theta(\frac{s}{2})^g} < \frac{1}{\theta(\frac{3s}{4})^g} \tag{22}$$

$\theta(\frac{3s}{4}) > \theta(n - \frac{3s}{4})$ implies

$$\frac{1}{\theta(\frac{3s}{4})^g} < \frac{1}{\theta(n - \frac{3s}{4})^g} \tag{23}$$

Rewrite (22) as

$$\frac{1}{\theta(\frac{s}{2})^g} < \frac{1}{2}\left\{\frac{1}{\theta(\frac{3s}{4})^g}\right\} + \frac{1}{2}\left\{\frac{1}{\theta(\frac{3s}{4})^g}\right\} \tag{24}$$

Take the RHS of (24). Because of (23), it must be the case that:

$$\frac{1}{2}\left\{\frac{1}{\theta(\frac{3s}{4})^g}\right\} + \frac{1}{2}\left\{\frac{1}{\theta(\frac{3s}{4})^g}\right\} < \frac{1}{2}\left\{\frac{1}{\theta(\frac{3s}{4})^g}\right\} + \frac{1}{2}\left\{\frac{1}{\theta(n - \frac{3s}{4})^g}\right\} \tag{25}$$

Now take the RHS of (25). It must be true that

$$\frac{1}{2}\left\{\frac{1}{\theta(\frac{3s}{4})^g}\right\} + \frac{1}{2}\left\{\frac{1}{\theta(n - \frac{3s}{4})^g}\right\} < \frac{3}{4}\left\{\frac{1}{\theta(\frac{3s}{4})^g}\right\} + \frac{3}{4}\left\{\frac{1}{\theta(n - \frac{3s}{4})^g}\right\} \tag{26}$$

Putting it all together, we see that

$$\frac{1}{\theta(\frac{s}{2})^g} < \frac{3}{4}\left\{\frac{1}{\theta(\frac{3s}{4})^g}\right\} + \frac{3}{4}\left\{\frac{1}{\theta(n - \frac{3s}{4})^g}\right\} \tag{27}$$

This shows that $\frac{\partial^2 W}{\partial s \partial g} > 0$. So, $\frac{\partial s^W}{\partial g} > 0$ ■

This proposition is intuitive. As patients become more risk averse, the utility gains from each correct patient/doctor match get larger. Meanwhile, the utility loss from each generalist who forgoes treating a patient to do a referrals gets smaller. These two effects push the social planner toward greater specialization.

3.2 Corner Solutions

Up to now, I have assumed an interior solution to the planner's problem. Yet s is bounded between 0 and $\frac{2n}{3}$. The optimal solution might lie at the corners: either all doctors serve as generalists ($s = 0$) or every doctor who can feasibly serve as a specialist does ($s = \frac{2n}{3}$).

$W'(s)$ is the net additional benefit from adding a specialist. If this benefit is less than zero, the social planner wants as few specialists as possible. By contrast, if this benefit is positive, the planner wishes to have as many specialists as possible.[8]. The likelihood of either corner solution depends on the slope of the benefit function, an unsurprising outcome since this slope reflects the gains from specialization. The following proposition states this result.

Proposition 3 *When the gains from doctor specialization are sufficiently small, the social planner sets prices such that no doctor specializes. On the other hand, when the gains from specialization are sufficiently large, the social planner sets prices such that every doctor who feasibly can specializes.*

Proof:

Because of symmetry of the benefit functions: $\Psi(\frac{3s}{4}) = \theta(n - \frac{3s}{4}))$. Given this, $W'(s)$ can be written as

$$v(\theta(\frac{s}{2})) - \frac{3}{4}(v(\theta(\frac{3s}{4})) - \frac{3}{4}v(\theta(n - \frac{3s}{4})) + \frac{1}{2}k(n - \frac{s}{2}) - k(s) \tag{28}$$

$W' < 0$ if

$$v(\theta(\frac{s}{2})) - \frac{3}{4}(v(\theta(\frac{3s}{4})) - \frac{3}{4}v(\theta(n - \frac{3s}{4})) < k(s) - \frac{1}{2}k(n - \frac{s}{2}) \tag{29}$$

$W' > 0$ if

$$v(\theta(\frac{s}{2})) - \frac{3}{4}(v(\theta(\frac{3s}{4})) - \frac{3}{4}v(\theta(n - \frac{3s}{4})) > k(s) - \frac{1}{2}k(n - \frac{s}{2}) \tag{30}$$

---

[8] This result can be easily derived. Set up the constrained maximization problem: $\max_s W(s)$ subject to $s \geq 0$ and $s \leq \frac{2n}{3}$. The Lagrangian is

$$\widetilde{L} = W(s) - \lambda_1[s - \frac{2n}{3}] + \lambda_2[s] \tag{A1}$$

The relevant FOCs are

$$\frac{\partial \widetilde{L}}{\partial s} = 0 \tag{A2}$$

$$\lambda_1[s - \frac{2n}{3}] = 0 \tag{A3}$$

$$\lambda_2[s] = 0 \tag{A4}$$

$$\lambda_1, \lambda_2 \geq 0 \tag{A5}$$

$\frac{\partial \widetilde{L}}{\partial s} = W'(s) - \lambda_1 + \lambda_2 = 0$. Suppose $W'(s) < 0$. For (A1) to hold $\lambda_2 > 0$. As a result, $s^W = 0$; otherwise (A4) won't hold. Suppose $W'(s) < 0$. For (A1) to bind $\lambda_1 > 0$. Given this positive multiplier, $s^* = \frac{2n}{3}$ or else (A3) won't hold.

Let $A = \left|\theta^{'}\right| = \left|\frac{\theta(\frac{3s}{4})-\theta(\frac{s}{2})}{\frac{3s}{4}-\frac{s}{2}}\right|$, where A is the benefit from specialization. By definition as A increases, $\theta(\frac{s}{2})-\theta(\frac{3s}{4})$ increases, meaning that $v(\theta(\frac{s}{2}))-\frac{3}{4}(v(\theta(\frac{3s}{4}))$ increases. So, the LHS of (29) and (30) are increasing in A. For a sufficiently large value of A, the LHS of (30) must be bigger than the RHS of (30). At $A=0$,there are no benefits from specialization, which gives the best chance for the inequality in (29) to hold ■

4 Maximizing Joint Profit

Now suppose that the doctors collectively set prices to induce the equilibrium they most prefer. What prices will they pick? Will they pick prices that restrict the number of specialists as compared to the social optimal? Maybe not. To understand this counterintutive result, note that the total profit from inducing an equilibrium with s referring doctors is

$$\Pi(s) = st_e^*(s) + (n - \frac{3s}{2})t_g^*(s) \tag{31}$$

Profits are concave. Taking the derivative and setting to zero gives

$$t_e^* + s\frac{\partial t_e^*}{\partial s} + (n - \frac{3s}{2})\frac{\partial t_g^*}{\partial s} - \frac{3}{2}t_g^* = 0 \tag{32}$$

Let s^{PM} be the profit maximizing number of medical specialists. The following threshold condition is used in the next proposition

$$\overline{k} = \frac{(\frac{9}{4}s^W - \frac{1}{2}n)k'}{2}$$

Proposition 4 *When doctors collectively set prices, three different outcomes are possible. (1) If $k(n-\frac{s^W}{2}) > \overline{k}$ doctors set prices to induce a level of specialization which is more than the social optimum; (2) If $k(n-\frac{s^W}{2}) < \overline{k}$ doctors set prices to induce a level of specialization which is less than the social optimum; and (3) If $k(n-\frac{s^W}{2}) = \overline{k}$, doctors set prices to induce a level of specializtion which is the social optimum.*

Proof:

Assume the patient indifferent between specialist and generalist treatment has illness 1. From equations (4) and (5), the following can be derived:

$$t_g^* = \frac{1}{n-\frac{3s}{2}}\int_{\frac{3s}{4}}^{n-\frac{3s}{4}}\frac{1}{2}\left[v(\theta(i)) + v(\Psi(i))\right]di - k(n-\frac{s}{2}) \tag{33}$$

$$t_e^* = \frac{1}{\frac{s}{2}}\int_0^{\frac{s}{2}}v(\theta(i))di - \frac{1}{n-\frac{3s}{2}}\int_{\frac{3s}{4}}^{n-\frac{3s}{4}}\frac{1}{2}\left[v(\theta(i)) + v(\Psi(i))\right]di - k(s) + t_g^* \tag{34}$$

Plugging in t_g^* and t_e^* into the profit equation and doing some algebra gives

$$\Pi(s) = 2\int_0^{\frac{s}{2}} v(\theta(i))di + \int_{\frac{3s}{4}}^{n-\frac{3s}{4}} \frac{1}{2}\left[v(\theta(i)) + v(\Psi(i))\right] di - sk(s) - (n - \frac{5}{2}s)k(n - \frac{s}{2}) \quad (35)$$

From (15), we know

$$W(s) = 2\int_{i=0}^{\frac{s}{2}} v(\theta(i))di + \int_{i=\frac{3s}{4}}^{n-\frac{3s}{4}} \frac{1}{2}\{v(\theta(i)) + v(\Psi(i))\}di - \int_{j=0}^{s} 2k(j)dj - \int_{j=s}^{n-\frac{s}{2}} k(j)dj \quad (36)$$

Rearranging (36) gives

$$W(s) + \int_{j=0}^{s} 2k(j)dj + \int_{j=s}^{n-\frac{s}{2}} k(j)dj = 2\int_{i=0}^{\frac{s}{2}} v(\theta(i))di + \int_{i=\frac{3s}{4}}^{n-\frac{3s}{4}} \frac{1}{2}\{v(\theta(i)) + v(\Psi(i))\}di \quad (37)$$

Subtract $sk(s) + (n - \frac{5}{2}s)k(n - \frac{s}{2})$ from both sides of (37). The result is

$$W(s) + \int_{j=0}^{s} 2k(j)dj + \int_{j=s}^{n-\frac{s}{2}} k(j)dj - sk(s) - (n - \frac{5}{2}s)k(n - \frac{s}{2}) = 2\int_{i=0}^{\frac{s}{2}} v(\theta(i))di + \int_{i=\frac{3s}{4}}^{n-\frac{3s}{4}} \frac{1}{2}\{v(\theta(i)) + v(\Psi(i))\}di - sk(s) - (n - \frac{5}{2}s)k(n - \frac{s}{2}) \quad (38)$$

Or

$$W(s) + H(s) = \Pi(s) \quad (39)$$

where $H(s) = \int_{j=0}^{s} 2k(j)dj + \int_{j=s}^{n-\frac{s}{2}} k(j)dj - sk(s) - (n - \frac{5}{2}s)k(n - \frac{s}{2})$. The derivative of (39) equals

$$\Pi'(s) = W'(s) + H'(s) \quad (40)$$

At the optimal value, s^W, $W'(s^W) = 0$. If $H'(s^W) = 0$, then, $\Pi'(s^W) = 0$ meaning that $s^W = s^{PM}$. Alternatively, if $H'(s^W) > 0$, then, $\Pi'(s^W) > 0$. Since Π is concave it must be that $s^{PM} > s^W$. Finally, if $H'(s^W) < 0$, then, $\Pi'(s^W) < 0$ and $s^{PM} < s^W$. After cancelling common terms, $H'(s^W)$ equals

$$H'(s^W) = 2k(n - \frac{s^W}{2}) - sk'(s^W) + \frac{1}{2}(n - \frac{5}{2}s)k'(n - \frac{s^W}{2}) \quad (41)$$

Since $k'' = 0$, we know that $k'(s) = k'(n - \frac{s}{2})$ for any value of s^W. Solving for $k(n - \frac{s^W}{2})$ gives the threshold condition

$$\overline{k} = \frac{(\frac{9}{4}s^W - \frac{1}{2}n)k'}{2} \quad (42)$$

If $k(n - \frac{s^W}{2}) > \overline{k}$ then $H'(s^W) > 0$ and $s^{PM} > s^W$. If $k(n - \frac{s^W}{2}) < \overline{k}$, $H'(s^W) < 0$ and $s^{PM} < s^W$. If $s^W \leq \frac{2}{9}n$, the RHS of (42) is always negative and doctors will choose prices such that there is too much specialization relative to the optimal amount ■

Why doesn't monopoly pricing necessarily result in too little specialization? Say we have 100 doctors. By selecting the prices, the doctors, in effect, decide how many doctors become specialists, generalists, and referring physicians. Suppose doctors decide on prices such that 20 patients demand speciality services. 10 referring doctors are required to support 20 specialists. That leaves 70 generalists treating patients. If the doctors restrict the output of specialists to, say, 10 that leaves 85 generalists treating patients. By restricting the output in the market for specialists, doctors increase the supply and lower the price in the generalist market. So, we don't have the standard monopoly story here. It is more complicated, depending on how the generalist price and specialist price respond to changes in quantity. If the generalist price increases a great deal with an increase in the number of specialists, doctors will set prices that result in too much specialization. If, on the other hand, the generalist price only increases a little bit with a change in the number of specialists, doctors will set prices that result in too little specialization.

Whether doctors pick prices to induce too few or too many specialists turns on the cost per visit for the last patient treated at the social optimum, $k(n - \frac{s^W}{2})$. To see this, plug t_g^* and t_e^* into (32). Collecting terms, we have

$$\frac{1}{\frac{s}{2}}\int_0^{\frac{s}{2}} v(\theta(i))di - \frac{1}{n - \frac{3s}{2}}\int_{\frac{3s}{4}}^{n-\frac{3s}{4}} \frac{1}{2}\left[v(\theta(i)) + v(\Psi(i))\right]di$$

$$- k(s) + (n - \frac{3s}{2})\frac{\partial t_g^*}{\partial s} = \frac{1}{2}t_g^* - s\frac{\partial t_e^*}{\partial s} \quad (43)$$

The LHS side of (43) is the marginal benefit to the doctors of adding another specialist. The RHS represents the marginal cost. Let's say the social optimal has 20 doctors providing speciality services. Suppose $k(n - \frac{20}{2})$ is quite large. Given the inverse relationship between this cost per visit and the generalist price, this means that $\frac{1}{2}t_g^*$ is quite small.

Evaluated at the social optimum, then, the marginal cost is small and the marginal benefit remains unchanged (it doesn't depend on $k(n - \frac{20}{2})$). And so, the doctors select more specialists than is socially optimal. The reverse holds if $k(n - \frac{20}{2})$ is small. In that case, the marginal cost of an additional specialist (again measured at the social optimum) is big and the marginal benefit unchanged, inducing the doctors to select fewer than the optimal amount of specialists.

5 Conclusion

Medical specialization by doctors is important. The role of referring doctors in facilitating specialization has not been the subject of much study by health economists. This paper is a step toward filling that void. To make the analysis tractable, the model ignores the role of education and doctor investment in specialized skills. The skill level of each doctor was taken as given. In a richer model, we might expect some doctors to make specialized investments via fellowships or additional training. Those considerations are left for future work.

References

Alchian, Armen A & Demsetz, Harold, 1972. "Production , Information Costs, and Economic Organization," 62 *American Economic Review* 777-795.

Association of American Medical Colleges. 2007. "2007 Medical School Graduation Questionaire – All Schools Summary Report," available at www.aamc.org/start.htm.

Baumgarder, James R. 1992. "Medicare Physician-Payment Reform and the Resource-Based Relative Value Scale: A Recreation of Efficient Market Prices," 82 *American Economic Review* 1027-1030.

Baumgardner, James R. 1988. "The Division of Labor, Local Markets, and Worker Organization," 96 *Journal of Political Economy* 509-527.

Becker, Gary and Murphy, Kevin. 1992. "The Division of Labor, Coordination Costs, and Knowledge," 107 *Quarterly Journal of Economics* 1137-60

Bolton, Patrict and Dewatripoint, Mathas. 1994. "The Firm as a Communication Network," 109 *Quarterly Journal of Economics* 809-839.

Brotherton et. al. 2005. "US Graduate Medical Education, 2004-2005," 294 *Journal of the American Medical Association* 1075-1082.

Carbona et. al. 2006. "Impact of Physician Asthma Care Education on Patient Outcomes," 117 *Pediatrics* 2149-2157.

Clark et. al. 1997. "Impact of Education for Physicians on Patient Outcomes," 101 *Pediatrics* 831-836.

Sujit Choudry, Sujit and Brennan, Troyen A. 2001. "Collective Bargaining By Physicians – Labor Law, Antitrust Law, and Organized Medicine," 345 *New England Journal of Medicine* 1141-1145.

Demange, Gabrielle and Gale, David. 1985. "The Strategy of Two-Sided Matching Models," 53 *Econometrica* 873-888.

Donini-Lenhoff, Fred G.and Hedrick, Hannah L. 2000. "Growth of Specialization in Graduate Medical Education." 204 J*ournal of the American Medical Association* 1284-1290.

Diamond, Peter A. 1982. "Wage Determination and Efficiency in Search Equilibrium," 49 *Review of Economic Studies* 217-227.

Gale, David and Shapley, Lloyd. 1962. "College Admissions and the Stability of Marriage," 69 *American Mathematical Monthly* 9-15

Gaynor, Martin and Vogt, William B. 2000. "Antitrust and Competition in Health Care Markets, " in *Handbook of Health Economics*, ed. Anthony J. Culyer and Joseph P. Newhouse.

Jones, Benjamin F.. 2008. "The Knowledge Trap: Human Capital and Development Reconsidered," (working paper, Kellogg School of Business).

Kim, Sunwoong. 1989. "Labor Specialization and the Extent of the Market." 97 *Journal of Political Economy* 692-705.

McGuire, T.G. 2000. "Physician Agency," in *Handbook of Health Economics*, ed. Anthony J. Culyer and Joseph P. Newhouse 461-528.

McGuire, T.G. and Ma, C.A. 1998. "Network Effects in Managed Health Care," 11 *Journal of Economics and Management Strategy* 1-35.

Mortensen, Dale T. 1982. "The Matching Process as a Noncoorperative Bargaining Game." *The Economics of Information and Uncertainty*. Chicago: University of Chicago Press for the National Bureau of Economic Research 233-258

Rochet, Jean-Charles & Tirole, Jean. 2002. "Cooperation among competitors: some economics of payment card associations" 33 *RAND Journal of Economics* 549-570.

Rochet, Jean-Charles & Tirole, Jean. 2006. "Two-Sided Markets: A Progress Report," 37 *Rand Journal of Economics* 645-667.

Rubinstein, Ariel, and Asher Wolinsky. 1987. "Middlemen," 102 *Quarterly Journal of Economics* 581-593.

Tirole, Jean. 1987. Industrial Organization (MIT Press).