ACCESS TO HEALTH CARE AS A CONGESTION GAME

YUNG-WEN LIU

Department of Industrial and Manufacturing Systems Engineering, University of Michigan, Dearborn, Michigan

TILMAN BÖRGERS

Department of Economics, University of Michigan, Ann Arbor, Michigan

ABSTRACT. We model health care access as a *congestion game* in which different options for access to health care become less convenient for patients as more patients use the same option. We focus on the choice between attending a walk-in clinic and making an appointment. We characterize a *Bayesian Nash equilibrium* of patients' interaction, and compare it to the optimal patient allocation. We focus then on policy tools that a health care system might use to steer patients to the optimal access choice. We also analyze how a health care system might acquire the information needed to implement the optimal policy.

1. INTRODUCTION

The delays that patients experience when trying to access primary healthcare are widely regarded as an important problem [4, 11]. The relevance of this problem derives from the costs to the healthcare system and the discomfort to patients that an imperfectly functioning access system causes [5]. Different systems for providing access that have been used in practice include scheduled appointments, open access appointments, and walk-in clinics. The literature on access systems has addressed the optimal operation of each of them, such as the optimal scheduling of appointments, and also the optimal proportion of patients assigned to different access systems (e.g. [16], [17]). In the existing literature, patients' behavior is assumed to be exogenous.

E-mail addresses: ywliu@umich.edu, tborgers@umich.edu.

Date: October 14, 2011.

We are grateful to participants at the INFORMS Healthcare conference in Montreal, June 2011, for their comments.

However, in practice patients choose which access point to health care they contact. In our paper, patients' choices will be the focus of the analysis.

Our model of patients' behavior is game theoretic because each patient's optimal choice depends on the other patients' choices. If the number of other patients who attend a walk-in clinic is very large, then it might be optimal to instead make an appointment, whereas, if the number of other patients who attend a walk-in clinic is low, then it might be better to go to the walk-in clinic. Specifically, we model the interaction among patients as a congestion game. In such games each option available to an agent becomes less attractive as more other agents make the same choice. The sizable literature on congestion games started with [13]. A typical application is car traffic: the more drivers take any particular route, the more congestion there is, and therefore the lower the utility from that choice (see, e.g., [14]). Central to this paper is the observation that when choosing their access point to healthcare, patients face a problem that is conceptually related to that faced by car drivers when choosing their commuting route. To our knowledge, ours is the first paper to develop a theory of congestion games in health care.

In our model patients differ in their cost of waiting. The disruption of a patient's work life that an extended stay in a waiting room causes, for example, will be different for different patients. Our congestion game will therefore be one with heterogeneous preferences. Congestion games with heterogeneous preferences have previously been considered in [9], [10], and [15]. The main results of this literature concern existence and uniqueness of equilibrium. We offer an independent, elementary proof of existence and uniqueness of equilibrium in our model, and then focus on a characterization of the equilibrium choices, as well as the welfare properties of the equilibrium. We show that the externality among agents causes equilibrium to be inefficient. We discuss how health care providers might go about repairing this inefficiency. Our discussion of health care providers' policies builds on the economics literature on mechanism design (e.g. [2]) and on optimal taxation (e.g. [6]), and applies the insights contained in that literature to our particular context.

To facilitate the focus on endogenous patient choices, we abstract in this paper from other features of the health care system. In particular, the precise organization of different access systems is ignored, and we use a reduced form approach to describe the relation between patients' attendance and waiting times. Also, unlike some of the existing literature (e.g. [1], [7], [16], [17]), we assume that the resources allocated to each access system are given and fixed, and that access systems cannot share resources among each other.

2. Model

There is a continuum of patients. Each patient i experiences pain at a level which differs from patient to patient. We model pain as a "cost per time interval," that is, the cost of suffering the pain in one time unit. We denote the pain cost per time interval for patient i as $p_i \geq 0$. Patients also have costs when they have to wait at a health care provider's location for treatment. These costs are distinct from their pain cost, and result, for example, from the disruption to work or family life. We refer to these costs as waiting costs. The waiting cost per time interval for patient i is $w_i \geq 0$. The joint distribution of (p_i, w_i) in the continuum size patient population is described by the cumulative distribution function G. We assume that G has support $[0, p_{\max}] \times [0, w_{\max}]$ for some finite $p_{\max}, w_{\max} > 0$. We also assume that G has a continuous density g on this support. We may think of each patient's characteristic (p_i, w_i) as independently distributed according to G. In a continuum size population the realized distribution is the same as the theoretical distribution, so that both distributions are given by G.

Patients choose simultaneously and independently between either calling the health care provider to make an appointment or attending a walk-in clinic. For simplicity we assume that each patient makes only one choice, and that this choice is irreversible. We think of the appointment clinic and the walk-in clinic as separate institutions the operations of which do not interfere with each other. A patient who attends a walk-in clinic has his or her pain alleviated quickly, but the waiting time at the walk-in clinic causes cost w_i per time unit. For simplicity, we set the pain cost of a patient who goes to a walk-in clinic equal to zero. A patient who attends an appointment clinic will have his or her pain alleviated with some delay, and therefore experiences what we refer to as pain cost of p_i per time unit. On the other hand, scheduled appointments cause less waiting cost. For simplicity we set the waiting cost of scheduled appointments equal to zero.

The average waiting times that patients anticipate when they choose to call the appointment clinic or attend the walk-in clinic depend on the choices that other patients make. The average waiting time at either of these choices are the longer the more patients make that choice. We assume that the average waiting time is, in fact, *equal* to the measure of patients making the choice.[1] We denote by $d(p_i, w_i) \in \{0, 1\}$ the choice of the patient with costs per time unit (p_i, w_i) where $d(p_i, w_i) = 1$ indicates that the patient makes an appointment, and $d(p_i, w_i) = 0$ indicates that the patient attends the walk-in clinic. Then the average waiting time for an appointment is:

$$t^A = \int_0^{w_{\max}} \int_0^{p_{\max}} d(p_i, w_i) g(p_i, w_i) dp_i dw_i \tag{1}$$

and the average waiting time in the walk-in clinic is:

$$t^C = \int_0^{w_{\max}} \int_0^{p_{\max}} (1 - d(p_i, w_i))\, g(p_i, w_i) dp_i dw_i = 1 - t^A. \tag{2}$$

[1]Setting average waiting time equal to the measure of patients making a choice would be without loss of generality if we made no specific assumptions about cost functions. In (3) and (4) we make, however, quite restrictive assumptions for cost functions. See footnote 2 where we explain the extent to which our analysis can be generalized.

Of course, different patients will in practice have different times to wait for their appointments, and different patients will also have different waiting times at the walk-in clinic. But as patients find out their actual waiting times only after they have made their choice we assume that it is the average waiting time on which they base their choice.

For a patient with costs per time unit (p_i, w_i) the anticipated cost of waiting for an appointment is:[2]

$$p_i t^A \tag{3}$$

and the anticipated cost of going to the walk-in clinic is:

$$w_i t^C. \tag{4}$$

Patients choose whatever option has the lower anticipated cost.

As we mentioned before, the game that we have described above is a *congestion game.* Specifically, our model falls into the class of *congestion games with player-specific payoff functions* introduced by [9]. This is because each agent has their own cost function. Different agents therefore have different costs from congestion. To our knowledge, the specific congestion game described in this Section has not been analyzed before.

3. Patients' Equilibrium Choices

We now study patients' choices if they make these choices independently and without co-ordination. A patient optimally calls for an appointment if (3) is less than (4), and a patient optimally attends the walk-in clinic if (4) is less than (3). The expressions in (3) and (4) involve the average waiting times t^A and t^C which, in turn, depend on the choices of patients other than i. This interaction among agents turns our model into a *game* in the sense of game theory. To predict patients' choices we shall use the notion of *Bayesian Nash equilibrium* [3, p. 215]. In a Bayesian Nash equilibrium each player correctly predicts all other players' choices. In our context this means that each patient i correctly predicts the values of t^A and t^C as determined by all patients' choices. One can think of equilibrium as a rest-point of a learning process that gradually takes patients to equilibrium. The precise concept that we use is Bayesian Nash equilibrium because our game is a game of incomplete information: no player knows the cost (p_i, w_i) of any other player. However, we do assume that players know the distribution of other players' cost parameters, and in a model with a continuum of players, because the realized distribution of players' cost parameters equals the theoretical distribution, there is no uncertainty left that players have to deal with. This is different in the finite population model that we investigate in Section 7.

[2] We use the linear cost functions in (3) and (4) for simplicity. The main results of our analysis remain qualitatively true if patients' costs are described by convex functions indexed by a one-dimensional parameter such that an increase in that parameter leads to an increase in the marginal cost of waiting for all t^A and t^C.

Definition 1. *A function d^* that maps every pair (p_i, w_i) into either 1 (appointment) or 0 (walk-in clinic) is an* equilibrium *if for every (p_i, w_i) the choice $d^*(p_i, w_i)$ is optimal if patients correctly anticipate the waiting times given by (1) and (2), where these waiting times are calculated using d^*.*

Our first proposition says that there is a unique equilibrium in our model. Strictly speaking, equilibrium is unique up to the a set of Lebesgue measure zero of patients who are indifferent between the two choices. We shall ignore such sets.

Proposition 1. *There is a unique equilibrium d^*. In this equilibrium there is a constant $\bar{r} > 0$ such that patients i for whom $p_i < \bar{r} w_i$ make an appointment, and patients i for whom $p_i > \bar{r} w_i$ attend the walk-in clinic. Here, $\bar{r} = t^C / t^A$.*

Proposition 1 is a special case of Proposition 4 below. Therefore, we don't offer a separate proof of Proposition 1. The existence of an equilibrium in our model also follows from [15, Theorem 2], and also from [9, Theorem 2]. It is also shown in [10, Theorem 4.3] that congestion games of the type considered here generically have a unique equilibrium. We shall offer an independent, elementary proof of Proposition 4 because it is more general[3], and because we obtain at the same time a characterization of patients' behavior in the equilibrium.

Figure 1 illustrates the equilibrium described in Proposition 1. We have simulated a population of 500 patients. Each patient's characteristics are indicated by a black dot in Figure 1. We have then calculated the equilibrium ratio $\bar{r}$. This ratio equals the slope of the solid line in Figure 1. Patients with characteristics above the solid line attend the walk-in clinic, whereas patients with characteristics below the solid line make an appointment.[4]

In equilibrium, patients sort themselves according to their privately observed characteristics into two categories. There is no need for the health care provider to observe patients' characteristics directly. As it is the cost ratio, not the absolute level of cost, that determines equilibrium behavior of

[3]Because it allows for fees.

[4]In Figure 1 the distribution of patients' characteristics is a product distribution, with the marginal distribution of w being a power distribution with parameter 0.5, and the marginal distribution of p being a power distribution with parameter 2. To find the equilibrium we have first constructed numerically the map that assigns to every *expected* number of patients making an appointment the corresponding number of patients *optimally* making an appointment if their expectations are given by the expected number. In equilibrium these two numbers have to be the same, that is, we need a fixed point of this mapping. The mapping is monotonically decreasing. To find a fixed point, we started at the lowest expected number, and then gradually increased the expected number until a fixed point was reached. All computations were carried out in R. The algorithm that was used is available from the authors upon request.

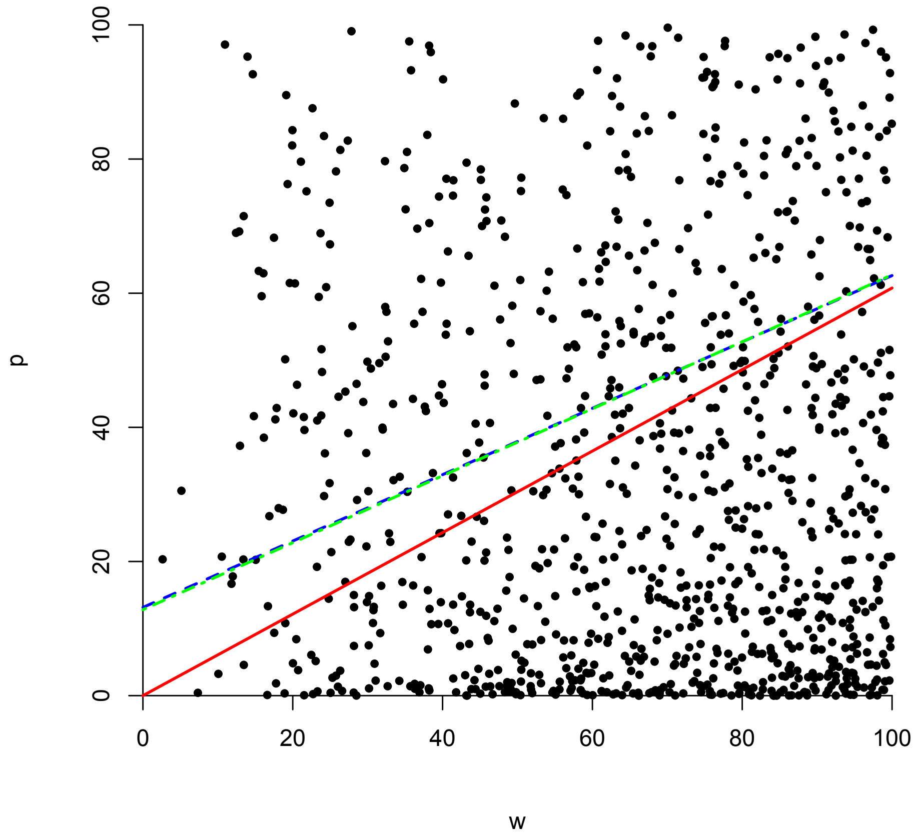

FIGURE 1. Equilibrium (solid line) and optimal (dashed line) allocations of patients.

patients though, even patients with arbitrarily low p_i will attend the walk-in clinic, and even patients with arbitrarily low w_i will make appointments. For example, a patient may attend the walk-in clinic even though their pain level is very low, if, relatively speaking, the interruption to their daily life due to the waiting time at the walk-in clinic is unimportant to the patient.

Careful reading of the proof of Proposition 4, which implies Proposition 1, reveals that the equilibrium is determined by the distribution of the ratio p_i/w_i that is implied by the distribution G of patients' characteristics. Denote this distribution by F. Suppose that this distribution is shifted to the right in the sense of *first order stochastic dominance*, that is for all possible values of the ratio r the value $F(r)$ decreases. The proof shows that then the equilibrium value of $\bar{r}$ will increase. Intuitively, as more patients experience higher pain levels, in equilibrium patients will anticipate more overcrowding at the walk-in clinic, and therefore become more willing to wait for an appointment.

4. Optimal Patient Allocation

Next, we investigate how a benevolent social planner would allocate patients to the appointment or the walk-in clinic, that is, how a social planner would choose the function d. We refer in this section to d as an *allocation* (of patients to treatment locations). We assume that the planner's objective is to minimize the average pain or waiting cost that patients experience. This average is given by:

$$\int_0^{w_{\max}} \int_0^{p_{\max}} d(p_i, w_i) p_i t^A + (1 - d(p_i, w_i))\, w_i t^C dp_i dw_i \tag{5}$$

where t^A and t^C are given by (1) and (2).

For simplicity we restrict attention to allocations d such that there are two open subsets D^C and D^A of $\mathbb{R}^2_+$ such that $(p_i, w_i) \in D^C$ implies $d(p_i, w_i) = 0$, $(p_i, w_i) \in D^A$ implies $d(p_i, w_i) = 1$, and such that the set $D^C \cup D^A$ has G-measure 1. We shall call such allocations d *regular*. Note that this is a very mild restriction. All plausible potential solutions to the planner's optimization problem are regular. The assumption of regularity simplifies the proof of the following proposition.

For any regular solution, define g^A to be the G-measure of D^A and g^C to be the G-measure of D^C.[5] We define $\bar{p}^A$ to be the expected value of p_i conditional on $(p_i, w_i) \in D^A$ and $\bar{w}^C$ to be the expected value of w_i conditional on $(p_i, w_i) \in D^C$.

Proposition 2. *For any regular allocation $\hat{d}$ that minimizes (5) among all regular solutions there are constants $\hat{a}$ and $\hat{r} > 0$ such that patients i for whom $p_i < \hat{r} w_i + \hat{a}$ make an appointment, and patients i for whom $p_i > \hat{r} w_i + \hat{a}$ attend the walk-in clinic. Here, $\bar{r} = t^C / t^A$ and $\hat{a} = \left(g^C \bar{w}^C - g^A \bar{p}^A\right) / t^A$.*

Note that the functional form of socially optimal allocations that Proposition 2 describes is more general than the functional form of equilibrium outcomes that Proposition 1 describes. Whereas Proposition 2 allows an intercept $\hat{a}$ that differs from zero, in Proposition 1 it is shown that in equilibrium this intercept is always zero. Figure 1 illustrates the optimal allocation. All patients above the dashed line attend the walk-in clinic, whereas all patients below the dashed line make an appointment.[6]

---

[5] Of course, by assumption, g^A is the same as t^A, and g^C is the same as t^C. However, the exposition is clearer if we use two separate symbols that indicate which interpretation of these variables is intuitively relevant.

[6] We determined the optimal allocation in Figure 1 using a grid search for the optimal fee. From Corollary 1 below we can infer that the optimal allocation is implemented by a fee. The algorithm that was used to find the optimal fee is available from the authors upon request.

Proof. We begin by noting that there cannot be an optimal regular allocation for which $D^C = \emptyset$ or $D^A = \emptyset$. If one of these sets were empty, the average waiting time at the corresponding location would be zero, whereas at the other location it would be 1. Therefore, average cost could be reduced by allocating some patients from the location with waiting times of 1 to the location with waiting times of 0. From now on, when referring to regular allocations, we shall assume that $D^C \neq \emptyset$ and $D^A \neq \emptyset$.

We next construct a necessary condition for a solution to the social planner's problem. Consider any $(p_i^*, w_i^*) \in D^C$. Because $D^C \neq \emptyset$ there is at least one such element of D^C. Because D^C is an open set, there is a neighborhood around (p_i^*, w_i^*) such that $d(p_i, w_i) = 0$ for every (p_i, w_i) in that neighborhood. Let the G-measure of that neighborhood be $\varepsilon > 0$. We denote the neighborhood by $\mathcal{N}_\varepsilon$. We choose $\mathcal{N}_\varepsilon$ sufficiently small so that $\mathcal{N}_\varepsilon \neq D^C$. A necessary condition for a solution to the social planner's problem is that the average cost are not reduced by setting $d(p_i, w_i) = 1$ for all $(p_i, w_i) \in \mathcal{N}_\varepsilon$. There are three effects of such a change. First, the average waiting time for an appointment is increased by ε. Second, the average waiting time at the walk-in clinic is reduced by ε. Finally, there is the direct effect on the average waiting costs of all patients belonging to $\mathcal{N}_\varepsilon$. The necessary condition is that the sum of these three effects has to be non-negative.

To write the necessary condition formally, we need some notation. We denote by g_ε^C the G-measure of $D^C \setminus \mathcal{N}_\varepsilon$. We denote by $(\bar{p}_\varepsilon, \bar{w}_\varepsilon)$ the expected value of (p_i, w_i) conditional on $(p_i, w_i) \in \mathcal{N}_\varepsilon$. We denote by $\bar{w}_\varepsilon^C$ the expected value of w_i conditional on $(p_i, w_i) \in D^C \setminus \mathcal{N}_\varepsilon$.

The change in average cost of switching the allocation of all members of $\mathcal{N}_\varepsilon$ from the walk-in clinic to appointments is:

$$\varepsilon \left(\bar{p}_\varepsilon \left(t^A + \varepsilon\right) - \bar{w}_\varepsilon t^C\right) + \left(g^A \bar{p}^A \varepsilon - g_\varepsilon^C \bar{w}_\varepsilon^C \varepsilon\right) \tag{6}$$

A necessary condition for optimality is that this expression is non-negative. Dividing the expression by ε, we see that this is equivalent to:

$$\left(\bar{p}_\varepsilon \left(t^A + \varepsilon\right) - \bar{w}_\varepsilon t^C\right) + \left(g^A \bar{p}^A - g_\varepsilon^C \bar{w}_\varepsilon^C\right) \geq 0 \tag{7}$$

Now we consider a sequence of open neighborhoods of (p_i^*, w_i^*) such that ε tends to zero. Taking limits for ε tending to zero on the left hand side of (7) we obtain:

$$\left(p_i^* t^A - w_i^* t^C\right) + \left(g^A \bar{p}^A - g^C \bar{w}^C\right) \geq 0 \Leftrightarrow \tag{8}$$

$$p_i^* \geq \frac{t^C}{t^A} w_i^* + \frac{g^C \bar{w}^C - g^A \bar{p}^A}{t^A} \tag{9}$$

We conclude that every $(p_i^*, w_i^*) \in D^C$ has to satisfy (9). A symmetric argument shows that every $(p_i^*, w_i^*) \in D^A$ has to satisfy the reverse of inequality (9). $\square$

Proposition 2 does not address the existence of an optimal allocation. A suitable technical strengthening of our regularity assumption for d yields the conclusion that an optimal allocation exists. We do not deal with the details of this argument here.[7]

The intercept $\hat{a}$ in Proposition 2 is negative if $g^A\bar{p}^A > g^C\bar{w}^C$. Intuitively, this condition means that the total cost per unit of time born by all patients who make appointments is larger than the total cost per unit of time born by all patients who attend the walk-in clinic. If the intercept is negative then there is a set of patients for whom the cost of waiting at the walk-in clinic w_i is so low that they are allocated to the walk-in clinic independent of their pain level. The intercept $\hat{a}$ in Proposition 2 is positive if $g^C\bar{w}^C > g^A\bar{p}^A$. Intuitively, this condition means that the total cost per unit of time born by patients who attend the walk-in clinic is larger than the total cost per unit of time born by patients with appointments. If the intercept is positive then there is a set of patients for whom the pain p_i is so low that they are allocated to appointments independent of their cost of waiting at the walk-in clinic. Thus, unlike in the equilibrium, the allocation is not exclusively determined by cost ratios, but also, to a limited extent, by the absolute cost level.

5. Comparing Equilibrium and Optimum

We now compare the equilibrium that we found in Proposition 1 to the optimal allocation as chosen by a social planner. The following result is a consequence of arguments found in the proof of Proposition 2. We therefore omit the proof of the following result.

Proposition 3. *If in the unique Nash equilibrium the total cost per unit of time, $g^C\bar{w}^C$, born by the patients at the walk-in clinic is larger than the total cost per unit of time, $g^A\bar{p}^A$, born by the patients who make appointments, then the unique Nash equilibrium is not optimal. Average patient cost can be lowered by re-allocating some patients from the walk-in clinic to appointments. If in the unique Nash equilibrium the total cost per unit of time, $g^C\bar{w}^C$, born by the patients at the walk-in clinic is smaller than the total cost per unit of time, $g^A\bar{p}^A$, born by the patients who make appointments, then the unique Nash equilibrium is not optimal. Average patient cost can be lowered by re-allocating some patients from appointments to the walk-in clinic.*

We note that typically one of the conditions in Proposition 3 will be satisfied by the unique Nash equilibrium. In general, there is no force in our model that would ensure that the Nash equilibrium satisfies the necessary condition for optimality in Proposition 2. If it violates this condition, then it will satisfy one of the conditions in Proposition 3. Typically, the Nash equilibrium will be inefficient.

[7]To prove existence, one may restrict attention to a class of regular allocations d that include those described in Proposition 2, but that is restrictive enough to be compact in a topology that makes the expected cost function continuous. Then one may appeal to the existence of a maximum of a continuous function on a compact space.

The reason for the discrepancy between equilibrium and optimal allocations in our model is that patients choose between attending a walk-in clinic and making an appointment according to their own interest, only, and do not take the consequences of this decision for other patients into account. These consequences, namely, that other patients' waiting times change, constitute an economic *externality*. By contrast, the social planner, when allocating a patient to either the walk-in clinic or to a scheduled appointment, takes into account the implications of these decisions for other agents.

In a Nash equilibrium, there will be some patients who are almost indifferent between going to a walk-in clinic and making an appointment. These are the patients for whom the inequality in Proposition 1 is almost satisfied as an equality. If these agents are re-allocated their own cost stay almost the same. However, the waiting times of all other patients are changed, and these other patients are not indifferent to changes in their waiting time. For example, if more patients are allocated to the walk-in clinic, then the increase in cost to all other patients in the walk-in clinic is proportional to these patients' total cost per unit of time from waiting at the walk-in clinic, and the reduction in cost to patients waiting for an appointment is proportional to those patients' total cost per unit of time of waiting for an appointment. If more patients are allocated to appointments, then the situation is the reverse. In each case, the re-allocation of patients leads to externalities that shift the cost from one group of patients to another. The social planner, taking the externalities into account, will want to allocate patients who are almost indifferent to being re-allocated to the location where the externality cost are smaller. This is the result stated in Proposition 3.

One may see that in the example in Figure 1 the optimal allocation has fewer patients attending the walk-in clinic than attend the walk-in clinic in the equilibrium allocation. This is graphically illustrated by the fact that the dashed line is above the solid line. All our numerical calculations have indicated that the equilibrium allocation is actually very close to the optimal allocation. The total cost in the optimal allocation are at most 1% lower than the cost in the equilibrium allocation. The welfare loss in equilibrium has been called the *price of anarchy* in the literature on congestion games. There is a substantial literature on the price of anarchy (e.g. [8]). In our model the price of anarchy seems to be low.

6. Optimal Fees

We now discuss how a social planner could change patients' incentives in a way that ensures that patients' decentralized choices in equilibrium improve on the equilibrium described in Proposition 1, and implement the optimal allocation described in Proposition 2. The specific policy that we consider first is that every patient who attends the walk-in clinic is charged a fee $f^C > 0$ or, alternatively, that every patient who makes an appointment is charged some fixed fee $f^A > 0$.[8] We assume now that patients' cost are the non-material costs described in Section 2 plus the fee, if any, that the patient has to pay. We maintain the assumption that the social planner's objective is solely to minimize average non-material costs, and hence we assume that the social planner ignores the effect of the fee payments on patients' cost. To motivate this we may imagine that the social planner returns the revenue that he or she collects from patients to those patients, dividing the revenue equally among all patients. The planner's payment to agents does not affect patients' incentives, because the payment is independent of agents' choices. Moreover, when calculating patients' average cost, the net average monetary cost are zero, and therefore in the calculation of patients' average cost the fee payment may be ignored, as we assume here.

As before, an equilibrium is a function d^* that maps every pair (p_i, w_i) into either 1 (appointment) or 0 (walk-in clinic) such that for every (p_i, w_i) the choice $d^*(p_i, w_i)$ is optimal if patients correctly anticipate the waiting times given by (1) and (2), where these waiting times are calculated using d^*. The following result generalizes Proposition 1.

Proposition 4. *Suppose every patient who attends the walk-in clinic is charged a fee $f^C > 0$ or, alternatively, that every patient who makes an appointment is charged some fixed fee $f^A > 0$. Then there is a unique equilibrium d^*.*

Proof. Consider first the case that a fixed fee $f^C > 0$ has to be paid for every visit to the walk-in clinic. A patient i optimally makes an appointment if and only if $p_i t^A < w_i t^C + f^C$. Because $t^C = 1 - t^A$, we can re-write this as: $p_i t^A < w_i(1 - t^A) + f^C$. Consider any possible value of t^A. We have an equilibrium if and only if the set of patients for whom the condition $p_i t^A < w_i(1 - t^A) + f^C$ holds, has G-measure t^A. Denote the G-measure of this set by $\mu(t^A)$. Note that $\mu(0) = 1$, that μ is continuous and decreasing in t^A, and that $\mu(1)$ is the G-measure of patients for whom p_i is less than f^C. Therefore, if $f^C \geq p_{\max}$, then in equilibrium all patients will make an appointment, and this will be the only equilibrium. If $f^C < p_{\max}$, then there will be a unique fixed point of μ. This fixed point corresponds to the unique equilibrium of the model. An analogous argument proves the Proposition for the case that a fixed fee $f^A > 0$ has to be paid for every scheduled appointment. □

If $f^C > 0$, the condition for a patient to optimally choose an appointment:

$$\begin{aligned} p_i t^A &< w_i t^C + f^C \Leftrightarrow \\ p_i &< \frac{t^C}{t^A} w_i + \frac{f^C}{t^A}. \end{aligned} \tag{10}$$

[8]The fees that we discuss in this section are the analogue in our model of the *Pigouvian taxes* that are discussed in the literature on competitive markets and externalities (e.g. [6, p. 355]).

Similarly, if $f^A > 0$, the condition for a patient to optimally choose an appointment is:

$$\begin{aligned} p_i t^A + f^A &< w_i t^C \Leftrightarrow \\ p_i &< \frac{t^C}{t^A} w_i - \frac{f^A}{t^A}. \end{aligned} \tag{11}$$

Note that the inequalities in Proposition 2 are identical to the inequalities above, with the fee chosen to be equal to the intercept $\hat{a}$, where a positive intercept corresponds to a fee f^C, and a negative intercept corresponds to a fee f^A. This motivates the following definition.

Definition 2. *Let $\hat{d}$ be a regular allocation that minimizes (5) among all regular allocations. The* optimal fee policy corresponding to $\hat{d}$ *charges a fee $f^C = g^C \bar{w}^C - g^A \bar{p}^A$ for every visit to the walk-in clinic if $g^C \bar{w}^C > g^A \bar{p}^A$, and it charges a fee $f^A = g^A \bar{p}^A - g^C \bar{w}^C$ for every scheduled appointment if $g^A \bar{p}^A > g^C \bar{w}^C$. Here, the relevant values of the variables $g^C, \bar{p}^C, g^A$ and $\bar{w}^A$ are the values which these variables take in the optimal allocation $\hat{d}$.*

From Propositions 2 and 4 we can infer the following result:

Corollary 1. *Let $\hat{d}$ be a regular allocation that minimizes (5) among all regular allocations. Then the unique equilibrium under the optimal fee policy corresponding to $\hat{d}$ is $\hat{d}$ itself.*

Proof. By Proposition 2 the optimal allocation satisfies precisely the inequalities required for an equilibrium. By Proposition 4 there is a unique such equilibrium. □

In the example in Figure 1, the planner will optimally charge a fee for every visit to the walk-in clinic. This fee moves the allocation towards the optimal allocation because it deters agents from attending the walk-in clinic.

7. Finding Optimal Fees

To implement the optimal fee policy the health care provider needs to know the values of $g^C, \bar{w}^C, g^A$ and $\bar{p}^A$ for an optimal allocation $\hat{d}$. To elicit this information a health care provide may use an incentive scheme known from game theory as a *Vickrey-Clarke-Groves (VCG) mechanism* ([2, Definition 5.4]). VCG-mechanisms are defined for finite populations of patients. Therefore we change the model at this point, and assume that the set of patients is the finite set : $I = \{1, 2, \ldots N\}$. Each patient $i \in I$ has characteristics $(p_i, w_i) \in [0, p_{\max}] \times [0, w_{\max}]$. Let d_i denote patient i's decision: $d_i \in \{0, 1\}$. The waiting time in each location equals the proportion of patients choosing that location:

$$t^A = \sum_{i=1}^{N} \frac{d_i}{N} \quad \text{and } t^C = 1 - t^A. \tag{12}$$

Except for these modifications, the model is as before.

The VCG-mechanism is now as follows: Each patient i reports the characteristic (p_i, w_i) to the planner. The planner then determines the efficient outcome on the basis of these reports, and assigns patients accordingly to either the walk-in clinic, or asks them to wait for an appointment. We write $\hat{d}_i \in \{0,1\}$ for the assignment that the planner determines for patient i. To provide incentives to patients to report their characteristics truthfully, the planner also charges fees. Define:

$$\mathcal{C}_{-i} = \left(\sum_{\{j \in I : j \neq i \text{ and } \hat{d}_j = 0\}} w_j - \sum_{\{j \in I : j \neq i \text{ and } \hat{d}_j = 1\}} p_j \right) \frac{1}{N} \tag{13}$$

This is the externality that patient i imposes on all other agents if i is given an appointment, and is not assigned to the walk-in clinic. If this is positive, then patient i pays $\mathcal{C}_{-i}$ if assigned to an appointment, and nothing if i attends the walk-in clinic. If $\mathcal{C}_{-i}$ is negative, then patient i pays $-\mathcal{C}_{-i}$ if assigned to the walk-in clinic, and nothing if i makes an appointment.[9] Note fees are allowed to be different for different patients.

It is well-known that the VCG-mechanism provides incentives to patients to report their characteristics truthfully in the following sense: Whatever other patients report it is always at least one of potentially several cost-minimizing strategies for patient i to report his or her characteristics (p_i, w_i) truthfully ([2, Proposition 5.2]). If all patients report their characteristics truthfully, then obviously the VCG-mechanism implements the optimal patient allocation.

It is interesting to relate the VCG-mechanism to the optimal fee policy described above. For this we consider the VCG-mechanism assuming that each patient's characteristics are an independent draw from the distribution G, and we consider the limit for $N \to \infty$, so that the finite population model approaches the continuum population model. It is easy to see that the fees paid by each patient in the VCG-mechanism converge, for $N \to \infty$, almost surely to the fees of the optimal fee policy. In particular, the fees will be the same for all patients. Thus the optimal fee policy is the limit of the fees imposed by the VCG-mechanism.

There is, as well, an important difference between the VCG-mechanism and the optimal fee policy. The VCG-mechanism is a centralized mechanism: it collects information from all agents, and then assigns agents to one of the two different options, walk-in clinic or appointments. By contrast, under the optimal fee policy, the choice between the two options is left to patients themselves. In this sense the optimal fee policy is decentralized.[10] However, as we pointed out earlier, the optimal fee policy does not clarify how the information that is needed to implement this policy is acquired.

[9] This applies if patient is optimal allocation changes, depending on i's report. If it does not change, then i pays no fee.

[10] For a discussion of related ideas see [14].

In practice, a mixture of both approaches might be realistic: A random sample of patients is asked to report their characteristics. Fees are then determined on the basis of the VCG-mechanism. Fees will be potentially different for different patients. But if the random sample is sufficiently large this difference will become negligible. The fees that are based on the VCG-mechanism can then be applied to all other patients, not included in the random sample. All other patients can then make their choices by themselves, without having to provide information to the health care provider.

We have numerically investigated an even simpler directed search algorithm that the healthcare provider might use to find optimal fees. In this algorithm the healthcare provider evaluates only the total cost per time unit suffered by those attending the walk-in clinic, and the total cost per time unit suffered by those waiting for an appointment. The health care provider then gradually adjust the fees so that the difference of fees reflects the difference of these cost per time unit. This algorithm has converged to a very close neighborhood of the optimal fees in all simulations that we have conducted.

8. Conclusion

We have modeled health care access as a congestion game, and have discussed policies that a health care provider might use to shift the equilibrium allocation to the optimal allocation. In our model it turns out that a simple fixed fee is sufficient to accomplish this objective. We have also discussed ways in which the health care provider might collect the information needed to determine the optimal fees. In future work we plan to investigate other systems of health care access, such as systems that rely on an initial evaluation of patients by a member of the health care system, or systems that accord each patient a fixed number of *credit points* for walk-in clinic visits.

References

[1] Ashton, R., Hague, L., Brandreth, M., Worthington, D. and Cropper, S. (2005), A simulation-based study of NHS walk-in centre, *Journal of the Operational Research Society*, 56, 153-161.

[2] Börgers, T. (2010), *An Introduction to the Theory of Mechanism Design*, forthcoming, *Oxford University Press*, available at: `http://www-personal.umich.edu/ tborgers/`.

[3] Fudenberg, D. and Tirole, J. (1991), *Game Theory*, Cambridge and London: MIT Press.

[4] Gallagher, M., Pearson, P., Drinkwater, C. and Guy, J. (2001), Managing Patient Demand: A Qualitative Study of Appointment Making in General Practice, *British Journal of General Practice*, 51(465), 280-285.

[5] Institute of Medicine (2001), *Crossing The Quality Chasm: A New Health System For The 21st Century*, Washington, DC: National Academy Press. Available at: `http://www.nap.edu/books/0309072808/html/index.html`.

[6] Mas-Colell, A., Whinston, M. and Green, J. (1995), *Microeconomic Theory*, New York and Oxford: Oxford University Press.

[7] Reilly, T.A., Marathe, V.P., and Fries, B.E. (1978), A delay-scheduling model for patients using walk-in clinic, *Journal of Medical Systems*, vol. 2, no. 4, 303-313.
[8] Bhawalkar, K., Gairing, M., and Roughgarden, T. (2010), Weighted Congestion Games: Price of Anarchy, Universal Worst-Case Examples, and Tightness, *Euopean Symposium on Algorithms*, Liverpool.
[9] Milchtaich, I. (1996), Congestion Games with Player-Specific Payoff Functions, *Games and Economic Behavior* 13, 111-124.
[10] Milchtaich, I. (2000), Generic Uniqueness of Equilibrium in Large Crowding Games, *Mathematics of Operations Research* 25, 349-364.
[11] Murray, M. and Berwick, D.M (2003), Advanced Access: Reducing Waiting and Delays in Primary Care, *Journal of the American Medical Association* 289(8), 1035-1040.
[12] Reid, P.P., Compton, W.D., Grossman, J.H. and Fanjiang, G. (editors) (2005), *Building a Better Delivery System: A New Engineering/Health Care Partnership*, Washington, D.C: National Academy Press.
[13] Rosenthal, R. (1973), A Class of Games Possessing Pure-Strategy Nash Equilibria, *International Journal of Game Theory* 2, 65-67.
[14] Sandholm, W. (2002), Evolutionary Implementation and Congestion Pricing, *Review of Economic Studies* 69, 667-689.
[15] Schmeidler, D. (1973), Equilibrium Points of Non-atomic Games, *Journal of Statistical Physics* 7, 295-301.
[16] Su, S. and Shih, C.L. (2003), Managing a mixed-registration-type appointment system in outpatient clinics, *International Journal of Medical Informatics*, 70, pp. 31-40.
[17] Wijewickrama, A., and Takakuma, S. (2005), Simulation analysis of appointment scheduling in an outpatient department of internal medicine, *Proceedings of 2005 Winter Simulation Conference*, 2264-2273.