A learning theory of referrals

Damien S. Eldridge
Department of Economics and Finance
La Trobe University
Bundoora, Vic, 3086
Australia
E-mail: d.eldridge@latrobe.edu.au

November 2007

Abstract

Many service industries, including the medical and legal professions in some countries, display a gated structure. Rather than approaching a final producer directly, a consumer will first seek a referral from an intermediary. Such an industry structure might help to alleviate adverse selection problems between parties that interact infrequently. Intermediaries aggregate many short-run transactions between various consumers and a particular producer. As such, they might be able to learn a producer’s level of proficiency more rapidly than an individual consumer. However, the presence of a positive information externality means that too few consumers will seek a referral. As such, some form of regulation to encourage consumers to seek a referral might be warranted.

1 Introduction

The potential for adverse selection problems to result in market failure is well understood. Nonetheless, there are a number of ways in which the parties to a transaction might be able to reduce the impact of adverse selection. The informed party in a transaction might attempt to signal his type. Alternatively, the uninformed party might attempt to design a screening contract that induces the informed party to truthfully reveal his type. However, in a static setting, such mechanisms might only be partially successful. If the transacting parties interacted repeatedly, then reputation effects might potentially reduce the

---

*This paper is a modified version of Chapter 3 of the author’s PhD dissertation (Eldridge [13], [15]). The author would like to thank Suren Basov and Max Stinchcombe for helpful advice and comments.

1 See, for example, Akerlof ([3]), Pauly ([27]), Rothschild and Stiglitz ([33]), Stiglitz and Weiss ([39]) and Wilson ([41]).

2 Useful surveys of the signalling and screening literature include Hirshleifer and Riley ([20], Chapter 11), Kreps ([22], Chapter 17), Mas-Colell et al ([26], Chapter 13), Riley ([32]) and Stiglitz and Weiss ([40]). The seminal papers in this literature include Banks and Sobel ([4]), Cho and Kreps ([8]), Riley ([30], [31]), Rothschild and Stiglitz ([33]), Spence ([36]) and Wilson ([41]).
incidence of adverse selection below the level that would occur in a static setting. Unfortunately, there are many occasions in which parties to a transaction do not repeatedly interact. In the absence of repeated interaction, reputation cannot be relied upon to overcome adverse selection. As such, we might expect adverse selection problems to be particularly severe in markets characterised by few or infrequent interactions between the trading parties. Is consumer protection regulation the only safeguard available in such settings or can institutions be devised that might capture the benefits of repeated interaction?

A gated industry structure might provide one possible solution to potential adverse selection problems between parties that interact infrequently. An industry has a gated structure when consumers seek a referral to a producer from an intermediary, rather than accessing the services of a producer directly. These intermediaries aggregate many short-run transactions between various consumers and a particular producer. This might enable an intermediary to learn the producer’s level of proficiency more rapidly than an individual consumer. A gated industry structure is observed in some professional service industries, including the medical and legal professions in some countries. In the medical industries of some Commonwealth countries, including Australia, it is unusual for patients to visit a specialist without first obtaining a referral from a general practitioner (GP). Indeed, while it is permissible for a patient to be treated by a specialist without a referral in Australia, there are financial incentives offered to patients who obtain a referral for treatment by a specialist. In most medical specialties, patients will be reimbursed a larger portion of their treatment costs under the Medicare system in Australia if they obtain a referral before seeking treatment. This raises two interesting questions. Why do we observe a gated structure for some professional service industries? Why might government regulation be required to support this gated industry structure? In this paper we provide an answer to both of these questions. As an aid to exposition, we will focus on the health care example.

This paper is organised as follows. First, we outline a competitive model of a health care market in which treatment outcomes depend on the ability of the treating specialist, which is private information. We analyse the outcomes in a static version of this health care market. In particular, we characterise the conditions under which the market will fail to exist because of the adverse selection problem. We then proceed to show that if this static market is repeated an infinite number of times, the resulting dynamic market is less likely to fail to exist than the static market. Following this, we consider the impact of introducing a gated structure to the dynamic health care market. This further reduces the potential for the health care market to fail to exist. A comparison between these three versions of the health care market is then provided. This comparison illustrates the benefits of both repetition and the presence of intermediaries. Following this, we show that the benefits of the gated structure might not be achievable without government intervention. The reason for this is the presence of a positive information externality. Finally, we conclude by comparing the results of this paper with those that we obtained in Eldridge (14) and with some features of actual health care markets.

3Commonwealth countries in which many patients obtain a referral from a general practitioner before seeking the services of a specialist include Australia ([28], p. 421; [10], part 2, p. 3), New Zealand ([34], section 2, p. 14) and the United Kingdom ([7]).

4See, for example, Commonwealth of Australia ([10], part 2, p. 3).
2 A competitive model of health care markets

Consider an economy with three groups of agents who live forever. These groups are patients, general practitioners (GPs) and medical specialists. Let patients be indexed by \( i \in \{1, 2, \ldots, I\} \), GPs by \( j \in \{1, 2, \ldots, J\} \) and specialists by \( k \in \{1, 2, \ldots, K\} \). We will assume that there are an infinite number of patients \((I \to \infty)\) and specialists \((K \to \infty)\), but only a finite number of GPs \((J < \infty)\). Patients are either well \((d = 0)\) or sick \((d = 1)\). In each period, a patient is randomly assigned a disease state, \( d \in \{0, 1\}\). Following this, each patient can choose whether or not to seek treatment if he is sick. Treatment can sometimes result in a cure, improving the patient’s health status for that period. The probability that a sick patient is cured by treatment increases with the ability of the treating specialist. Patients can seek a referral to the specialist from a GP if they believe that this will increase the probability that they are treated by a high ability specialist. Both referrals and treatments come at a price. For budget-constrained patients, the benefits of an increased probability of good health need to be weighed against the foregone consumption of other goods that expenditure on health care entails. We will assume that patients visit neither a GP nor a specialist when they are healthy.\(^5\)

All agents in this economy are price takers who behave as though the existing prices are exogenously specified. We will focus on stationary equilibria for this economy, so that prices don’t change over time. The price per referral from any GP is \( w \), while the price per treatment from a medical specialist is \( r \). We will assume throughout that specialist ability is neither observable nor verifiable by outside parties, although it may be learned by patients and GPs that interact with the specialist. As such, the treatment price cannot vary with specialist ability.

For payoff purposes, time is assumed to be discrete in this economy. Time periods are indexed by \( t \in \{0, 1, 2, \ldots\} \), with payoffs occurring at the end of each period. In each period, the market opens and the agents interact within the market. Note that not all agents move at once in the market. The market process involves sequential moves by various agents. Thus the timing of the moves in the market process is important. We will maintain the assumption that time is discrete and index time within a period by \( s \in \{0, 1, 2, \ldots\} \). In this fashion, each point in time can be given a unique time stamp of the form \((t, s) \in \{0, 1, 2, \ldots\}^2 = \mathbb{Z}^2_+\).

2.1 The timing of the market in each period

Prior to the beginning of the first stage game, at time \( t = -1 \), Nature selects an ability level for each potential specialist as a sequence of independent draws from a common distribution. Any given specialist has either high ability \((\lambda = 1)\) or low ability \((\lambda = 0)\). Each specialist’s ability level, which is fixed for all time, is observed by nobody except for that specialist. However, it is common knowledge that the probability that any given specialist has a high level of ability is given by \( \mu \in (0, 1) \).

\(^5\)If the equilibrium prices for referrals and treatment are positive, this assumption is not needed. Even if these prices are zero, we could avoid making this assumption by introducing an opportunity cost of time (perhaps in the form of foregone leisure) into the model.
At the beginning of each period \((s = 0)\), Nature randomly chooses a disease state for each patient, \(d_i \in \{0, 1\}\). The disease state for each patient is chosen as a random draw from some common distribution, \(\Pi\). The probability that any given patient is sick in any given period is \(\pi \in (0, 1)\), while the probability that a patient is well in any given period is \((1 - \pi)\). The distribution from which these disease states are drawn is common knowledge.

At \(s = 1\), having observed their disease state, patients choose whether or not to seek treatment and, if they seek treatment, whether or not to seek a referral from their GP. If they seek a referral, they choose which GP to visit at \(s = 2\). At \(s = 3\), GP’s choose the specialists to which they will refer their patients. At this point in the stage game, any patients who chose top self-refer at \(s = 2\) will also choose the specialist from whom they will treat. We will assume that GPs follow up on the outcomes from treatment of any of the patients they refer. In this fashion, the GP knows the entire history of outcomes for each of his previous referrals at the start of each period. This allows him to use this information when making his current referral decisions.

Following this, at \(s = 4\), specialists treat each patient that has been referred to them. Finally, at \(s = 5\), Nature chooses whether or not each patient is cured. If a patient is cured, he will have good health in that period \((h = 1)\), while if the patient is not cured, he will have bad health \((h = 0)\). We will assume that treatment by a high ability specialist always results in a patient being cured, while treatment by a low ability specialist never results in a patient being cured. Furthermore, any patient who chose not to seek treatment will not be cured.

### 2.2 Player objectives

Every agent in this game is assumed to maximise the discounted present value of a sequence of per-period von Neuman Morgernstern expected utility functions. Furthermore, they all have a common rate of time preference, represented by the stationary discount factor \(\delta \in [0, 1)\). Thus differences in the preferences of the three groups of agents arise from differences in their per-period preferences. These are outlined below.

#### 2.2.1 Patients

Patients all have identical per-period preferences defined over their expenditure on health care \((p)\) and their health state \((h)\). These preferences may be represented by a quasi-linear per-period Bernoulli utility function of the form

\[
u(h_t, p_t) = B(h_t) - p_t,
\]

where \(B(0)\) is normalised to zero and \(B(1) = B > 0\).

The health state in each period is a random variable and may vary across patients. It depends on whether or not the patient is sick, whether or not treatment is sought and, if so, the ability of the treating specialist. If a sick patient receives treatment from a high ability specialist, he will definitely be cured. If a sick patient receives treatment from a low ability specialist, he will definitely not be cured. Since each patient knows his disease status before having to make any decisions about treatment, the probability of good health in period \(t\) is given by...
t =
\begin{align*}
1 & \quad \text{if either } d = 0 \text{ or high ability treatment is received with certainty when } d = 1; \\
\mu & \quad \text{if } d = 1 \text{ and treatment is sought from a specialist whose ability is not known;} \\
0 & \quad \text{if } d = 1 \text{ and either low ability treatment is received or no treatment is sought.}
\end{align*}

Expenditure on health care in any given period may also vary across patients. It will depend on whether or not the patient seeks treatment and, if so, whether or not the patient also seeks a referral. We will assume that patients do not seek a referral if they do not also desire treatment. Thus a patient’s expenditure on health care is given by
\[
p_t =
\begin{cases}
0 & \quad \text{if neither treatment nor referral is sought;} \\
r & \quad \text{if treatment is sought without a referral;} \\
w + r & \quad \text{if both treatment and referral are sought.}
\end{cases}
\]

This allows us to express a patient’s per-period expected utility as
\[
E u(h_t, p_t) = \theta_t B - p_t.
\]

### 2.2.2 General practitioners

GP's are assumed to be risk-neutral. As such, they maximise their expected profits. The Bernoulli utility function that represents their per-period preferences is simply their per-period profit. Let \( n_{j,t} \) denote the number of referrals a particular GP makes in period \( t \). Assuming that they have a constant marginal cost of \( C_{GP} \) per referral and no fixed costs, their per-period profits are
\[
\Pi_j = n_{j,1}(w - C_{GP}),
\]
where we have dropped the time subscripts for convenience.

Note that \( n_{j,1,t} \) is a random variable if a GP has a finite patient pool of size \( n_{j,t} \). However, when GP has an infinite patient pool, this source of uncertainty disappears. The reason for this is that \( n_{j,1,t} \) converges almost surely to \( \pi n_{j,t} = \infty \) as \( n_{j,t} \to \infty \).

**Proposition 1** \( n_{j,1,t} \) converges almost surely to \( \pi n_{j,t} = \infty \) as \( n_{j,t} \to \infty \).

**Proof.** First, note that
\[
n_{j,1} = \frac{n_{j,1}}{n_j} n_j = \alpha_j n_j,
\]
Furthermore,
\[
\alpha_j = \frac{n_{j,1}}{n_j} = \frac{1}{n_j} \sum_{i(j)=1}^{n(j)} 1_{i(j),1},
\]
where \( 1_{i(j),1} \) is an indicator variable that takes the value of one if the patient in question is sick and the value zero otherwise. Note that GP \( j \) has a patient pool consisting of \( n_j \) patients, including those that do not need the GP’s services in the current period. These patients are indexed by \( i_j \in \{1, 2, \cdots, n_j\} \). Note that
each of these indicator variables is a Bernoulli random variable that takes on
the value one with probability \( \pi \) and zero otherwise. As such, \( \{1_{ij}, i}\}_{i=1}^{n(j)} \) is a
sequence of independent and identically distributed Bernoulli random variables
in which

\[
E(1_{ij,1}) = \pi(1) + (1-\pi)(0) = \pi \quad \text{for all } i_j \in \{1, 2, \cdots, n_j\}.
\]

Thus, from the strong law of large numbers\(^6\), we know that

\[
\Pr\left( \lim_{n(j) \to \infty} \alpha_j = \pi \right) = 1.
\]

This allows us to conclude that \( \alpha_j \) converges almost surely to \( \pi \). Finally, note
that

\[
\lim_{n(j) \to \infty} n_j = \infty.
\]

Thus we can conclude that

\[
n_{j,1} = \alpha_j n_j \xrightarrow{a.s.} (\pi)(\infty) = \infty.
\]

Hence we know that \( n_{j,1,t} \) converges almost surely to \( \pi n_{j,t} = \infty \) as \( n_{j,t} \to \infty \).

\[\text{\textbullet}\]

2.2.3 Medical Specialists

Like GPs, medical specialists are assumed to be risk-neutral. As such, a spe-
cialist’s per-period Bernoulli utility function is simply his profit. This profit
will depend on the treatment price \( r \) and the cost of constant marginal cost
of treatment \( C_S \). We will assume that there are no fixed costs of treatment.
Each specialist’s per-period per-patient profit is given by \( r - C_S \). Let \( n_{k,t} \)
denote the number of patients a particular specialist treats in period \( t \). Given this, the
specialist’s profit in period \( t \) is simply

\[
\Pi_{k,t} = n_{k,t}(r - C_S).
\]

In each period, specialists observes the number of patients seeking treatment
from them before actually treating any patients. As such, the only uncertainty
that affects medical specialists relates to the number of patients that will seek
their treatment services in future periods. While this may be a function of the
outcomes that result from their current and past treatment of patients, there is
nothing they can do to influence these outcomes. Thus specialists will simply
maximise their per-period profits.

\(^6\)See Billingsley ([5], pp. 85-86) for a discussion of the strong law of large numbers. Note
that when GP patient pools are finite, the number of members in a GP’s patient pool is an
integer. As such, when we take the limit as this number approaches infinity, we are restricting
our attention to the set of natural numbers. In effect, there is a one-to-one correspondence
between each member of a GP’s patient pool and each element of the set of natural numbers
when that GP has an infinite patient pool. As such, each GP has a countable number of
patients. This ensures that the standard version of the strong law of large numbers applies
in the model considered in this paper. If we had assumed that each GP had a continuum of
patients instead of a countably infinite number of patients, we would have needed to use the
techniques mentioned in Judd ([21]). The reason for this is that each GP would have had an
uncountable number of patients in that case.

6
3 Static outcomes in competitive health care markets

Before analysing the dynamic model of competitive health care markets, it is useful to consider what would happen in the absence of any repetition whatsoever. To do this, we will initially assume that patients can be afflicted with the disease at most once. As such, patients will only need the services of medical specialist at most once. We will focus on a representative patient \((i)\) whose has the disease and a representative specialist \((k)\). We will assume that all agents in this economy are price takers and that all prices are exogenously determined.

Specialist will receive the treatment price \((r)\) from each patient that they treat. However, they will also incur a treatment cost equal to \(C_S\) for each patient that they treat. This treatment cost is independent of their ability.

**Proposition 2 (The participation constraint for specialists):** Specialists will offer their treatment services if and only if \(r \geq C\).

**Proof.** Medical specialists are profit maximisers. They can guarantee themselves zero profits by refusing to treat any patients. As such, they will only treat patients if the profit per patient is at least zero. This requires that the treatment price either matches or exceeds the cost of treatment for each patient.

Now consider a sick patient. If the patient is cured, then he will be in good health \((h = 1)\) for the remainder of the current period. This will yield him benefits equal to \(B(h) = B(1) = B > 0\). If the patient is not cured, he will be in bad health \((h = 0)\) for the remainder of the current period. This yields him benefits equal to \(B(h) = B(0) = 0\). If the patient is to have any chance of being cured of this disease, he will require treatment by a medical specialist. There are two types of medical specialists, high ability specialists \((\lambda = 1)\) and low-ability specialists \((\lambda = 0)\). If the patient is treated by a high ability specialist, he is guaranteed to be cured. If the patient is treated by a low ability specialist, he is guaranteed not to be cured. Unfortunately, each specialist’s ability is private information, known only to that specialist. It is common knowledge, however, that the probability of any given specialist having high ability is \(\mu \in (0, 1)\).

**Proposition 3 (The participation constraint for patients):** Patients will seek treatment if and only if \(r \leq \mu B\).

**Proof.** Since the participation constraint for specialists is independent of their ability, patients with rationally believe that the probability of a cure following treatment is equal to the probability that a specialist has high ability. As such, if a patient obtains treatment, his expected utility is:

\[
EU_i(\text{treatment}) = \mu B + (1 - \mu)0 - r = \mu B - r.
\]

If a patient does not obtain treatment, he will neither be cured nor will he have to pay the treatment price. As such, his expected utility will be zero. Thus a patient will seek treatment if and only if \(\mu B - r \geq 0\). This requires that \(r \leq \mu B\).

The outcomes in this static health care market will vary with the treatment price. If the price is too low, no specialists will offer their treatment services.
As such, no patients will be cured. If the treatment price is too high, no patients will seek treatment and hence no patients will be cured. In both cases, patients receive zero expected utility and specialists receive zero profit. There will sometimes be an intermediate range of prices in which all patients will seek treatment and all specialists will offer their treatment services. In these cases, patients will receive non-negative expected utility and specialists will earn non-negative profits. However, some patients will be disappointed with the outcome of their treatment. These are the patients that will have been unfortunate enough to be treated by a low ability specialist.

**Proposition 4** (Static market existence): If \( B < C_S \), then the set of prices at which both patients demand treatment and specialists supply treatment is empty. If \( B \geq C_S \), then the set of prices at which both patients demand treatment and specialists supply treatment is non-empty.

**Proof.** Both the specialist participation constraint and the patient participation constraint are satisfied if and only if \( C_S \leq r \leq \mu B \). This clearly requires that \( C_S \leq \mu B \). Thus, if \( C_S > \mu B \), there are no values for the treatment price that will satisfy both participation constraints. If \( C_S = \mu B \), then there is a unique value for the treatment price that will satisfy both participation constraints. This value is \( r = C_S = \mu B \). Finally, if \( C_S < \mu B \), then there is a range of values for the treatment price that will satisfy both participation constraints. These are \( r \in [C_S, \mu B] \).

**Proposition 5** (Static market outcomes): If the health care market exists, then patients will receive non-negative expected utility and specialists will make non-negative profits. However, some patients might not be cured following treatment.

**Proof.** If the health care market exists, then \( C_S \leq r \leq \mu B \). This means that

\[
EU_i = \mu B - r \geq \mu B - \mu B = 0.
\]

It also means that

\[
\Pi_{k(d)} = r - C_S \geq C_S - C_S = 0.
\]

Thus patients receive non-negative expected utility and specialists receive non-negative profits. However, since both high ability specialists and low ability specialists are prepared to offer their treatment services for this range of prices, some patients might have sought treatment from a low-ability specialist. Any such patients will not be cured.

### 4 Dynamic outcomes without general practitioners

Suppose we now play an infinitely repeated version of the stage game in the absence of GPs. In this version of the dynamic model, any patient that wants to be treated by a specialist needs to seek the services of a specialist without a referral. Before observing his disease state in period zero, a patient’s lifetime expected utility is

\[
J_i = \sum_{t=0}^{\infty} \delta^t \pi EU_{i,t}(sick) + \sum_{t=0}^{\infty} \delta^t (1 - \pi) EU_{i,t}(well).
\]
The expected utility that patients receive in periods when they are well is not affected by their treatment choices when they are sick. As such, we can ignore these terms when considering the impact of treatment decisions on a patient’s lifetime expected utility. Given this, for the remainder of this paper we will only focus on the payoffs that a patient receives in periods when he is sick.

The first time a patient is afflicted with the disease, he will not have any information about the ability of any of the specialists. As such, he may as well randomly select a specialist from whom to seek treatment if he decides to seek treatment.

**Proposition 6 (Treatment payoff):** If it is optimal for a patient to seek treatment when he is first afflicted with the disease, then it will be optimal for him to seek treatment whenever he is afflicted with the disease. Furthermore, his lifetime expected utility, net of at that point in time is

\[
V = \left( \frac{(1 - \delta) + \delta \pi}{(1 - \delta)^2 + (1 - \delta) \mu \delta} \right) \mu B - \left( \frac{(1 - \delta) + \mu \delta \pi}{(1 - \delta)^2 + (1 - \delta) \mu \delta} \right) r.
\]

**Proof.** The patient does not know the ability of whichever specialist treats him the first time he is afflicted with the disease. As such, his expected utility in that period is simply \(\mu B - r\). Following treatment, the patient is either cured or not cured. As such, he learns the ability of the treating specialist. If he is cured, then he knows the treating specialist has a high level of ability. If it is ever optimal to seek treatment, then it must be optimal to do so when you know that you will be cured. As such, the patient will seek treatment from that specialist whenever he is sick in future periods. If the patient is not cured in the current period, then he knows the treating specialist has a low level of ability. As such, he will never seek treatment from that specialist again. However, he might still get sick in some future periods. In the next such period, he will need to start from scratch. Since there an infinite number of potential medical specialists, the problem he will face in that period will be identical to the current one. As such, the lifetime expected utility from that period onwards will be identical to the lifetime expected utility in the current period. Hence, the patient’s lifetime expected utility the first time he gets sick, prior to treatment, must satisfy the following equation:

\[
V = (\mu B - r) + \mu \sum_{t=1}^{\infty} \delta^t \pi (B - r) + (1 - \mu) \sum_{t=1}^{\infty} \delta^t (1 - \pi)^{t-1} V.
\]

Note that this equation can be rewritten as

\[
V = (\mu B - r) + \mu \delta \sum_{t=0}^{\infty} \delta^t \pi (B - r) + (1 - \mu) \delta \sum_{t=0}^{\infty} \delta^t (1 - \pi)^t \pi V,
\]

which becomes

\[
V = (\mu B - r) + \mu \delta \left( \frac{\pi (B - r)}{1 - \delta} \right) + (1 - \mu) \delta \left( \frac{\pi V}{1 - \delta + \delta \pi} \right).
\]

Solving this equation for the patient’s lifetime expected utility \(V\) yields:

\[
V = \left[ \frac{(1 - \delta + \delta \pi)^2}{(1 - \delta + \delta \pi \mu)(1 - \delta)} \right] \mu B - \left[ \frac{(1 - \delta + \delta \pi)}{(1 - \delta)} \right] r.
\]
While we have characterised a patient’s lifetime expected utility if he chooses to seek treatment, we have not yet established the conditions under which seeking treatment will be optimal. In order to derive these conditions, we first need to lifetime expected utility of a sick patient who does not seek treatment.

**Proposition 7 (Non-treatment payoff):** If it is optimal for a patient not to seek treatment when he is first afflicted with a disease, then it will never be optimal for him to seek treatment. Furthermore, his lifetime expected utility at that point in time is zero.

**Proof.** If the patient does not seek treatment when he is first afflicted with the disease, then he will neither be cured in that period nor incur any medical expenses in that period. As such, his expected utility in that period is zero. Furthermore, since he did not seek treatment, he will not learn anything about the ability of any of the medical specialists. As such, he will face an identical problem the next time he gets sick, assuming such an event occurs. If it is optimal for him not to seek treatment the first time he gets sick, it must therefore be optimal for him not to seek treatment the next time he gets sick. Thus, by mathematical induction, if it is not optimal for a patient to seek treatment the first time he is sick, it will never be optimal for him to seek treatment when he is sick. The lifetime expected utility of such a patient will be zero.

Now that we have characterised the payoffs to a patient, both when he chooses to seek treatment and when he does not, we are in a position to derive the conditions under which it will be optimal for him to seek treatment when he is sick.

**Proposition 8 (Participation constraint for patients):** A sick patient will seek treatment in this dynamic health care market if and only if

\[ r \leq \left[ \frac{(1 - \delta + \delta \pi)}{(1 - \delta + \delta \pi \mu)} \right] \mu B. \]

**Proof.** We have already established that a patient will either always seek treatment when he sick or never do so. Furthermore, we have calculated the lifetime expected utility at the point in time where the patient first discovers that he sick for both of these cases. Thus we know that a sick patient will choose to seek treatment if and only if his lifetime expected utility from doing so is at least as large as his lifetime expected utility from not seeking treatment. This requires that

\[ \left[ \frac{(1 - \delta + \delta \pi)^2}{(1 - \delta + \delta \pi \mu)(1 - \delta)} \right] \mu B - \left[ \frac{(1 - \delta + \delta \pi)}{(1 - \delta)} \right] r \geq 0. \]

This inequality can be rearranged to obtain

\[ r \leq \left[ \frac{(1 - \delta + \delta \pi)}{(1 - \delta + \delta \pi \mu)} \right] \mu B. \]

Thus patients will seek treatment whenever the treatment price is not too high.
It is worth noting that the range of treatment prices for which patients will be willing to seek treatment in this dynamic health care market is larger than that in a static health care market. This makes intuitive sense because the benefits from seeking treatment in the dynamic market are larger than they are in the static market. As well as the expected benefits from treatment in the current period, which occur in both markets, patients who seek treatment will also learn the ability level of the treating specialist. While this has no value in a static model, it yields positive expected utility in a dynamic market.

**Proposition 9** *(The benefits of repetition)*: The maximum treatment price at which a patient will seek treatment is higher in a dynamic market than it is in a static market.

**Proof.** Let $\tilde{r}_d$ denote the maximum price at which patients will seek treatment in a dynamic market and $\tilde{r}_s$ denote the maximum price at which patients will seek treatment in a static market. These prices are given by least upper bounds of the patient participation constraints in each model. Note that

$$\tilde{r}_d - \tilde{r}_s = \left[ \frac{(1 - \delta + \delta \pi)}{(1 - \delta + \delta \pi \mu)} \right] \mu B - \mu B,$$

which can be rearranged to obtain

$$\tilde{r}_d - \tilde{r}_s = \left[ \frac{(1 - \mu) \delta \pi}{(1 - \delta + \delta \pi \mu)} \right] \mu B > 0.$$

Thus we know that $\tilde{r}_d > \tilde{r}_s$. 

There is nothing that specialists can do to influence treatment outcomes, treatment costs or treatment prices. This means that repetition does not affect their participation decisions. As such, the participation constraint facing specialists in dynamic health care markets will be the same as that facing them in static health care markets. We are now in a position to characterise the conditions under which a dynamic health care market will exist. As with a static health care market, the outcomes in this dynamic health care market will vary with the treatment price. If the price is too low, no specialists will offer their treatment services. As such, no patients will be cured. If the treatment price is too high, no patients will seek health care treatment and hence no patients will be cured. In both cases, patients receive zero expected utility and specialists receive zero profit. There will sometimes be an intermediate range of prices in which all patients will seek treatment and all specialists will offer their treatment services. In these cases, patients will receive non-negative lifetime expected utility and specialists will earn non-negative profits. However, some patients will be disappointed with the outcome of their treatment on at least one occasion. These are the patients that will have been unfortunate enough to be treated by a low ability specialist in at least one period.

**Proposition 10** *(Dynamic market existence)*: If $\tilde{r}_d < C_S$, then the set of prices at which both patients demand treatment and specialists supply treatment is empty. If $\tilde{r}_d \geq C_S$, then the set of prices at which both patients demand treatment and specialists supply treatment is non-empty.
Proof. Both the specialist participation constraint and the patient participation constraint are satisfied if and only if $C_S \leq r \leq \hat{r}_d$. This clearly requires that $C_S \leq \hat{r}_d$. Thus, if $C_S > \hat{r}_d$, there are no values for the treatment price that will satisfy both participation constraints. If $C_S = \hat{r}_d$, then there is unique value for the treatment price that will satisfy both participation constraints. This value is $r = C_S = \hat{r}_d$. Finally, if $C_S < \hat{r}_d$, then there is a range of values for the treatment price that will satisfy both participation constraints. These are $r \in [C_S, \hat{r}_d]$.

**Proposition 11 (Dynamic market outcomes):** If the health care market exists, then patients will receive non-negative expected utility and specialists will make non-negative profits. However, some patients might not be cured following treatment.

Proof. If the health care market exists, then the participation constraints of both patients and specialists must be satisfied. Thus patients receive non-negative expected utility and specialists receive non-negative profits. However, since both high ability specialists and low ability specialists are prepared to offer their treatment services for this range of prices, some patients might have sought treatment from a low-ability specialist on at least one occasion when they were sick. Any such patients will not have been cured on those occasions.

We showed earlier that the repetition present in this dynamic health care market expands the set of treatments prices for which patients will be willing to seek treatment compared to the set of such prices in a static health care market. Since the set of treatment prices for which specialists will offer their services is the same in both markets, this means that there is a larger set of circumstances in which a dynamic health care market will exist than in which a static health care market will exist.

**Proposition 12 (The relationship between static market existence and dynamic market existence):** A dynamic health care market will exist whenever a static health care exists. Furthermore, a static health care market will not exist whenever a dynamic health care market does not exist. However, there are some cases where a dynamic health care market will exist but a static health care market will not exist.

Proof. First, we will show that the existence of a static health care market implies the existence of a dynamic health care market. We have already established that a static health care market will exist if and only if $C_S \leq \hat{r}_s$. We have also already established that $\hat{r}_d > \hat{r}_s$. Thus, if a static health care market exists, we know that $C_S < \hat{r}_d$. Thus the condition that guarantees the existence of a dynamic health care market ($C_S \leq \hat{r}_d$) is satisfied. Thus the existence of a static health care market does indeed imply the existence of a dynamic health care market. Now we show that the non-existence of a dynamic health care market implies the non-existence of a static health care market. If a dynamic health care market does not exist, then we know that $C_S > \hat{r}_d$. Since $\hat{r}_d > \hat{r}_s$, this means that $C_S > \hat{r}_s$ as well. This means that, if a dynamic health care market cannot exist, then nor can a static health care market. Finally, we will show that there are some cases in which a dynamic health care market will exist, but a static health care market will not exist. Suppose that $\hat{r}_s < C_S \leq \hat{r}_d$. 

12
In this case, the necessary and sufficient condition for the existence of a static health care market is not satisfied, but the necessary and sufficient condition for the existence of a dynamic health care market is satisfied. Hence there are cases where a dynamic health care market can exist but a static health care market cannot exist.

5 Dynamic outcomes with general practitioners

Having established what happens in a dynamic health care market without general practitioners, we are now in a position to analyse the impact of introducing them. Suppose, for the moment, that access to treatment by medical specialists is subject to regulation. Specifically, we will assume that all sick patients will need to seek a referral before obtaining treatment in the first period (period zero). In all subsequent periods, sick patients will be able to choose whether or not to seek a referral before obtaining treatment. A regulation along these lines is needed because of the presence of an information externality. This issue is discussed in more detail later in this paper. Market outcomes in period zero will be deferred until later in this paper as well. In this section, we will focus on outcomes in this form of a dynamic health care market after period zero has finished.

Suppose that there are a finite number of GPs, each of whom has an infinite patient pool in period zero of this dynamic health care market. In these circumstances, each GP must have an infinite number of sick patients in period zero.

Proposition 13 (Patient numbers): Each GP has and infinite number of sick patients in period zero.

Proof. We showed earlier that \( n_{j,d} \) converges almost surely to \( \pi_d n_j = \infty \) as \( n_{j,t} \to \infty \). Given that all GPs have infinite patient pools in period zero, they will almost surely have an infinite number of sick patients in that period.

Since each GP has an infinite number of sick patients in period zero, he can refer a single patient to each of an infinite number of medical specialists. As such, every GP will find at least one medical specialist who has high ability.

Proposition 14 (GP learning): In period zero, every GP will find at least one medical specialist who has high ability.

Proof. Recall that any given specialist has high ability with probability \( \mu \) and low ability with probability \( 1 - \mu \). Furthermore, recall that GPs observe the treatment outcomes for all of the patients for whom they provide a referral. Since the ability level of each specialist is perfectly revealed by the outcome of any treatment that they provide, a GP will learn the ability level of any specialist to whom he refers at least one patient. Since the GP has an infinite number of sick patients, he can refer a single patient to each of an infinite number of medical specialists. The probability that at least one of these specialists has high ability is simply one minus the probability that none of the specialists who treat a patient referred by the GP has high ability. This is given by

\[
\Pr \left\{ \#_{j,k}(\lambda = 1) > 0 \right\} = 1 - \lim_{n(j,1) \to \infty} \left( \mu^{n(j,1)} \right) = 1 - 0 = 1.
\]
Thus, in period zero, each GP will find at least one medical specialist who has
high ability.

Thus it is possible for every GP to find at least one high ability specialist
in period zero. Since the ability of specialists is fixed for all time prior to
the opening of the dynamic health care market in period zero, every GP can
guarantee a patient that he will be cured if he seeks a referral from that GP in
any time period after period zero.

A potential problem with referrals is that GPs might have an incentive to
refer patients to a low ability specialist. The reason for this is that patients
will no longer need a referral after they learn the identity of a high ability
specialist. If GPs are earning positive profits on each referral they make, they
might attempt to induce further demand for services by initially making poor
referrals. However, patients can deter such a strategy by threatening to dump
any GP who refers them to a low ability specialist after period zero. This will
remove any incentive that GPs might have to make poor referrals after period
zero.

Proposition 15 (GP incentive compatibility constraint): General practitioners
cannot profit by referring patients to specialists who have low ability after period
zero.

Proof. If all consumers employ a strategy that involves never again using the
referral services of a GP who refers them to a low ability specialist after period
zero, then GPs will not gain anything by referring any patient to a low ability
specialist after period zero. As such, after period zero, GPs will be indifferent
between referring patients to high ability specialists and referring them to low
ability specialists. In these circumstances, it is reasonable to assume that GPs
will refer patients to high ability specialists after period zero.

In order to analyse market existence and market outcomes under these cir-
cumstances, we need to consider the behaviour of two groups of consumers. The
first group of consumers are those that are fortunate enough to have been treated
by a high ability specialist in period zero. The second group of consumers con-
sists of all consumers who are not members of group one. This includes both
patients who were not sick in period zero, patients who were sick in period zero
but were unfortunate enough to be treated by a low ability specialist and pa-
tients who were sick but chose not to seek treatment in that period. The lifetime
utility after period zero will be different for these two groups of patients. The
first time a patient in group one gets sick after period zero, his lifetime expected
utility if he seeks treatment will be

\[ V_{i,1} = (B - r) + \sum_{t=1}^{\infty} \delta^t \pi (B - r) = \frac{1 - \delta + \delta \pi}{1 - \delta} (B - r). \]

If a patient in group two seeks a referral after period one, he knows that he will
be referred to a high ability specialist. As such, he will learn the identity of a
high ability specialist. Thus, the first time a patient in group two gets sick after
period zero, his lifetime expected utility if he seeks a referral and treatment is

\[ V_{i,2} = (B - r - w) + \sum_{t=1}^{\infty} \delta^t \pi (B - r) = \frac{1 - \delta + \delta \pi}{1 - \delta} (B - r) - w. \]
With these lifetime expected utilities in hand, we can characterise the circumstances under which members of each will choose to seek treatment when they are sick.

**Proposition 16 (Participation constraint for informed patients):** A patient who knows the identity of a high ability medical specialist will treatment whenever he is sick if and only if \( r \leq B \). Furthermore, if \( r > B \) he will never seek treatment.

**Proof.** The remaining lifetime expected utility of a sick patient who knows the identity of a high ability specialist is given by \( V_{i,1} \). Furthermore, this is true for any period in which he is sick. This patient will seek treatment when he is sick if and only if \( V_{i,1} \geq 0 \). This requires that \( r \leq B \). Thus a patient who knows the identity of a high ability specialist will seek treatment whenever he is sick if and only if \( r \leq B \).

**Proposition 17 (Participation constraint for uniformed patients who have access to an informed GP):** A patient who does not know the identity of a high ability specialist but whose GP does know the identity of such a specialist will weakly prefer to seek both a referral and treatment over no treatment when he is first sick if and only if

\[
\frac{1 - \delta}{1 - \delta + \delta \pi} w.
\]

Furthermore, if such a patient seeks both a referral and treatment the first time he is sick and \( w \geq 0 \), he will seek treatment whenever he is sick from that point in time onwards.

**Proof.** The remaining lifetime expected utility of a sick patient who does not know the identity of a high ability specialist but whose GP does know the identity of a high ability specialist is given by \( V_{i,1} \) if that patient seeks both a referral and treatment. This patient will prefer to seek both a referral and treatment if and only if \( V_{i,1} \geq 0 \). This requires that

\[
\frac{1 - \delta + \delta \pi}{1 - \delta} (B - r) - w \geq 0,
\]

which can be rearranged to yield

\[
r \leq B - \left( \frac{1 - \delta}{1 - \delta + \delta \pi} \right) w.
\]

Finally, note that following the referral and treatment, such a patient will know the identity of a high ability specialist. We know that informed patients will be willing to seek treatment whenever they are sick if \( r \leq B \). Thus if \( w \geq 0 \) and uninformed patients with access to an informed GP seeks a referral after period zero, then it must the case that \( r \leq B \).

Recall that repetition does not affect the participation decision of specialists. As such, the participation constraint facing specialists in this dynamic health care markets with GPs will be the same as that facing them in a static health care market. We are now in a position to characterise the conditions under which a dynamic health care market with GPs will exist.
Proposition 18  (Treatment market existence): If all sick patients seek treatment in period zero, then the treatment market will exist after period zero whenever \( C_S \leq B \). However, some patients will choose not to seek treatment if

\[
\begin{align*}
    r &\in \left( \max \left\{ C_S, \frac{(1 - \delta + \delta \pi)}{(1 - \delta + \delta \pi \mu)} \right\} \mu B, \mu B - \frac{1 - \delta}{1 - \delta + \delta \pi} w \right), B \right].
\end{align*}
\]

Proof. Recall that patients are not allowed to seek treatment without a referral in period zero. As such, if every sick patient in period zero seeks treatment, then every sick patient will also seek a referral. Since each GP has an infinite patient pool, this means that each GP will almost surely have an infinite number of sick patients in period zero. We have already shown that this ensures that each GP will be able to discover the identity of at least one high ability specialist. Since a GP only discovers the identity of a high ability specialist when at least one of his patients is treated by a high ability specialist, we know that at least one patient from the patient pool of each GP must also learn the identity of a high ability specialist. This means that there will be at least \( J \) informed patients at the end of period zero. If these patients are ever sick from the beginning of period one onwards, they will seek treatment whenever \( r \leq B \). Specialists will be willing to provide a referral if and only if \( r \geq C_S \). As such, the treatment market will exist if and only if \( C_S \leq B \). On the other hand, if \( C_S > B \), then the treatment market will not exist. The mere existence of the treatment market does not mean that all sick patients will choose to seek treatment. Clearly informed patients will seek treatment. However, uninformed patients might not do so. Recall that uninformed patients would be willing to seek treatment without a referral if and only if

\[
\begin{align*}
    r &\leq \left( \frac{1 - \delta + \delta \pi}{1 - \delta + \delta \pi \mu} \right) \mu B = \left[ \frac{\mu (1 - \delta) + \delta \pi \mu}{(1 - \delta) + \delta \pi \mu} \right] B < B.
\end{align*}
\]

Furthermore, assuming that \( w \geq 0 \), uninformed patients will be willing to seek treatment with a referral if and only if

\[
\begin{align*}
    r &\leq B - \left( \frac{1 - \delta}{1 - \delta + \delta \pi} \right) w < B.
\end{align*}
\]

Thus uninformed patients will only participate in the treatment market if

\[
\begin{align*}
    \max \left\{ \left\{ \frac{(1 - \delta + \delta \pi)}{(1 - \delta + \delta \pi \mu)} \right\} \mu B, \mu B - \frac{1 - \delta}{1 - \delta + \delta \pi} w \right\} &\geq C_S.
\end{align*}
\]

If the treatment market exists but this condition does not hold, then it must be the case that

\[
\begin{align*}
    \max \left\{ C_S, \left\{ \frac{(1 - \delta + \delta \pi)}{(1 - \delta + \delta \pi \mu)} \right\} \mu B, \mu B - \frac{1 - \delta}{1 - \delta + \delta \pi} w \right\} &= C_S
\end{align*}
\]

and \( r \in [C_S, B] \). This means that uninformed patients will not participate in the treatment market despite the fact that it exists. ■

Since GPs are also present in the health care sector now, we also need to consider the existence of a referral market. This requires us to examine participation constraints for patients and GPs in the referral market after period zero.
Proposition 19 (Patient participation constraint for referrals): Informed patients will never seek a referral. Uninformed patients will seek a referral if and only if both
\[
 r \leq B - \left( \frac{1 - \delta}{1 - \delta + \delta \pi} \right) w
\]
and
\[
 w \leq \left( \frac{1 - \delta + \delta \pi}{1 - \delta + \delta \pi \mu} \right) (1 - \mu) B.
\]

Proof. Recall that
\[
 \hat{r}_d = \left[\frac{(1 - \delta + \delta \pi)}{(1 - \delta + \delta \pi \mu)}\right] \mu B.
\]
Let
\[
 \hat{r}_{GP} = B - \left( \frac{1 - \delta}{1 - \delta + \delta \pi} \right) w.
\]
Consider an uninformed patient who is sick at some time after period zero. If this patient seeks both a referral and treatment, his remaining lifetime expected utility will be
\[
 V_i(\text{referral}) = \begin{cases} 
 (\frac{1 - \delta + \delta \pi}{1 - \delta})(B - r) - w & \text{if } r \leq \hat{r}_{GP}; \\
 0 & \text{if } r > \hat{r}_{GP}.
\end{cases}
\]
If the patient seeks treatment without a referral, his remaining lifetime expected utility will be
\[
 V_i(\text{treatment}) = \begin{cases} 
 \left[\frac{(1 - \delta + \delta \pi)}{(1 - \delta + \delta \pi \mu)}\right] \mu B - \left[\frac{(1 - \delta + \delta \pi)}{(1 - \delta)}\right] r & \text{if } r \leq \hat{r}_d; \\
 0 & \text{if } r > \hat{r}_d.
\end{cases}
\]
Finally, if the patient seeks neither treatment nor referral, his remaining lifetime expected utility will be zero. A patient will prefer to seek a referral and treatment to no treatment whatsoever if and only if
\[
 \left( \frac{1 - \delta + \delta \pi}{1 - \delta} \right)(B - r) - w \geq 0.
\]
This can be rearranged to obtain
\[
 r \leq B - \left( \frac{1 - \delta}{1 - \delta + \delta \pi} \right) w.
\]
Furthermore, a patient will prefer to seek both a referral and treatment over treatment alone if and only if $V_i(\text{referral}) \geq V_i(\text{treatment})$. If $r \leq \min \{\hat{r}_d, \hat{r}_{GP}\}$, this requires that
\[
 \left( \frac{1 - \delta + \delta \pi}{1 - \delta} \right)(B - r) - w \geq \left[\frac{(1 - \delta + \delta \pi)}{(1 - \delta + \delta \pi \mu)}\right] \mu B - \left[\frac{(1 - \delta + \delta \pi)}{(1 - \delta)}\right] r.
\]
This can be rearranged to obtain
\[
 w \leq \left( \frac{1 - \delta + \delta \pi}{1 - \delta + \delta \pi \mu} \right) (1 - \mu) B.
\]
\[\blacksquare\]
Proposition 20 (GP participation constraint): GPs will offer their referral services whenever \( w \geq C_{GP} \).

Proof. We have already established that patients can deter GPs from referring them to low ability specialists. As such, repetition does not affect the participation decision of GPs. Recall that there are no fixed referral costs and constant marginal referral costs. As such, the average cost of a referral is constant. Indeed, it is simply the marginal referral cost. Hence GPs will participate if and only if the referral price exceeds this referral cost. This requires that \( w \geq C_{GP} \).

With the participation constraints for patients and GPs in hand, we are now in a position to establish the conditions under which a referral market will exist.

Proposition 21 (Referral market existence): A referral market will exist if and only if
\[
C_{GP} \leq \left( \frac{1 - \delta + \delta \pi}{1 - \delta + \delta \pi \mu} \right) (1 - \mu) B.
\]

Proof. Note that only uninformed patients will seek a referral after period zero. These patients will seek a referral if and only if
\[
w \leq \left( \frac{1 - \delta + \delta \pi}{1 - \delta + \delta \pi \mu} \right) (1 - \mu) B.
\]
GPs will offer their referral services if and only if \( w \geq C_{GP} \). As such, the set of referral prices for which both patients and GPs are willing to participate in the referral market will be non-empty if and only if
\[
C_{GP} \leq \left( \frac{1 - \delta + \delta \pi}{1 - \delta + \delta \pi \mu} \right) (1 - \mu) B.
\]

Finally, we are in a position to comment on the impact that the introduction of GPs has on the threshold treatment price for market existence.

Proposition 22 (Treatment price when a referral market exists): If the referral market exists, then the threshold treatment price will satisfy the following condition:
\[
\hat{r}_{GP} \in [\hat{r}_d, B - \left( \frac{1 - \delta}{1 - \delta + \delta \pi} \right) C_{GP}] \subseteq [\hat{r}_d, B].
\]

Proof. Recall that, for the referral market to exist, we need
\[
C_S \leq w \leq \left( \frac{1 - \delta + \delta \pi}{1 - \delta + \delta \pi \mu} \right) (1 - \mu) B.
\]
As such, we know that
\[
\hat{r}_{GP} \leq B - \left( \frac{1 - \delta}{1 - \delta + \delta \pi} \right) C_{GP} \leq B.
\]
Furthermore, we also know that
\[
\hat{r}_{GP} \geq B - \left( \frac{1 - \delta}{1 - \delta + \delta \pi} \right) \left( \frac{1 - \delta + \delta \pi}{1 - \delta + \delta \pi \mu} \right) (1 - \mu) B,
\]
which can be rearranged to obtain

\[ \hat{\tau}_{GP} \geq \left[ \frac{(1 - \delta + \delta \pi)}{(1 - \delta + \delta \pi \mu)} \right] \mu B = \hat{\tau}_d. \]

6 A comparison of treatment market outcomes

We can now compare the various circumstances in which a treatment market will exist and the treatment outcomes in each of these circumstances. Recall that a static treatment market will exist if \( C_S \leq r \leq \mu B = \hat{\tau}_s \), while a dynamic treatment without GPs will exist if

\[ C_S \leq r \leq \left[ \frac{(1 - \delta + \delta \pi)}{(1 - \delta + \delta \pi \mu)} \right] \mu B = \hat{\tau}_d. \]

Furthermore, a dynamic treatment market with GPs in which all sick patients seek treatment will exist if

\[ C_S \leq r \leq B - \left( \frac{1 - \delta}{1 - \delta + \delta \pi} \right) C_{GP} = \hat{\tau}_{GP}. \]

Finally, a dynamic treatment market with GPs in which only informed patients seek treatment will exist if

\[ B - \left( \frac{1 - \delta}{1 - \delta + \delta \pi} \right) C_{GP} = \hat{\tau}_{GP} \leq r \leq B. \]

The relationship between these various threshold treatment prices is illustrated in Figure 1. It is clear that repetition expands the range of treatment prices for which the treatment market will exist. Repetition and the presence of GPs expands the range of treatment prices that are consistent with treatment market existence even further.

Treatment market outcomes will vary with the nature of the market and the prevailing prices. In both the static market and the dynamic market without GPs, all patients will seek and obtain treatment whenever the treatment price lies between the cost of treatment and the relevant threshold treatment price. However, in both cases, patients will not know whether or not they will be cured. Some patients will be cured, while some patients will not be cured. In a dynamic market with GPs, all patients will seek treatment if the treatment price lies between the cost of treatment and the threshold price for uninformed patients. However, if the treatment price lies between the threshold treatment price for uninformed patients and the benefit from good health, then only the informed patients will seek treatment. In both of these cases, all of the patients who seek treatment will be cured.

7 Information externalities and the need for regulation

While characterising the outcomes in dynamic markets with GPs, we assumed that all sick patients were required to seek a referral and treatment in period
zero. This allowed every GP to learn the identity of at least one high ability specialist before the start of the next period. In this section, we will examine patients referral choices in period zero. We will show that all sick patients choosing to seek a referral in that period is not an equilibrium outcome. The intuition for this result involves the presence of a positive information externality. Individual patients bear the entire cost of obtaining a referral. However, they do not capture any of the benefits from that referral. The patient would learn the ability level of one of the specialists if he sought treatment regardless of whether or not he also sought a referral. The benefit from the referral is that the GP also learns the ability level of one of the specialists for each referral that he makes. Given the presence of this positive externality, it is not surprising that patients may choose to consume too few referrals in period zero, from a social welfare point of view. As such, some policy to correct for this may be warranted. Potential policies include a requirement that patients seek a referral before obtaining treatment in period zero or some form of subsidy for patients who seek a referral. As we noted in the introduction to this paper, the cost of treatment is often subsidised for patients who seek a referral in Australia.

**Proposition 23** *(The need for regulation — Part 1)*: All sick patients in period zero choosing to seek a referral is not part of an equilibrium outcome if referrals are not free.

**Proof.** Suppose that all but one of the sick patients in period zero choose to seek both a referral and treatment. Consider the choice confronting the remaining sick patient. There are an infinite number of sick patients in period zero who choose to seek a referral from each GP regardless of this patient’s referral decision. As such, every GP will still be able to learn the identity of at least one high ability GP at the end of period zero. Furthermore, the probability of this patient becoming an informed patient at the end of period zero is not altered by his referral choice. As such, the continuation payoff for this patient will not be altered by his referral decision. This means that he can simply
maximise his period zero expected utility when making a referral decision. The expected utility for this patient in period zero if he seeks both a referral and treatment is

\[ EU_i(\text{referral}) = \mu B - r - w. \]

The expected utility for this patient in period zero if he seeks only treatment is

\[ EU_i(\text{treatment}) = \mu B - r. \]

Clearly

\[ EU_i(\text{referral}) \geq EU_i(\text{treatment}), \]

with the inequality being strict if \( w > 0 \). As such, this patient will choose not to seek a referral in period zero if referrals are not free. Thus it cannot be an equilibrium outcome for all sick patients to seek both a referral and treatment in period zero if referrals are not free.

**Proposition 24** *(The need for regulation — Part 2):* All sick patients in period zero choosing not to seek a referral is part of an equilibrium outcome if referrals are not free.

**Proof.** Suppose that all but one of the sick patients in period zero choose not to seek both a referral and treatment. Consider the choice confronting the remaining sick patient. If seeks a referral, he will not alter the probability that he learns the identity of a high ability specialist in period zero. Furthermore, he will not alter his expected continuation payoff. If he learns the identity of a high ability specialist, then he will not need to seek a referral in future. If he does not learn the identity of a high ability specialist, then nor will any GPs, regardless of whether or not this patient sought a referral in period zero. Since the expected continuation payoff for this patient will not be altered by his referral decision, he will simply maximise his period zero expected utility when making a referral decision in period zero. The expected utility for this patient in period zero if he seeks both a referral and treatment is

\[ EU_i(\text{referral}) = \mu B - r - w. \]

The expected utility for this patient in period zero if he seeks only treatment is

\[ EU_i(\text{treatment}) = \mu B - r. \]

Clearly

\[ EU_i(\text{referral}) \geq EU_i(\text{treatment}), \]

with the inequality being strict if \( w > 0 \). As such, this patient will choose not to seek a referral in period zero if referrals are not free. Hence, when referrals are not free, it is an equilibrium outcome for no patients to seek a referral in period zero.

It appears possible that sick patients might not seek referrals in sufficient numbers to allow GPs to become informed at the end of period zero. As such, some form of regulation might be needed if each GP is to be able to learn the identity of at least one high ability specialist. Suppose that any sick patient who wants to obtain treatment in period zero is required to obtain a referral prior to seeking treatment. Will such patients still choose to obtain treatment in period zero?
Proposition 25 \textit{(Patient participation constraint for period zero):} If sick patients are required to obtain a referral before seeking treatment in period zero, then they will seek treatment if and only if
\[
W \leq \left( \frac{1 - \delta + \delta \pi}{1 - \delta + 2\delta \pi} \right) \left[ \left( \frac{\mu - \delta \mu + \delta \pi}{1 - \delta} \right) B - \left( \frac{1 - \delta + \delta \pi}{1 - \delta} \right) r \right].
\]

\textbf{Proof.} The lifetime expected utility of a sick patient in period zero who chooses to seek both a referral and treatment can be decomposed into three terms. These terms are the patient’s expected utility in period zero, his continuation utility if he becomes informed at the end of period zero and his continuation utility if he does not become informed at the end of period zero. The continuation payoffs will also need to be weighted by the probability of their occurrence. We will assume throughout that all of the other sick patients during period zero choose to seek a referral. Since we are deriving a condition under which this will be true, this assumption will be valid if that condition holds. The patient’s expected utility in period zero is
\[
EU_i(\text{referral}) = \mu B - r - w.
\]
His continuation utility if he learns the identity of a high ability specialist during period zero is
\[
V_{i,1} = \sum_{t=1}^{\infty} \delta^t \pi (B - r) = \frac{\delta \pi (B - r)}{(1 - \delta)}.
\]
His continuation utility if he does not learn the identity of a high ability specialist during period zero is
\[
V_{i,2} = \sum_{t=1}^{\infty} \delta^t (1 - \pi)^{t-1} \pi w.
\]
This expression simplifies to
\[
V_{i,2} = \frac{\delta \pi (B - r)}{(1 - \delta)} - \frac{\delta \pi w}{(1 - \delta + \delta \pi)}.
\]
Thus the patients lifetime expected utility if he seeks a referral is
\[
V_i = \mu B - r - w + \mu \left( \frac{\delta \pi (B - r)}{(1 - \delta)} \right) + (1 - \mu) \left( \frac{\delta \pi (B - r)}{(1 - \delta)} \right) - \frac{\delta \pi w}{(1 - \delta + \delta \pi)},
\]
which can be rearranged to obtain
\[
V_i = \left( \frac{\mu - \delta \mu + \delta \pi}{1 - \delta} \right) B - \left( \frac{1 - \delta + \delta \pi}{1 - \delta} \right) r - \left( \frac{1 - \delta + 2\delta \pi}{1 - \delta + \delta \pi} \right) w.
\]
Clearly, the patient will choose to seek a referral if and only if \( V_i \geq 0 \), which requires that
\[
\left( \frac{\mu - \delta \mu + \delta \pi}{1 - \delta} \right) B - \left( \frac{1 - \delta + \delta \pi}{1 - \delta} \right) r - \left( \frac{1 - \delta + 2\delta \pi}{1 - \delta + \delta \pi} \right) w \geq 0.
\]
This expression can be rearranged to obtain
\[
w \leq \left( \frac{1 - \delta + \delta \pi}{1 - \delta + 2\delta \pi} \right) \left[ \left( \frac{\mu - \delta \mu + \delta \pi}{1 - \delta} \right) B - \left( \frac{1 - \delta + \delta \pi}{1 - \delta} \right) r \right].
\]
Finally, we need to establish the conditions under which the treatment and referral markets will exist in period zero.

**Proposition 26** *(Treatment market existence in period zero):* If there is a regulation requiring any patient that wants treatment in period zero to obtain a referral as well, then the treatment market will only exist if

\[
C_S \leq \left( \frac{\mu(1-\delta) + \delta\pi}{1-\delta + \delta\pi} \right) B - \left( \frac{(1-\delta)(1-\delta + 2\delta\pi)}{(1-\delta + \delta\pi)^2} \right) w.
\]

**Proof.** Recall that specialists will offer their treatment services if and only if \( r \geq C_S \). We can rearrange the period zero participation constraint for patients to obtain

\[
r \leq \left( \frac{\mu(1-\delta) + \delta\pi}{1-\delta + \delta\pi} \right) B - \left( \frac{(1-\delta)(1-\delta + 2\delta\pi)}{(1-\delta + \delta\pi)^2} \right) w.
\]

Thus the participation constraints for patients and GPs can be simultaneously satisfied if and only if

\[
C_S \leq \left( \frac{\mu(1-\delta) + \delta\pi}{1-\delta + \delta\pi} \right) B - \left( \frac{(1-\delta)(1-\delta + 2\delta\pi)}{(1-\delta + \delta\pi)^2} \right) w.
\]

**Proposition 27** *(Referral market existence in period zero):* Even with a regulation requiring any patient that wants treatment in period zero to obtain a referral as well, the referral market will only exist if

\[
C_{GP} \leq \left( \frac{1-\delta + \delta\pi}{1-\delta + 2\delta\pi} \right) \left[ \left( \frac{\mu - \delta\mu + \delta\pi}{1-\delta} \right) B - \left( \frac{1-\delta + \delta\pi}{1-\delta} \right) r \right].
\]

**Proof.** If this condition is not satisfied, then the participation constraints for patients and GPs cannot be simultaneously satisfied.

8 Conclusion

Some professional service industries display a gated structure, notably including the medical industry in some Commonwealth countries. The main focus of this paper has been on explaining both the existence of gatekeeping intermediaries who refer consumers to one of many ultimate producers and providing a rationale for regulations that encourage the use of referrals. The results in this paper are complementary to those obtained in Eldridge ([14]). In that paper, the gated industry structure observed in some professional service industries provided an artificial long-run relationship between patients and specialists when, in the absence of GPs, they would only have a short-run relationship. The artificial long-run relationship between patients and specialists enabled them to avoid a market failure resulting from shirking on the part of specialists. This industry structure was largely driven by the demands of patients, although there may have been some circumstances in which the presence of GPs improved the welfare of both patients and specialists. While we provided an explanation for the
gated structure of some professional service industries in Eldridge ([14]), that explanation did not provide a rationale for regulations that encouraged such a structure. In this paper, we have provided a rationale for such regulations. However, while patients in the model employed in Eldridge ([14]) had an incentive to repeatedly seek a referral for the treatment of non-chronic diseases, patients in the model employed in this paper will seek a referral at most twice. The reason for this difference relates to the underlying market failure. In Eldridge ([14]), the underlying market failure is a moral hazard problem. Specialists could alter their effort choices from period to period. As such, they constantly needed to be induced to provide high effort treatment. In this paper, the underlying market failure is an adverse selection problem. The ability level of a specialist is private information, known only by that specialist. However, this ability level is fixed for all time. Thus, if a patient learns the identity of a high effort specialist, he will obtain no additional benefits from seeking further referrals.

In actual health care markets, patients might well seek a referral on a number of occasions. As such, it would appear that the limited number of referrals that are predicted by the model in this paper is somewhat unrealistic. However, that result is generated by the stationary population of agents in the model employed in this paper. In actual health care markets, the populations of patients, GPs and specialists will be in a constant state of flux. In each period, some new agents will arrive and some old agents will leave. Thus we would expect the outcomes in actual health care markets to reflect aspects of both the period zero outcomes and the later period outcomes of the model employed in this paper.
References


[19] Fudenberg, D and J Tirole (1995), Game theory, Massachusetts Institute of Technology Press, USA.


[40] Stiglitz, J and A Weiss (1990), Sorting out the differences between signaling and screening models, NBER Technical Working Paper Number 93, National Bureau of Economic Research, USA.

Recent Discussion Papers

School of Business Discussion Papers are available from the Research Officer, School of Business, La Trobe University VIC 3086, Australia.

07.01 Rosaria Burchielli and Timothy Bartram – What makes organising work? A model of the stages and facilitators of organising.

07.02 László Kónya and Jai Pal Singh - Causality between Indian Exports, Imports, and Agricultural, Manufacturing GDP.

07.03 Buly A. Cardak and Chris Ryan – Participation in Higher Education: Equity and Access – Are Equity-based Scholarships an Answer?

07.04 Samantha Farmakis-Gamboni and Divid Prentice – Does Reducing Union Bargaining Power Increase Productivity

07.05 Damien S. Eldridge – A Shirking Theory of Referrals