Essays in the Microeconomics of Medical Specialization

Suzette Applegate Baker

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctorate in Philosophy in the Department of Economics.

Chapel Hill 2010

Approved by:

Gary Biglaiser

Sergio Parreiras

Helen Tauchen

Tracy Lewis

Donna Gilleskie

ABSTRACT

Suzette Applegate Baker: Essays in the Microeconomics of Medical Specialization (Under the direction of Gary Biglaiser)

This dissertation consists of two essays. In the first essay, I examine the interaction between medical specialization and patient referrals. 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 patients 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 specialist services, and (3) price of generalist services. I find that, when doctors collectively set prices for both specialist and generalist treatment. Depending on the parameters, the joint maximization problem can result in (a) too many specialists and two few generalists; (b) too many generalists and too few specialists; or (c) the optimal number of specialists and generalists. This ambiguous result shares similarities with the textbook model of the quantity decisions of a multi-market monopolist.

The second essay considers the role played by fellowship programs in the training of medical researchers. Many hospitals hire senior researchers straight out of their own or another hospital's fellowship program. As a result, medical programs both "train" fellows and provide a supply of medical talent for that hospital in the next period. Using an overlapping generation model, I derive three results linking the underlying features of the medical marketplace, the size of fellowship programs, and the quality of medical research. First, when the hospital's time horizon or discount rate increases, the hospital tends to employ more fellows each period. Second, when the fellow's outside option depends on their skill level, the hospital employs fewer fellows each period. Finally, when the fellow's outside option depends on their skill level and the number of other fellows in the private-sector market, the hospital employs more fellows than in the case where the outside option depends on the fellow's skill level only.

ACKNOWLEGEMENT

I would like to thank my husband, Scott Baker, and our three children, Luke, Marissa, and Rachel for their patience and support. I am also indebted to several friends and family members who supported me along the way.

TABLE OF CONTENTS

LIST OF FIGURES

Figures

MEDICAL SPECIALISTS I: COORDINATION OF SPECIALISTS AND PATIENTS

1.1 Introduction

Specialization by physicians has transformed the way medical care is provided. By focusing on a single illness or injury, a specialist 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 factors influence the optimal degree of medical specialization, taking into account the cost of coordinating patients and physician-specialists.[1]

According to a recent survey of U.S. patients, "[a]lmost 63 million adults—nearly three in 10—said they needed a new specialist in the previous year, with 46 million actually seeing a new specialist." (Tu and Launer 2008 p. 2). The survey goes on to estimate that 86.5% of patients use primary care referral or another doctor's referral when selecting a specialist.

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,

[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).

Becker and Murphy 1992, Alchian and Demestz 1972). I study a different coordination problem here – the matching of patients with a specific illness 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 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 risk associated with generalized care and, as a result, patients will spend more for that service.

Third, as prices increase, the patients with the highest cost of seeking care forgo treatment. The optimal amount of specialization balances the cost of this patient exit against the gains from each additional patient-doctor match. Finally, I show that doctors acting in concert but assuming full employment cannot set prices to simultaneously restrict the amount of specialist services and generalist services.[2] This result shares similarities with

[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

the output decisions of a standard multi-market monopoly.

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.[3] 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 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, 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.

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] 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.

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 possible 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 a 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. Along related lines, 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.

The paper closest to mine is Garciano (2000). He considers the problem of communication within organizations. In his model, workers specialize in the production of knowledge or the transmission of knowledge. His key result is that harder problems are those most likely to be referred up the chain of command. The higher up in the chain a worker is, the more difficult the problems the organization will ask him to solve. A pyramid scheme results. This organization form minimizes communication costs while ensuring that problems can be solved (only hard problems are continually referred upward, which reduces the cost of transmission). My model differs because some "low-level" doctors decide to treat patients (i.e., solve the problem themselves), even though they could refer to a more qualified doctor for treatment. Patients don't complain about this practice because, in equilibrium, the price adjusts to reflect the lower quality of treatment generalists offer. In addition, the "communication cost" is the cost to the patient of making an additional trip to the specialist.

Finally, there is some recent empirical work on medical specialists. Johnson (2009) finds that, over time, primary care physicians learn about the quality of specialists to which they refer. As a result, she finds some evidence that lower quality specialists are more likely to drop out of the market. The sorting mechanism in Johnson's framework is that referrring physicians learn about specialist talent. In my model, fewer patients are treated because a subset of doctors opt to focus on referring patients rather than treating them. Since fewer doctors are available for treatment, some patients don't receive any care whatsoever.

Part 1.2 develops the model. Part 3 explores the level of specialization that maximizes social welfare. Part 3.1 considers corner solutions, where welfare is maximized by having either no specialists, as many specialists as feasibly possible, or failing to use all the available doctors. In part 4, doctors set prices. The prices set 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.

1.2 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 outcome in monetary terms of medical treatment depends on the doctor's skill level. If the patient has a type 1 illness, $\theta(i)$ is the patient's monetary value associated with treatment by doctor i. If the patient has a type 2 illness, $\Psi(i)$ is the patient's outcome 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 outcome function for illness one, $\theta(i)$, is linear and decreasing in the doctor index. The outcome 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.

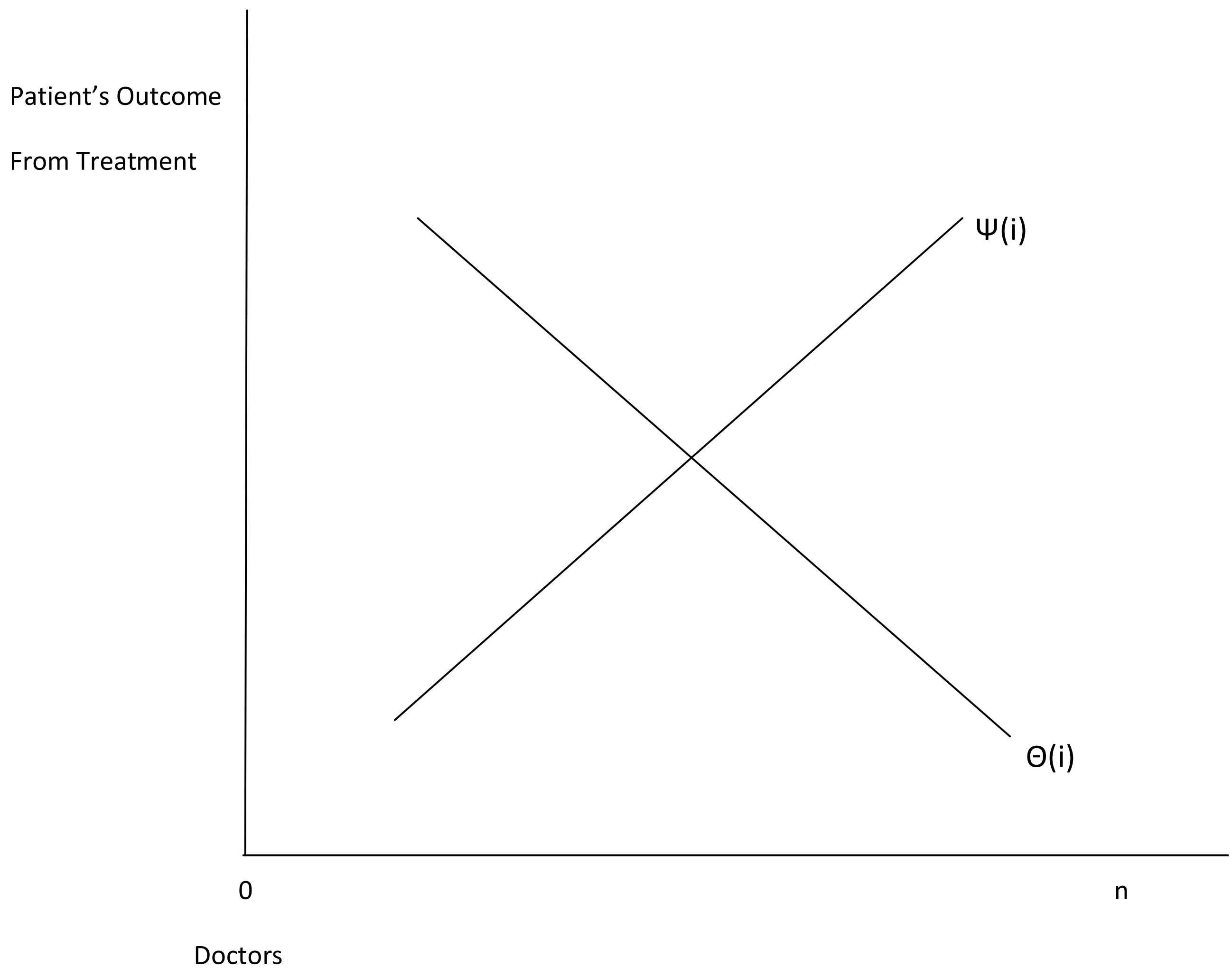

Figure 1: The Relationship Between Patient Benefit and Doctor Skill Levels

1.2.1 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.[4] 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 outcome from treatment, (2) the price paid for medical services; (3) whether she sees a generalist or, via a referral sees a specialist;

[4] 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.

(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

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

The utility functions show that patients differ in the "net" benefit from medical treatment. Because $k(n) > k(0)$, patients close to n have a lower net benefit from treatment than patients close to 0.

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 signaling their practice area 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. 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.[5]

[5] 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 $i \leq \underline{\underline{i}}$.

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

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

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

1.2.3 Timing and Initial Results

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

Assume that doctors do not set prices for now (i.e., take prices as given). 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 one of many 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 has 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

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

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. Therefore these patients chose to forgo treatment.

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 simultaneously 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, a sufficient, but not necessary 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 $\blacksquare$

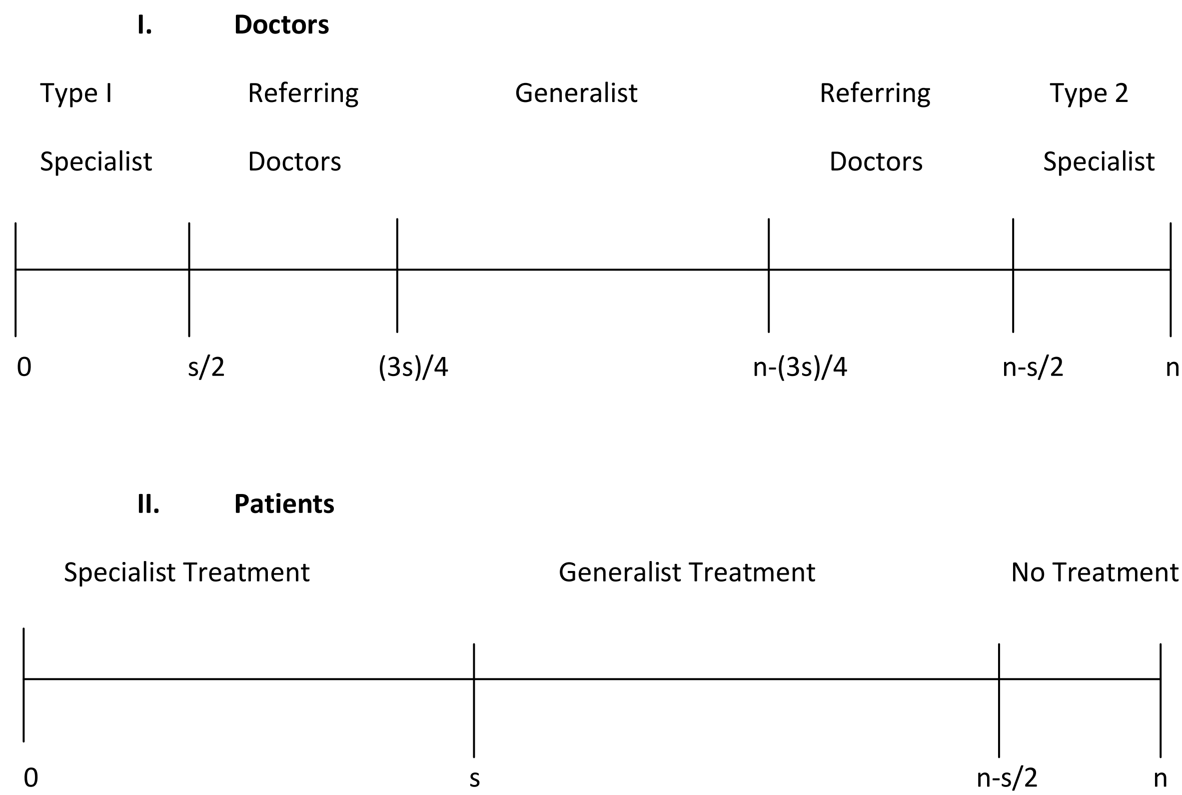

Figure 2: Equilibrium Allocation of Patients and Doctors

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 at least as much or more than generalists or referring doctors (the specialist's net payment, $t_e^* - \frac{t_g^*}{2}$ must be greater than or equal to 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, but 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^*, t_e^*, F^*) depend on the anticipated generalist pool. Because consumers are risk averse, they are willing to pay more for generalist treatment by a doctor with a lower spread in outcomes.[6] Among the

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

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

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 $\frac{n}{2}$. Because 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.

possible lineups of 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.

1.3 Welfare Analysis

Now consider the welfare effects of specialization. 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 patient outcome functions for type 1 and type 2 illnesses, welfare can be rewritten as

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

Welfare is concave.[7] The first order condition is

$$v(\theta(\frac{s}{2})) - \frac{3}{8}\left\{(v(\theta(n-\frac{3s}{4})) + v(\Psi(n-\frac{3s}{4}))\right\} - \frac{3}{8}\left\{(v(\theta(\frac{3s}{4})) + v(\Psi(\frac{3s}{4})\right\} + \frac{1}{2}k(n-\frac{s}{2}) - k(s) = 0$$

Since $\theta(n-\frac{3s}{4}) = \Psi(\frac{3s}{4})$ and $\Psi(n-\frac{3s}{4}) = \theta(\frac{3s}{4})$, we can rewrite the FOC condition as

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

The solution to (16) provides the optimal number of medical specialists, s^W. To increase the number of specialists by one unit requires an additional 1/2 unit of referral services, leading to the rationing of 1/2 unit of patient care. The left hand side of (16) represents

[7] To see concavity, note that $W''(s)$ equals

$$\frac{1}{2}\theta'(\frac{s}{2})v'(\theta(\frac{s}{2})) - \frac{9}{16}\left[\theta'(\frac{3s}{4})v'(\theta'(\frac{3s}{4})) + \Psi^{'}(\frac{3s}{4})v'(\Psi(\frac{3s}{4}))\right] - \frac{1}{4}k'(n-\frac{s}{2}) - k'(s)$$

Since $\theta'(\frac{3s}{4}) = -\Psi^{'}(\frac{3s}{4})$, $W''(s)$ can be rewritten

$$\frac{1}{2}\theta'(\frac{s}{2})v'(\theta(\frac{s}{2})) - \frac{9}{16}\theta'(\frac{3s}{4})\left[v'(\theta(\frac{3s}{4})) - v'(\Psi(\frac{3s}{4}))\right] - \frac{1}{4}k'(n-\frac{s}{2}) - k'(s)$$

Notice that $\theta'(\frac{s}{2}) < 0$ and $k' > 0$. Since $\theta(\frac{3s}{4}) > \Psi(\frac{3s}{4})$, it must be true that $v'(\theta'(\frac{3s}{4})) < v'(\Psi(\frac{3s}{4})$. Taken all together, the entire expression is negative.

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.[8]

1.3.1 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 ($s = \frac{2n}{3}$). In addition, the social planner might not want to employ all the doctors.

$W'(s)$ is the net additional benefit from adding a specialist. Welfare is concave. If $W'(0) < 0$, social welfare will be decreasing as the social planner employs any more than zero specialists. And so, it is optimal to have no specialists. Alternatively, if the marginal gain from adding a specialist is still positive at $s = \frac{2n}{3}$ (i.e., $W'(\frac{2n}{3}) > 0$), it makes sense to employ the maximum number of specialists.[9] These results are described in figures 3

[8] In this analysis, I only consider one class of welfare functions, where each patient is weighted by the social planner equally. The results might differ with different weights for different patients (like, say, weighting the patients with the highest benefit from treatment the most). That is to say, the analysis might not be robust to different welfare functional forms.

[9] 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

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

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

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

and 4. Figure 3 represents the situation where it is optimal to have the maximum number of specialists ($s = \frac{2n}{3}$). Note that the slope of the tangent line at $\frac{2n}{3}$ is positive. Figure 4 represents the situation where it is optimal to have no specialists. Note that the slope of the tangent line at 0 is negative.

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

$\frac{\partial \tilde{L}}{\partial s} = W'(s) - \lambda_1 + \lambda_2 = 0$. Suppose $W'(s) < 0$. For (A2) to hold $\lambda_2 > 0$. As a result, $s^W = 0$; otherwise (A4) won't hold. Suppose $W'(s) > 0$. For (A2) to hold $\lambda_1 > 0$. Given this positive multiplier, $s^* = \frac{2n}{3}$ or else (A3) won't hold.

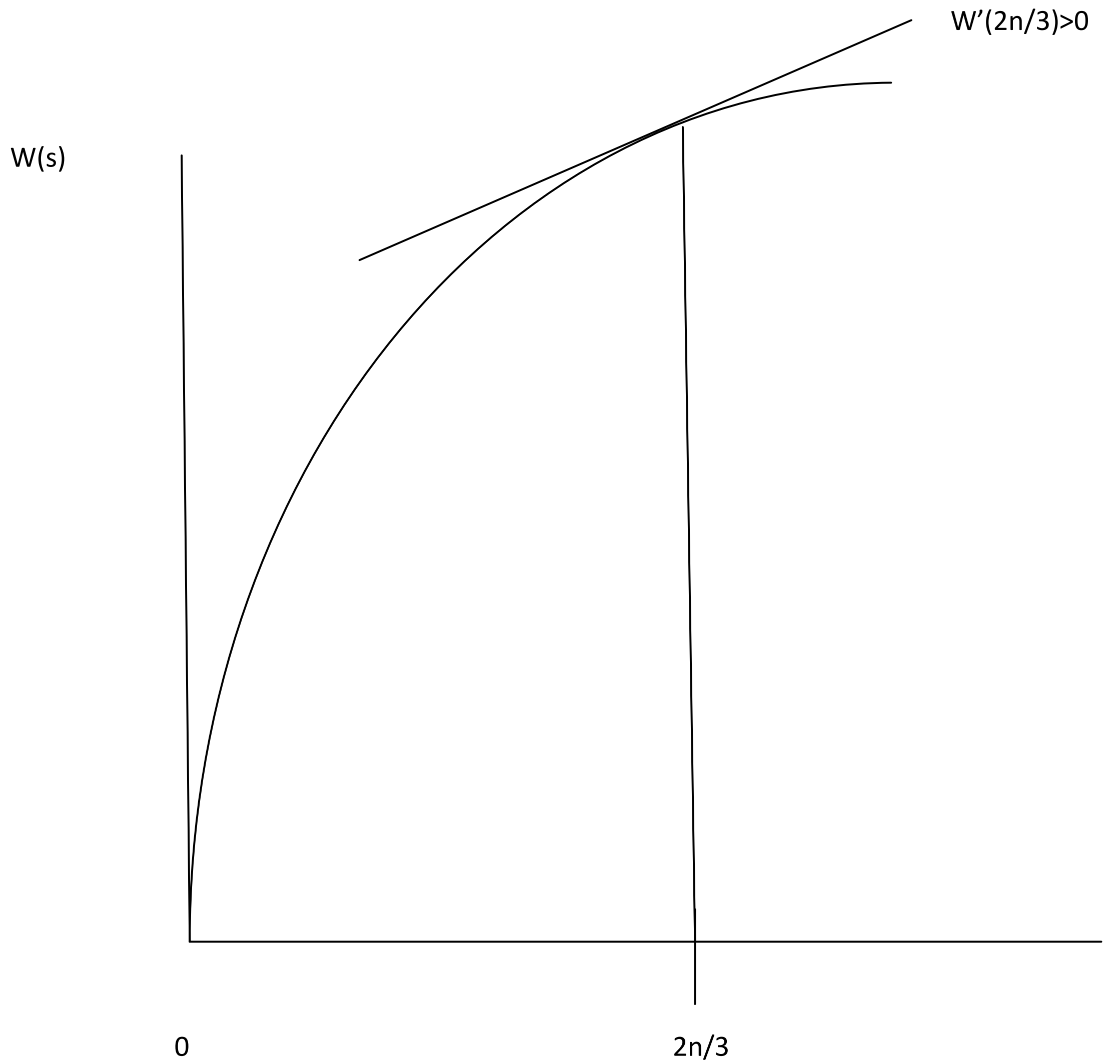

Figure 3: Corner Solution with Maximum Specialists

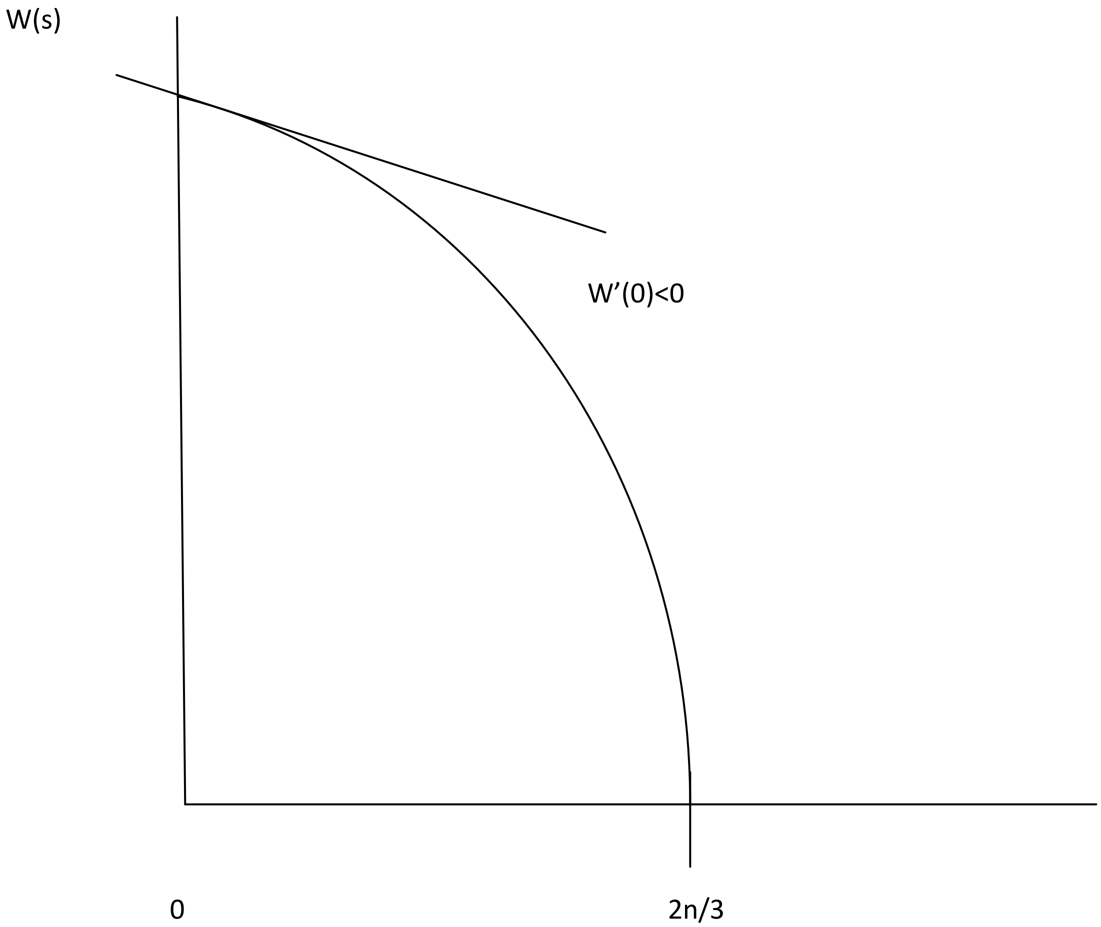

Figure 4: Corner Solution with No Specialists

The likelihood of either corner solution depends on the slope of the patient outcome function and the cost of each visit, a logical outcome since this slope reflects the gains from specialization. In defining formally the proposition, the following definition is useful

$$k^* = 2v(\theta(\frac{n}{3})) - 3(v(\theta(\frac{n}{2}))$$

Notice that k^* is increasing as the gains from specialization increase. That is, as $\theta(\frac{n}{3})$ grows with respect to $\theta(\frac{n}{2})$, k^* gets larger. We have the following result:

Proposition 2 *(i) If $k(0) > \frac{1}{2}k(n)$, as the gains from doctor specialization vanish, the social planner sets prices such that no doctor specializes . (ii) When $k(\frac{2n}{3}) \leq k^*$ the social planner sets prices such that every doctor who feasibly can specializes. (iii) When the cost per visit is sufficiently high, the social planner sets the number of specialists equal to zero and employs fewer than n doctors.*

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

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

Proof of (i).

$W'(0) < 0$ if the following condition holds

$$v(\theta(0)) - \frac{3}{4}(v(\theta(0)) - \frac{3}{4}v(\theta(n)) < k(0) - \frac{1}{2}k(n) \tag{29}$$

Suppose the gains from specialization vanish. That means that $\theta(0) = \theta(n)$. The LHS of the above equation must be less than zero. A sufficient condition for the inequality to hold is that the RHS is positive, which happens whenever $k(0) > \frac{1}{2}k(n)$.

Proof of (ii)

$W'(\frac{2n}{3}) > 0$ if the following holds

$$v(\theta(\frac{n}{3})) - \frac{3}{4}(v(\theta(\frac{n}{2})) - \frac{3}{4}v(\theta(\frac{n}{2})) + \frac{1}{2}k(n - \frac{n}{3}) - k(\frac{2n}{3}) > 0 \tag{30}$$

which reduces to

$$v(\theta(\frac{n}{3})) - \frac{3}{2}(v(\theta(\frac{n}{2})) \geq \frac{1}{2}k(\frac{2n}{3})$$

The LHS defines k^*, which completes the proof.

Proof of (iii)

The welfare associated with employing one less doctor (the best generalist), but still no specialists is

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

This welfare is greater if the following condition holds true

$$\frac{1}{n}\int_{i=0}^{n}\frac{1}{2}\{v(\theta(i)) + v(\Psi(i))\}di - k(n) < 0$$

In that event, the social planner wants to employ no specialists and also restricts the number of generalists treating patients ■

Depending on the shape of the welfare function, we might or might not have corner solutions. It depends on the shape of the utility function and its relationship to the size of the cost of the patient visit.

1.3.2 Maximizing Joint Profit

Now suppose that the doctors collectively set prices to induce their most-preferred equilibrium. Will they pick prices that result in too many specialists or too many generalists when compared to the social optimum? In answering this question, I assume that each doctor must earn a positive price in equilibrium. In other words, doctors can't constrain the total number of doctors participating in the overall market (as generalists, specialists, or referring doctors). Instead, the choice of prices simply shifts the proportion of doctors in each practice.

Doctors maximize total profit by choosing an equilibrium with s referring doctors. Because of the "full employment" assumption, this choice determines the number of specialists, the number of generalists and the number of referring doctors. Profit equals

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

Profits are concave. Setting the first order condition equal 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 3 *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 specialization 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

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

From (15), we know

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

Rearranging (36) gives

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

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

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

Or

$$W(s) + H(s) = \Pi(s) \tag{39}$$

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

$$\Pi'(s) = W'(s) + H'(s) \tag{40}$$

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

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

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

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

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 ■

Compare this result to the model of a monopolist producing substitutes goods in two separate markets. It is well-known that this multi-market monopolist will restrict prices in both markets. Indeed, he will do so more than the standard monopolist (Tirole 1987 p. 70). At the same time, the effect on output in each market is indeterminate. By raising prices in market A, the multi-market monopolist increases quantity in market B. At the same time, the multi-market monopolist raises the price in market B, restricting output in that market. The end result in terms of output is indeterminate. Depending on the parameters of the model, the above proposition suggests the same style of result in my model. Doctors will set prices to induce (1) too many specialists and too few generalists; (2) too few specialists and too many generalists; or (3) the optimal number of specialists and generalists.

The key difference in my model is that the total number of doctors employed is fixed at n. By manipulating the prices, the doctors adjust how many doctors are doing each activity. Unlike a multi-market monopolist, even if they wanted to, the doctors can't restrict the total supply of services. When the doctors induce more specialists, they necessarily induce fewer generalists. When they set prices to induce fewer specialists, more generalists come as a by-product.

Take an example. Say we have 100 doctors. Suppose doctors decide on prices such that 20 patients demand specialist services. 10 referring doctors are required to support 20 specialists. That choice 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. 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 an increase in the number of specialists (because the risk averse patients are, say, willing to pay very little for the better generalist pool), 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.

1.4 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.

MEDICAL SPECIALISTS II: THE ROLE OF FELLOWSHIPS IN MEDICAL TRAINING

2.1 Introduction

Unlike many training situations, hospitals both train fellows and hire former fellows onto their staffs. Because, in staffing fellowship programs, the hospital partially determines the quality of its future workforce, the decisions about how many fellows to hire differs from the standard one studied in the literature. The hospital both provides educational benefits and reaps at least some of those educational benefits as employers. Yet any educational benefit provided to fellows might spillover – the fellows might cash in on their training themselves and not work for the hospital in the next period. This externality complicates the hospital's decision about how many fellows to employ in a given period.

Fellowship programs in the United States play a large role in the advanced training of doctors. They are the last step in the training of specialists, lasting between one and three years. Most major-university hospitals have fellowship programs. The fellow has finished his residency and, in theory, works with a more senior doctor to learn the most advanced treatment and research in that area of medicine. Unlike residents, fellows can serve as attending physicians. They are fully registered and licensed. If need be, they can perform surgeries without supervision. They can help manage and train residents. More important, hospitals look to the fellowship programs to hire the next medical researcher in their practice. Fellowship programs, in other words, are the stream of medical specialist quality.

I start by investigating a single hospital and an overlapping generation of doctors. Each period the hospital hires one seasoned medical researcher out of the pool of fellows who just completed its fellowship program. Further, the hospital decides how many newly-minted residents to bring into the fellowship program. The process repeats: each period the hospital hires a senior doctor and matches them with a fresh set of fellows. The result from this simple model is this: The longer the time horizon of the hospital has, the more fellows it hires in any one period and, correspondingly, the higher the quality of the research produced by the hospital. The longer time horizon gives the hospital a greater chance of recouping the benefits from the fellowship training program. Likewise, the higher the hospital's discount rate – the more patient the hospital is – the more fellows it hires and the higher the average quality of the seasoned doctor's research.

Next, I consider the situation where the fellow can trade on their fellowship training by treating patients in the private sector. I examine two cases. In the first case, the price for treatment in the private sector is exogenous. In the second case, the price for treatment services in the private sector is endogenous. More precisely, with an exogenous treatment price, the price is independent of the number of fellows produced in the prior period, but does depend on the quality of the fellow. In this circumstance, I show that the hospital has less of an incentive to employ fellows. The reason: Program expansion increases the odds that a fellow will be of high quality and, as a result, able to receive a lucrative outside option. Therefore, by training additional fellows the hospital confers a positive externality on the fellows.

With an endogenous treatment price, the price depends on (1) the number of fellows flowing out of the program in the prior period and (2) the fellow's quality. With an endogenous treatment price, program expansion drives down the price of treatment services (by increasing the supply of fellows and, as a result, the number of doctors in the private sector). The hospital has more of an incentive to employ fellows than in the exogenous price case.

The upshot of the model is that the hospital's time horizon, its discount rate, and the features of the non-academic, private sector market determine the size and scope of the fellowship programs. The size and scope of the fellowship program affects the quality of medical research. And so, we get a positive relationship between a hospital's concern about the future and doctor quality and medical research. But this effect doesn't come from the conventional story of the hospital wanting to maintain a good reputation. Instead, it arises out of the unique training grounds associated with medicine.

I use an overlapping generation model (OLG), which has been called one of the two building blocks of modern macroeconomics (La Croix and Michel 2002, p. 1). Macroeconomists have used the model to study the accumulation of capital (Diamond 1965), the funding of education (Glomn and Ravikumar 1992; de la Croix and Monfort 2000), tax versus debt financing of public expenditures (Diamond 1965, King 1992, Grossman and Yanagawa 1993), and the impact of altruism on intergenerational transfers (Barro and Becker 1989).

Samuelson (1958) is a foundational paper for this style of model. He shows that a pure exchange economy can have multiple equilibria. In one equilibria, the interest rate equals the growth rate of the population; in the second equilibria, there is no trade whatsoever. Samuelson, then, argues that a social agreement to use money as a store of value can eliminate the "no trade" equilibrium.

In the other seminal OLG paper, Diamond (1965) finds convergence of the capital in an economy, but that convergence might be inefficient. In other words, an economy under perfectly competitive conditions can accumulate too little or too much capital. That result has been the subject of an extensive follow-on literature (see, for example, Hu 1979; Nourry 2001; Fanti and Spataro 2006).

Micro-economists have picked up the overlapping generation model to examine a number of issues. Tadelis (2002), for example, shows how a "market" for reputations that are distinct from the employees of the firm can maintain incentives for short-lived agents. Cremer (1986) demonstrates how the development of social norms in an infinitely-lived corporation can encourage "good" behavior by short-lived agents. Cremer also provides a rationale for why the youngest workers are often given the most difficult tasks. The reason: If they shirk, young workers are subject to more years of punishment at the hands of other workers in the organization.

The paper proceeds as follows: Section 2 develops the benchmark model with a single hospital and overlapping generations of doctors. Section 3 expands the model, containing the results when the fellow's outside option depends on his skill level and the price he gets for treating people in the private sector. Section 4 concludes.

2.2 Benchmark Case

Consider a single hospital that lives for multiple periods. The hospital produces medical research, which results from the combined efforts of senior researchers and medical fellows. Because medical research is capital intensive, I assume that the hospital has the capacity to employ at most one senior doctor to engage in research each period. The hospital can, however, match or employ as many fellows as it wants to work with this senior doctor.

Senior doctors have either a high skill level θ^H or a low skill level θ^L. Let the skill level be the hospital's payoff from employing a doctor, with θ^H greater than θ^L. The payoff from employing the seasoned doctor depends on the number of fellows matched with that doctor. These payoffs are increasing and concave in the number of fellows.[1]

In the first period, the senior doctor has a low skill level.[2] The fellowship program trains a discrete number of fellows, $s \in [0, m]$. The hospital also observes the fellows in action, which means the hospital can distinguish between fellows likely to have a high skill level as a senior researcher and those likely to have a low skill level. Whether the training program produces at least one high quality researcher depends on the number of fellows in the program. To capture this in the easiest way, assume that each fellow transforms into a high quality researcher with probability p. He becomes a low quality researcher with probability $1 - p$. These probabilities are independent of the number of fellows in the program.[3] The independence assumption means that there is no negative effect on fellow training associated with "crowding" i.e., multiple fellows attempting to learn from a single senior researcher. Under these assumptions, the probability that a pool of size s fails to produce at least one high type is

$$prob\{no\ high\ type\ |s\} = (1 - p)^s$$

This is the probability of "no successes" from a binomial distribution, where the draws are

[1] Obviously, employees of the hospital both treat patients and conduct research. I abstract away from the treatment decision here. The same results apply if we view θ^H and θ^L as the hospital's payoff from the combination of treatment and research.

[2] I make this assumption for simplicity. At the beginning there are no fellowship programs and hence no training. As a result, the first doctor hired has low skill. Realistically, the first doctor could be a high type or stochastically determined. If I allowed for this, the results would remain substantially similar.

[3] I also assume that a fellow is equally likely to be a high type when matched with a high skill doctor or a low skill doctor (that is, the value of p is the same). We might include different probabilities of producing a high or low type fellow as a function of the senior doctor's skill level (say p and p', where p' is the probability of a high type when matched with a high type senior doctor, and $p' > p$). While more realistic, this addition to the model would complicate the notation without adding new insights.

the number of the fellows in the pool. Clearly, this probability is decreasing in s.

The timing of the game follows:

1. The hospital hires a single seasoned doctor to engage in medical research.

2. The hospital decides on a number of fellows, s_1, to assign to the seasoned doctor in the fellowship program.

3. The period ends.

4. The hospital hires a single seasoned doctor out of the pool of fellows from the prior period.

5. The hospital decides how many new fellows, s_2, to assign to that doctor through the fellowship program.

2.2.1 One Period Model

In a one period model, the hospital will choose the number of fellows, s_1, to maximize

$$\theta^L(s_1) - ws_1$$

Since the maximization problem involves selecting a discrete number of fellows, it is useful to define the following function

$$G(w, s_1) = [\theta^L(s_1) - ws_1] - [\theta^L(s_1 - 1) - w * (s_1 - 1)]$$

$G(w, s_1)$ represents the incremental gain in profit from moving from $s_1 - 1$ fellows to s_1 fellows. The concavity of $\theta^L(s)$ implies that the additional benefit of increasing s by one unit is always decreasing.[4] It follows that, for all s_1,

$$G(w, s_1) > G(w, s_1 + 1)$$

At the maximum, $\widetilde{s}(w)$, the following condition must hold:

$$G(w, \widetilde{s}_1) \geq 0 \geq G(w, \widetilde{s}_1 + 1) \tag{1}$$

This equation says that the hospital continues to employ fellows until the incremental gain from adding one more fellow falls below zero.[5] The relationship is illustrated in figure five. The profit function reaches its maximum at 5 fellows. Notice that $G(5)$ – the difference in profit from employing five rather than four fellows – is positive. While $G(6)$ – the incremental profit from employing six fellows rather than five – is negative. Condition 1 picks out the maximum number of fellows, given the hospital's choice is discrete.

[4] Note that θ^H and θ^L are continuous in s. The trouble is that the hopsital is restricted to selecting an integer number of fellows. Concavity means that $\frac{\partial \theta^H}{\partial s} > 0$; $\frac{\partial \theta^L}{\partial s} > 0$; $\frac{\partial \theta^H}{\partial s \partial s} < 0$; and $\frac{\partial \theta^L}{\partial s \partial s} < 0$. The second order conditions mean that the marginal improvement in payoff from an additional fellow is decreasing. In the discrete choice case, this reduces to condition (1).

[5] For a full description of the conditions for maximization where the choice variable is discrete see Sah and Zhoa (1998).

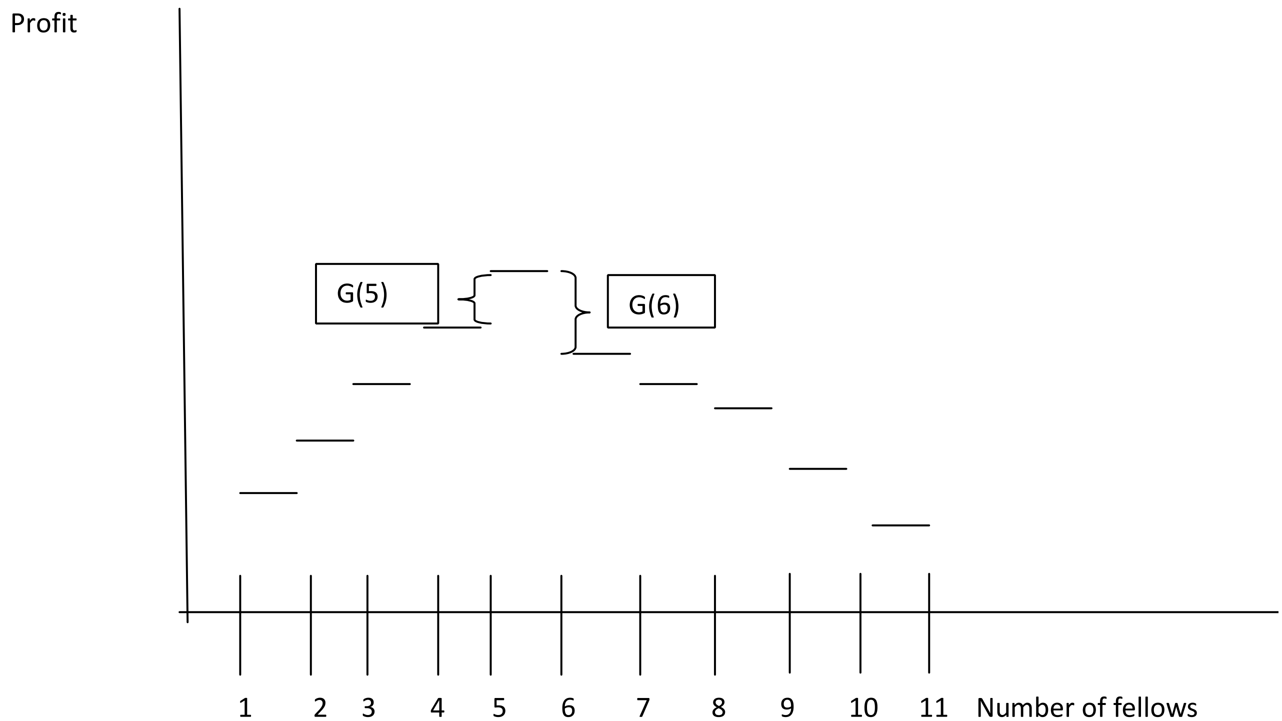

Figure 5: Graph of the G-Function

2.2.2 N-Period Model

Suppose now that the hospital lives for n periods and discounts the future at a constant rate δ. What will happen to the number of fellows employed in period one? We solve by backward induction. If the fellowship program in period $n-1$ produced at least one high type researcher, the hospital selects its newly minted fellows to maximize

$$\theta^H(s_n) - ws_n$$

Denote the solution to this equation as s_n^H. If the fellowship program failed to produce at least one high type, the hospital maximizes

$$\theta^L(s_n) - ws_n$$

Denote the solution to this equation as s_n^L. Plugging in the optimal number of fellows generates the profit associated with employing a high type and low type respectively.

$$\theta^H(s_n^H) - ws_n^H$$

$$\theta^L(s_n^L) - ws_n^L$$

To make the model interesting, assume that, no matter the time period, the hospital prefers to employ a high quality researcher when one is available (i.e., $\theta^H(s_t^H) - ws_t^H > \theta^L(s_t^L) - ws_t^L$ for all t). Take a step backward and consider the hospital's choice of fellows in period $n-1$. The expected payoff in period n is

$$V^n(s_{n-1}, s_n^H, s_n^L) = [1-(1-p)^{s_{n-1}}][\theta^H(s_n^H) - ws_n^H] + (1-p)^{s_{n-1}}[\theta^L(s_n^L) - ws_n^L]$$

In period $n-1$, there are two possibilities: either the pool of fellows in period $n-2$ produced at least one high quality type or it didn't. Denote the two possibilities as $i \in [H, L]$. At $n-1$, the hospital will maximize

$$\theta^i(s_{n-1}) - ws_{n-1} + \delta V^n(s_{n-1}, s_n^H, s_n^L) \quad (2)$$

The incremental profit function is

$$G^i_{n-1}(w, s_{n-1}) = [\theta^i(s_{n-1}) - ws_{n-1} + \delta V^n(s_{n-1}, s^H_n, s^L_n)] -$$
$$[\theta^i(s_{n-1} - 1) - w * (s_{n-1} - 1) + \delta V^n(s_{n-1} - 1, s^H_n, s^L_n)]$$

Under the same reasoning as above, at the maximum s^i_{n-1} the following condition must be true

$$G^i_{n-1}(w, s^i_{n-1}) \geq 0 \geq G^i_{n-1}(w, s^i_{n-1} + 1)$$

The solution to this problem gives a set of fellow choices, $\{s^L_{n-1}, s^H_{n-1}\}$, depending on whether the prior hospital inherited a high type or not from the prior period. Plugging these values into equation (2) gives the maximum obtainable two period value, $V^{n-1}(s_{n-2}, s^L_{n-1}, s^H_{n-1})$. This value is

$$V^{n-1}(s_{n-2}, s^L_{n-1}, s^H_{n-1}) = [1 - (1-p)^{s_{n-2}}][\theta_H(s^H_{n-1}) - ws^H_{n-1}] +$$
$$(1-p)^{s_{n-2}}[\theta_L(s^L_{n-1}) - ws^L_{n-1}] + \delta V^n(s_{n-1}, s^L_n, s^H_n)]$$

Do the same analysis all the way back to time period one. In so doing, we see that the hospital in period one wants to maximize

$$\theta^L(s_1) - ws_1 + \delta V^2(s_1, s^H_2, s^L_2)$$

The incremental profit function from this equation is

$$G_1^L(w, s_1) = [\theta^L(s_1) - ws_1 + \delta V^2(s_1, s_2^H, s_2^L)] - [\theta^L(s_1 - 1) - w*(s_1 - 1) + \delta V^2(s_1 - 1, s_2^H, s_2^L)] \tag{3}$$

The superscript, L recognizes that, in period one, we assumed that only low types were available to the hospital. So, we can restrict attention to just G_1^L (i.e., there is no G_1^H).[6] At the maximum, s_1^L, the following condition must hold:

$$G_1^L(w, s_1^L) \geq 0 \geq G_1^L(w, s_1^L + 1) \tag{4}$$

A close inspection of this condition provides the first result.

Proposition 1 *A hospital with a n-period time horizon employs at least as many and often more fellows in period one than a hospital with a one period time horizon (that is, $s_1^L \geq \widetilde{s}_1$).*

Proof:

We want to show that s_1^L - the maximum number of fellows associated with the n-period problem – is greater than or equal to $\widetilde{s}_1$ – the optimal number of fellows from the one period problem. Proving this result reduces to checking whether the following two conditions hold:

$$G_1^L(\widetilde{s}_1 + 1) > G(\widetilde{s}_1 + 1) \tag{a1}$$

$$G_1^L(\widetilde{s}_1) > G(\widetilde{s}_1) \tag{b1}$$

To see why, note that the maximization condition of the one period model implies that

[6] In the latter parts of the paper, the superscript is dropped to simplify the notation.

$G(\widetilde{s}_1+1) \leq 0$. Condition (a1) states that $G_1^L(\widetilde{s}_1+1) > G(\widetilde{s}_1+1)$. Given this inequality, there are two possible cases.

In case one, $G_1^L(\widetilde{s}_1+1) > 0$. If this is true, equation (4) teaches that $\widetilde{s}_1$ cannot be the maximum value associated with G_1^L. Moreover, because G_1^L is concave, the incremental profit function decreases as s_1 increases. Therefore, in case one, concavity implies that the maximum value, s_1^L, must be bigger than $\widetilde{s}_1$.

In the second case, $G_1^L(\widetilde{s}_1+1) < 0$ which means that $\widetilde{s}_1$ might be the maximum for the n-period problem. But to be sure, we need to know that $G_1^L(\widetilde{s}_1) > 0$. If so, then $\widetilde{s}_1$ is the maximum associated with the n-period problem (this follows from equation (4)). Condition (b1) states that $G_1^L(\widetilde{s}_1) > G(\widetilde{s}_1)$. Equation (1) tells us that $G(\widetilde{s}_1) > 0$. When condition (b1) holds, then, it must be true that $G_1^L(\widetilde{s}_1) > 0$. And so, $\widetilde{s}_1$ is the maximum in the n-period problem. The proof thus reduces to checking for condition (a1) and condition (b1).

Start with condition (a1). We know that $V^2(\widetilde{s}_1+1, s_2^H, s_2^L) = [1-(1-p)^{\widetilde{s}_1+1}][\theta^H(s_2^H) - ws_2^H] + (1-p)^{\widetilde{s}_1+1}[\theta^L(s_2^L) - ws_2^L] + \delta V^3(\cdot)$. We also know that $V^2(\widetilde{s}_1, s_2^H, s_2^L) = [1-(1-p)^{\widetilde{s}_1}][\theta^H(s_2^H) - ws_2^H] + (1-p)^{\widetilde{s}_1}[\theta^L(s_2^L) - ws_2^L] + \delta V^3(\cdot)$. Because (i) $(1-p)^{\widetilde{s}_1} > (1-p)^{\widetilde{s}_1+1}$ and (ii) the profits from employing a high type are always greater than the profits from employing a low type, it follows that $V^2(\widetilde{s}_1+1, s_2^H, s_2^L) > V^2(\widetilde{s}_1, s_2^H, s_2^L)$.
After collecting terms and rearranging equation (3), we can write $G_1^L(w, \widetilde{s}_1+1)$ as

$$G_1^L(w, \widetilde{s}_1+1) = G(w, \widetilde{s}_1+1) + \delta[V^2(\widetilde{s}_1+1, s_2^H, s_2^L) - V^2(\widetilde{s}_1, s_2^H, s_2^L)]$$

Because the second term is positive, it follows that $G_1^L(w, \widetilde{s}_1+1) > G(w, \widetilde{s}_1+1)$. As a result, condition (a1) holds.

Turn now to condition (b1). Reasoning as above, we know that $V^2(\widetilde{s}_1, s_2^H, s_2^L) = [1-$

$(1-p)^{\widetilde{s}_1}][\theta^H(s_2^H) - ws_2^H] + (1-p)^{\widetilde{s}_1}[\theta^L(s_2^L) - ws_2^L] + \delta V^3(\cdot)$ and $V^2(\widetilde{s}_1 - 1, s_2^H, s_2^L) = [1-(1-p)^{\widetilde{s}_1 - 1}][\theta^H(s_2^H) - ws_2^H] + (1-p)^{\widetilde{s}_1 - 1}[\theta^L(s_2^L) - ws_2^L] + \delta V^3(\cdot)$. Again, since (i) $(1-p)^{\widetilde{s}_1 - 1} > (1-p)^{\widetilde{s}_1}$ and (ii) employing the high type is always more profitable, it must be the case that $V^2(\widetilde{s}_1, s_2^H, s_2^L) > V^2(\widetilde{s}_1 - 1, s_2^H, s_2^L)$. Rewrite $G_1^L(w, \widetilde{s}_1)$ as

$$G_1^L(w, \widetilde{s}_1) = G(w, \widetilde{s}_1) + \delta[V^2(\widetilde{s}_1, s_2^H, s_2^L) - V^2(\widetilde{s}_1 - 1, s_2^H, s_2^L)]$$

The bracketed term must be positive, which implies that $G_1^L(w, \widetilde{s}_1) > G(w, \widetilde{s}_1)$ – condition (b1) holds, which completes the proof■

In light of this proposition, we can also derive some results about the quality of medical research in period one. The value of medical research in period one is $\theta^L(s_1)$. This value is an increasing function of the number of fellows employed. As such, the hospital with an n-period time horizon provides better research in period one than a hospital with a one-period time horizon.

The second insight from the benchmark model involves the relationship between the discount rate and the number of fellows employed. As the discount rate gets bigger (the hospital weighs future consequences more heavily), the number of fellows it employs increases. Each additional fellow provides a kicker in terms of the likelihood of finding a high quality doctor in the next round. The benefit of this kicker depends on how much the hospital cares about the future. More formally we have the following proposition.

Proposition 2 *When the hospital has an n-period time horizon, there is an increasing relationship between the number of fellows employed and the discount rate.*

Proof:

$G_1^L(w, s_1^L + 1)$ is an increasing function of δ. As δ increases the cutoff point (where the incremental gain from one more fellow is just positive) increases. This means that for a sufficiently large increase in δ, the optimal number of fellows increases$\blacksquare$

2.3 The Model with A Valuable Outside Option

Up until this point, we have assumed that senior researcher, no matter their quality, had an outside option of zero. Why would this be so? In a more realistic setup, the high quality fellows would have better outside options than the low quality fellows. Imagine that the doctor can decide to work for the hospital or work in the private sector. In the private sector, they treat patients. The value of this outside option depends on two factors: (1) the price for treatment services and (2) the number of patients the doctor can treat. As noted in the introduction, I focus on two cases. In the first subsection, the price for treatment is independent of the number of fellows pumped by the hospital into the market. In the second subsection, this assumption is relaxed. The cases correspond to situations where the research hospital is one of many small producers of medical talent and a situation where the research hospital is the single producer of medical talent in a particular field.

2.3.1 Exogenous Treatment Price

Suppose that the senior researcher can treat patients in the private sector for a price r per unit. In addition, assume that low quality doctors can treat one patient per period. High quality doctors can treat α patients per period, where $\alpha > 1$. I restrict attention to a two period model.

In period two, the hospital selects the size of its second period fellowship program to maximize

$$\theta^H(s_2) - ws_2 \qquad \text{if a high type is available}$$

$$\theta^L(s_2) - ws_2 \qquad \text{if a high type is unavailable}$$

Like above, denote the values of s_2 that solve these equations as s_2^H and s_2^L respectively. In this extension, the hospital has to pay the high skill researcher more to forgo their outside option. In period two, the profit from employing a high type and low type respectively are thus

$$\theta^H(s_2^H) - ws_2^H - \alpha r \qquad \text{if a high type is available}$$

$$\theta^L(s_2^L) - ws_2^L - r \qquad \text{if a high type is unavailable}$$

To keep the model interesting, assume that the profit from employing a high type in period two is always higher than the profit from employing a low type (the thinking, here, is that the hospital could always choose a low type if it wanted to).[7] In period one, the hospital selects the size of its first period fellowship program to maximize

$$\theta^L(s_1) - ws_1 + \delta\left([1-(1-p)^{s_1}](\theta^H(s_2^H) - ws_2^H - \alpha r) + (1-p)^{s_1}(\theta^L(s_2^L) - ws_2^L - r)\right) \quad (5)$$

Denote the solution to this equation as $\bar{s}_1$. It is easy to show the relationship between the number of fellows employed in the extension and the number of fellows employed in the benchmark model. The next proposition formalizes that result.

[7] This assumption holds so long as α is not too large, more precisely that $\alpha < 1 + \frac{(\theta^H - ws_2^H) - (\theta^L - ws_2^L)}{r}$

Proposition 3 *In an environment where the senior researcher's outside option is positive, exogenously set, and depends on the fellow's skill level, the hospital employs the same number of fellows or fewer in period one than in the n-period model (that is, $s_1^L \geq \overline{s}_1$).*

Proof:

Denote as $\overline{G}_1$ the incremental profit function associated with equation (5). To prove that $s_1^L \geq \overline{s}_1$, we need to check for the following two conditions:

$$G_1^L(w, \overline{s}_1 + 1) > \overline{G}_1(w, r, \overline{s}_1 + 1) \tag{a2}$$

and

$$G_1^L(w, \overline{s}_1) > \overline{G}_1(w, r, \overline{s}_1) \tag{b2}$$

where $\overline{s}_1$ is the optimal number of fellows derived from equation (5). Start with condition (a2). After some rearranging, we can write $G_1^L(\overline{s}_1 + 1)$ as[8]

$$G_1^L(\overline{s}_1{+}1) = \overline{G}_1(\overline{s}_1{+}1){+}\delta\left[((1-(1-p)^{\overline{s}_1+1})\alpha r + (1-p)^{\overline{s}_1+1}r) - ((1-(1-p)^{\overline{s}_1})\alpha r + (1-p)^{\overline{s}_1}r)\right]$$

Focus on the bracketed term. Because (i) $(1-p)^{\overline{s}_1} > (1-p)^{\overline{s}_1+1}$ and (ii) $\alpha r > r$, it follows that the bracketed term must be positive; and so, condition (a2) holds.

Turn now to condition (b2). After similar rearranging as above, we can write $G_1^L(\overline{s}_1)$ as

$$G_1^L(\overline{s}_1) = \overline{G}_1(\overline{s}_1){+}\delta\left[((1-(1-p)^{\overline{s}_1})\alpha r + (1-p)^{\overline{s}_1}r) - ((1-(1-p)^{\overline{s}_1-1})\alpha r + (1-p)^{\overline{s}_1-1}r)\right]$$

Like above, because (i) $(1-p)^{\overline{s}_1-1} > (1-p)^{\overline{s}_1}$ and (ii) $\alpha r > r$, the bracketed term must be positive. This means that condition (b2) holds, which completes the proof■

[8] In the remainder of the paper, I drop the arguments in G other than the number of fellows.

The idea is that investing in additional fellows increases the chance that a fellow will have a high type and thus a lucrative outside option. The gains from training, then, produce a positive externality, captured by the senior researcher (they get a bigger share of the rents from the hospital than they previously received). Anticipating the positive externality, the hospital employs fewer fellows in period one.

2.3.2 Endogenous Treatment Price

Let the demand for medical treatment be linear and equal to $A - Br$. Suppose that the only doctors treating patients come out of the fellowship program from the hospital. The supply of treating physicians in period two is thus determined by (1) the size of the period one fellowship program and (2) the realization of the random variable, X, which reflects the number of high types in the pool. For a given value of s_1 and a realization of high types X (which will be a number between 0 and s_1), the supply of treatment is

$$\begin{array}{ll} s_1 - 1 & if\ X = 0\ or\ X = 1 \\ (X-1)\alpha + [s_1 - X] & otherwise \end{array}$$

To illustrate the mechanics behind this supply schedule, consider the case where the pool fails to produce a high type ($X = 0$). In that case, there are $s_1 - 1$ low quality doctors treating in the private sector and one low quality doctor working for the hospital. Suppose that $X = 1$ (the pool produced exactly one high quality type). In that case, the hospital hires the one high quality type and the remaining $s_1 - 1$ low quality fellows work in the private sector. Suppose that X=2. In that case, one high quality type works in the hospital; one high quality type treats α patients in the private sector, the remaining $s_1 - 2$ low quality fellows treat one patient that period. Finally, suppose that the pool produces all high types

$(X = s_1)$ There, the hospital hires a high type and the remaining $s_1 - 1$ high types each treat $\alpha > 1$ patients. With this supply schedule in hand, we can derive the equilibrium price by setting supply equal to demand. And so,

$$r^* = \frac{A - (s_1 - 1)}{B} \qquad if\ X = 0\ or\ X = 1$$
$$r^* = \frac{A - [(X-1)\alpha + [s_1 - X]]}{B} \qquad otherwise$$

The equilibrium price will depend on demand parameters and three other things. First, it will depend on s_1 – the number of fellows in the program and hence the number of doctors treating patients in the private sector. Second, it will depend on X – the realized number of doctors with a high skill level. Third, it will depend on α – the number of treatment procedures the high-skilled researcher can do. As each of these values increase, there is an increased supply of treatment and, as a result, a decrease in the equilibrium price.

At the time the hospital decides on the size of its fellowship program, the value of X is unknown and hence the equilibrium price, $r^*(X, s_1)$ is also unknown. The hospital will select s_1 to maximize its expected profit, which can be written as

$$\theta^L(s_1) - ws_1 + \delta\binom{s_1}{0}(1-p)^{s_1}[\theta^L(s_2^L) - ws_2^L - r(0, s_1)] + \tag{6}$$
$$\delta\binom{s_1}{1}p(1-p)^{s_1-1}[\theta^H(s_2^H) - ws_2^H - \alpha r(1, s_1)] + ...\delta\binom{s_1}{s_1}p^{s_1}[\theta^H(s_2^H) - ws_2^H - \alpha r(s_1, s_1)]$$

Denote the solution to this equation as $\bar{\bar{s}}_1$.

By increasing the size of the training program, the hospital reduces the equilibrium price (in expectation). This results in a lower outside option, meaning that the hospital must pay less to attract the medical talent (whether low or high quality). When the treatment price is endogenous, there are thus two benefits to increasing pool size: (1) it increases the chance of getting at least one high quality researcher in the second period and (2) it drives down the outside option price, increasing the payoff whether or not the training program produces a high type. This, in turn, drives the hospital to increase the size of the pool relative to the environment where the treatment price is exogenous. More formally, we have the following final result.

Proposition 4 *In an environment where the treatment price is endogenously set, for any value of $\overline{s}_1$ such that $r > \max\{r^C(\overline{s}_1 - 1), r^C(\overline{s}_1)\}$, the hospital selects at least as many or more fellows in the first period as in the environment where the treatment price is exogenous. That is, we have $\overline{\overline{s}}_1 \geq \overline{s}_1$.*

Proof:

Let $\overline{\overline{G}}$ be the incremental profit function associated with equation (6). Applying the method from proposition 1, we need to check two conditions:

$$\overline{\overline{G}}(\overline{s}_1 + 1) > \overline{G}_1(\overline{s}_1 + 1) \tag{a3}$$

$$\overline{\overline{G}}(\overline{s}_1) > \overline{G}_1(\overline{s}_1) \tag{b3}$$

Start with condition (a3). Recall that $\overline{G}_1(\overline{s}_1 + 1)$ is defined as

$$\begin{aligned}\overline{G}_1(\overline{s}_1 + 1) \quad = \quad & G_1 + \delta[(1-(1-p)^{\overline{s}_1+1})[\theta^H(s_2^H) - ws_2^H - \alpha r] + (1-p)^{\overline{s}_1+1}(\theta^L(s_2^L) - ws_2^L - r)] \\ & -\delta[(1-(1-p)^{\overline{s}_1})[\theta^H(s_2^H) - ws_2^H - \alpha r] + (1-p)^{\overline{s}_1}(\theta^L(s_2^L) - ws_2^L - r)]\end{aligned}$$

Likewise, we can define $\overline{\overline{G}}(\overline{s}_1+1)$ as

$$\overline{\overline{G}}_1(\overline{s}_1+1) = G_1 + \delta\left[\begin{array}{c}\left(\sum_{X=1}^{\overline{s}_1+1}\binom{\overline{s}_1+1}{X}p^X(1-p)^{\overline{s}_1+1-X}[\theta^H(s_2^H) - ws_2^H - \alpha r(X,\overline{s}_1+1)\right)+ \\ (1-p)^{\overline{s}_1+1}(\theta^L(s_2^L) - ws_2^L - r(0,\overline{s}_1+1))\end{array}\right] - \delta\left[\begin{array}{c}\left(\sum_{X=1}^{\overline{s}_1}\binom{\overline{s}_1}{X}p^X(1-p)^{\overline{s}_1-X}[\theta^H(s_2^H) - ws_2^H - \alpha r(X,\overline{s}_1)\right)+ \\ (1-p)^{\overline{s}_1}(\theta^L(s_2^L) - ws_2^L - r(0,\overline{s}_1))\end{array}\right]$$

Generate a new variable, $\overline{\overline{H}}(\overline{s}_1+1)$. Do so, by replacing in $\overline{\overline{G}}_1(\overline{s}_1+1)$ all the values of $r(X,\overline{s}_1+1)$ and $r(X,\overline{s}_1)$ with $r(0,\overline{s}_1)$. Since $r(0,\overline{s}_1)$ is bigger than $r(X,\overline{s}_1)$ and $r(X,\overline{s}_1+1)$ for all values of X, it follows that $\overline{\overline{H}}(\overline{s}_1+1) < \overline{\overline{G}}(\overline{s}_1+1)$ (basically, we have increased the cost of the outside option in every state of the world). If $\overline{\overline{H}}(\overline{s}_1+1) > \overline{G}_1(\overline{s}_1+1)$, it follows that $\overline{\overline{G}}(\overline{s}_1+1) > \overline{G}_1(\overline{s}_1+1)$. Since we replaced all the different values of $r(X,\overline{s}_1)$ with a common value, we can rid ourselves of the factorials and write $\overline{\overline{H}}(\overline{s}_1+1)$ as

$$\begin{aligned}\overline{\overline{H}}_1(\overline{s}_1+1) = G_1 + {} & \\ & \delta\{(1-(1-p)^{\overline{s}_1+1})[\theta^H(s_2^H) - ws_2^H - \alpha r(0,\overline{s}_1)] + \\ & (1-p)^{\overline{s}_1+1}(\theta^L(s_2^L) - ws_2^L - r(0,\overline{s}_1)\} \\ & - \delta\{(1-(1-p)^{\overline{s}_1}[\theta^H(s_2^H) - ws_2^H - \alpha r(0,\overline{s}_1)] \\ & + (1-p)^{\overline{s}_1}(\theta^L(s_2^L) - ws_2^L - r(0,\overline{s}_1))\}\end{aligned}$$

Note that $\overline{\overline{H}}(\overline{s}_1+1)$ is equivalent to $\overline{G}_1(\overline{s}_1+1)$, when the outside option price, r equals

$r(0, \overline{s}_1)$. Move back now to $\overline{G}_1(\overline{s}_1 + 1)$ and take its derivative with respect to r

$$\frac{\partial \overline{G}_1(\overline{s}_1 + 1)}{\partial r} = -\delta \left((1 - (1-p)^{\overline{s}_1+1})\alpha - (1-p)^{\overline{s}_1+1} + (1 - (1-p)^{\overline{s}_1})\alpha + (1-p)^{\overline{s}_1}\right)$$

Rearranging and collecting terms, the derivative reduces to

$$\delta \left((1-p)^{\overline{s}_1+1}(\alpha - 1) - (1-p)^{\overline{s}_1}(\alpha - 1)\right)$$

Because $(1-p)^{\overline{s}_1} > (1-p)^{\overline{s}_1+1}$ and $\alpha > 1$, it follows that $\frac{\partial \overline{G}_1(\overline{s}_1+1)}{\partial r} < 0$. If it turns out, then, that $r > r(0, \overline{s}_1)$, it must be true that $\overline{\overline{H}}(\overline{s}_1 + 1) > \overline{G}_1(\overline{s}_1 + 1)$ There are two reasons. As noted, $\overline{\overline{H}}(\overline{s}_1 + 1)$ is equivalent to $\overline{G}_1(\overline{s}_1 + 1)$ when $\overline{G}_1(\overline{s}_1 + 1)$ is evaluated at $r(0, \overline{s}_1)$. Second, $\overline{G}_1(\overline{s}_1 + 1)$ is decreasing in r. Denote the threshold value that makes this true as $r^C(\overline{s}_1)$. Condition (a3) holds whenever $r > r^C$.

Move now to condition (b3). We can write $\overline{G}_1(\overline{s}_1)$ as

$$\begin{aligned}
\overline{G}_1(\overline{s}_1) = G_1 + \\
&\delta\{(1 - (1-p)^{\overline{s}_1})[\theta^H(s_2^H) - ws_2^H - \alpha r] + (1-p)^{\overline{s}_1}(\theta^L(s_2^L) - ws_2^L - r)]\} \\
&- \delta\{(1 - (1-p)^{\overline{s}_1-1})[\theta^H(s_2^H) - ws_2^H - \alpha r] + (1-p)^{\overline{s}_1-1}(\theta^L(s_2^L) - ws_2^L - r)\}
\end{aligned}$$

Doing the same manipulation as above, compute $\overline{\overline{H}}(\bar{s}_1)$ as

$$\begin{aligned}\overline{\overline{H}}_1(\bar{s}_1) = G_1 + \\ &\delta\{(1-(1-p)^{\bar{s}_1})[\theta^H(s_2^H) - ws_2^H - \alpha r(0,\bar{s}_1-1)] \\ &+ (1-p)^{\bar{s}_1}(\theta^L(s_2^L) - ws_2^L - r(0,\bar{s}_1-1))\} \\ &- \delta\{(1-(1-p)^{\bar{s}_1-1})[\theta^H(s_2^H) - ws_2^H - \alpha r(0,\bar{s}_1-1)] + \\ &(1-p)^{\bar{s}_1-1}(\theta^L(s_2^L) - ws_2^L - r(0,\bar{s}_1-1))\}\end{aligned}$$

Focus again on $\overline{G}_1(\bar{s}_1)$ and take its derivative with respect to r

$$\frac{\partial \overline{G}_1(\bar{s}_1)}{\partial r} = -\delta\left((1-(1-p)^{\bar{s}_1})\alpha - (1-p)^{\bar{s}_1} + (1-(1-p)^{\bar{s}_1-1})\alpha + (1-p)^{\bar{s}_1-1}\right)$$

Rearranging terms again, we see that this derivative equals

$$\delta(\left(1-p\right)^{\bar{s}_1}(\alpha-1) - (1-p)^{\bar{s}_1-1}(\alpha-1)\right)$$

which must be less than zero because $(1-p)^{\bar{s}_1-1} > (1-p)^{\bar{s}_1}$ and $\alpha > 1$. Because the derivative is negative, if $r > r(0,\bar{s}_1-1)$, it follows that $\overline{\overline{H}}(\bar{s}_1) > \overline{G}_1(\bar{s}_1)$. Denote this value $r^{CC}(\bar{s}_1-1)$. Condition (b3) holds whenever $r > r^{CC}$. As a result, condition (a3) and condition (b3) both hold whenever $r \geq \max\{r^C(\bar{s}_1-1), r^C(\bar{s}_1)\}$, which completes the proof■

2.4 Conclusion

The medical fellowship program is a unique labor market. The hospital uses fellows to enhance the research of the current senior doctors on its staff. At the same time, the hospital understands that it is training the next generation of researchers. Yet, because of constraints on capital, the hospital will not be able to hire all of the fellows as researchers. Some will go to the private sector. As a result, the hospital will have trouble capturing all the returns to its investment in the fellowship program. On the one hand, the larger the program is, the more likely the hospital is to find at least one high quality researcher in the fellow pool, pushing the hospital toward an expansive program. If, however, the high quality researcher can trade on that quality in the private sector, the hospital will be forced to pay him more to come on board. This fact reduces the benefit from having a high skilled researcher in the pool and limits the incentive to expand the pool. If the hospital's program provides the bulk of treating physicians in a specific area, this effect is muted. By expanding the program, the hospital drives down the price received for services in the private sector, reducing the value of the outside option. This, in turn, limits the amount of surplus the high quality researcher can extract from the hospital. Interestingly, the more physicians the hospital supplies to the private sector, the better its research is each period.

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.

Barro, Robert and Becker, Gary. 1989. "Fertility Choice in a Model of Economic Growth." 57 *Econometrica* 481-501.

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

Cremer, Jacques. 1986. "Cooperation in Ongoing Organizations." 101 *Quarterly Journal of Economics* 33-49.

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.

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

Diamond, Peter. A. 1965. "National Debt in a Neoclassical Growth Model." 55 *American Economics Review* 1126-1150.

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

Donini-Lenho, Fred G.and Hedrick, Hannah L. 2000. "Growth of Specializationin Graduate Medical Education." 204 *Journal of the American Medical Association* 1284-1290.

Fanti, Luciano and Spataro, Luca. 2006. "Endogenous labor supply and Diamond's (1965) model: a reconsideration of the debt role." 28 *Journal of Macroeconomics* 428-438.

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

Garicana, Luis. 2000. "Hierarchies and the Organization of Knowledge of Production." 108 *Journal of Political Economy* 874-904.

Glomm, Gerhard and Ravikumar, B. 1992. "Public versus Private Investment in Human Capital: Endogenous Growth and Income Inequality." 100 J*ournal of Political Economy* 818-834.

Grossman, Gene and Yanagawa, Noriyuki. 1993. "Asset Bubbles and Endogeneous Growth." 31 *Journal of Monetary Economics* 3-19.

Hu, Sheng. C. 1979. "Social Security, the Supply of Labor, and Capital Accumulation." 69 *American Economic Review* 274-283.

Hugman, Wallace E. and Robert E., Evenson. 2006. "Do Formula or Competitive Grant Funds Have a Greater Impacts on State Agricultural Productivity?" 88 *American Journal of Agricultural Economics* 783-798.

Jacob, Brian and Lefgren, Lars. 2007. "The Impact of Research Grant Funding on Scientific Productivity." working paper 13519, National Bureau of Economic Research.

Johnson, Erin M. 2009. "Ability, Learning and the Career Path of Cardiac Specialists." (working paper, University of California at Berkeley, Department of Economics).

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.

King, Ian. 1992. "Endogeneous Growth and Government Debt." 59 *Southern Economic Journal* 15-21.

La Croix, David De and Michel, Phillipe. 2002. A Theory of Economic Growth: Dynamics and Policy in Overlapping Generations. Cambridge Univ. Press, Cambridge UK.

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 Noncooperative Bargaining Game." The Economics of Information and Uncertainty. Chicago: University of Chicago Press for the National Bureau of Economic Research 233-258.

Nourry, Carine. 2001. "Stability of equilibria in the overlapping generations model with endogenous labor supply." 25 *Journal of Economic Dynamics and Control* 1647-1663

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

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

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

Sah, Raaj and Zhoa, Jingang. 1998. "Some Envelope Theorem Results for Integer and Discrete Choice Variables." 39 *International Economic Revie*w 623-634.

Samuelson, Paul A. 1958. "An Exact Consumption-Loan Model of Interest with or without the Social Contrivance of Money." 66 *Journal of Political Economy* 467-482.

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.

Tadelis, Stephen. 2002. "The Market for Reputations as an Incentive Mechanism." 110 *Journal of Political Economy* 854-88.

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

Tu, Ha T. and Johanna Lauer. 2008. Health Tracking Household Survey, Research Brief No. 9 Center for Studying Health System Change.