Provider–client interactions and quantity of health care use

Hsien-Ming Lien [a,*], Ching-To Albert Ma [b], Thomas G. McGuire [c]

[a] *Department of Public Finance, National Cheng-Chi University; 64, Sec. 2, Zhi-nan Road, Wenshan, Taipei 116, Taiwan, ROC*
[b] *Department of Economics, Boston University, 270 Bay State Road, Boston, MA 02215, USA*
[c] *Department of Health Care Policy, Harvard Medical School, 180 Longwood Avenue, Boston, MA 02115, USA*

Received 16 October 2002; received in revised form 24 March 2004; accepted 24 March 2004
Available online 21 July 2004

Abstract

This paper considers three types of provider–client interactions that influence quantity of health care use: rationing, effort, and persuasion. By rationing, we refer to a quantity limit set by a provider; effort, the productive inputs supplied by a provider to increase a client's demand; persuasion, the unproductive inputs used by a provider to induce a client's demand. We construct a theoretical model incorporating all three mechanisms as special cases. When the general model is specialized into one of three mechanisms, a set of empirical implications emerges. We test for the presence of each mechanism using data of patients receiving outpatient treatment for alcohol abuse in the Maine Addiction Treatment System. We find evidence for rationing and persuasion, but not effort.

Keywords: Physician induced demand; Rationing and effort

1. Introduction

The theoretical and empirical literatures in health economics take different approaches to the issue of provider induced demand. In the empirical literature, the term "provider induced demand" (PID) is used broadly to account for a range of empirical findings. Many studies have shown that when common demand-side variables (such as demand price, income and clinical needs) are controlled for, consumer demands continue to be affected by physician

* Corresponding author.
E-mail addresses: hmlien@nccu.edu.tw (H.-M. Lien), ma@bu.edu (C.-T. Albert Ma), mcguire@hcp.med.harvard.edu (T.G. McGuire).

or provider variables. These supply-side variables include the number of competitors in a market (Cromwell and Mitchell, 1986), supply price (Rice, 1983; Yip, 1998), incentives for self-referral because of ownership of complementary inputs (Hillman et al., 1992), physician attitudes toward earnings (Rizzo and Blumenthal, 1994), demand shocks due to demographic changes (Gruber and Owings, 1996), incentives due to malpractice liability (Kessler and McClellan, 1996), and partnership incentives (Gaynor and Gertler, 1995). Empirical papers generally attribute these effects to PID.

The theoretical literature uses the term PID more narrowly. Referring to Fuchs (1978), Pauly (1980), Eisenberg (1986), Cuyler (1989) and Williams (1998), we say that PID exists when the physician influences a patient's demand for care not necessarily in the patient's best interest. PID may be regarded as a form of manipulation: the patient is persuaded to demand health services. Besides persuasion, other ways by which a physician influences utilization have also been identified. A provider can set quantity directly; this is possible because health care services are nonretradable (Farley, 1986; Gaynor, 1994; De Jaegher and Jegers, 2000). When a provider's quantity setting power is exercised, deliberate influence on the patient's demand is unnecessary; the patient simply accepts or rejects the provider's decision. Third, the provider can undertake an action ("effort" or "quality"), noncontractible but observable to the patient. The effort influences the patient's valuation of services, therefore affecting demand (Ma and McGuire, 1997). Unlike the inducement behind PID, the physician's action or effort is productive.

The three mechanisms are conceptually distinct (see McGuire, 2000 for a review). Under PID, the key word is *induce*; the patient's use of health services may be unrelated to his true demand. In the quantity-setting mechanism, when the physician unilaterally picks quantity, the patient may only accept or refuse. In the action–effort mechanism, a patient reacts rationally against the physician's effort. These mechanisms have different implications for a normative view of how health care markets function. It is important, therefore, for both academic research and policy, to apply carefully the ideas of theoretical papers to guide empirical work. In this paper, we take a step in this direction by analyzing a health care setting in which providers may influence quantities. Specifically, we use the data to test for the presence of the three mechanisms of influence.

Section 2 reviews the three mechanisms of provider influence on patient decision making. The theoretical model presented there incorporates all three mechanisms as special cases. When the general model is specialized into one of the three mechanisms, a specific set of predictions emerges. These predictions allow us to test for the presence of each mechanism. The statistical method for identifying the relative roles of the three mechanisms is presented in Section 3. Our data set, described in Section 4, is from public sector drug and alcohol treatment facilities. It contains more information than common health care utilization studies. In particular, we have information on whose decision (the client's or the provider's) it was to terminate a course of treatment, as well as on treatment outcomes. The first piece of information helps to distinguish the quantity setting mechanism from effort and persuasion; the second separates productive effort from unproductive persuasion. Our empirical results are presented in Section 5. We find evidence for two of three mechanisms: direct quantity setting and persuasion. The mix of these three mechanisms is likely to differ across health care settings, but the perspective and the empirical methods proposed here can be used elsewhere. We discuss these prospects in Section 6.

2. Mechanisms of provider influence on health care use

The health economics literature places significant emphasis on the modeling of physician–patient interactions. Such interactions determine the quantities of care supplied to patients. Table 1, adapted from McGuire (2000), summarizes the three mechanisms in the literature. Common to almost all papers on this topic is a monopolistically competitive market structure, but different informational and contractibility assumptions have been made. Perhaps the most straightforward account of quantity determination is persuasion, associated with the writings of Dranove (1988), Evans (1974), Fuchs (1978), Rice (1983) and many others (see the column under "Persuasion" in Table 1). Persuasion is simply regarded as a form of inducement that shifts the entire demand curve. Inducement itself has no effect on the health outcome, though any quantity demanded following from inducement may affect outcome. To complete the persuasion theory, some cost to the provider of inducement must be specified. Dranove (1988) contains a model in which the profit-maximizing amount of inducement is limited by the patient's suspicion of the provider's aggressiveness; other approaches include the resource cost of the activity (Stano, 1987), and increasing professional discomfort (Evans, 1974; McGuire and Pauly, 1991).

The second mechanism regards the health care provider as a quantity setter; see the column under "Nonretradability" in Table 1. Health care providers supply a nonretradable service (Farley, 1986; Gaynor, 1994). A monopolistic competitor selling a nonretradable service sets a quantity to maximize profit. Recognition of physician quantity setting power stemming from nonretradability accords well with most patients' experiences with physicians who, quite simply, "tell them what to do". Of course, the theory must also recognize that consumers need not always comply with physicians' instructions. Generally, one can interpret the quantity setting model as one in which quantity restrictions are placed by physicians, and patients respond to these restrictions. Empirically, this theory says that observed quantity results from the physician's quantity restriction and the patient's response. Also, like the persuasion mechanism, the quantity setting model does not involve any other input into health care production.

Table 1
Determinants of provider–patient interactions on quantity of health care

	Persuasion	Nonretradability allows quantity setting	Choice of noncontractible input
Main features	Physicians take actions to persuade; constrained by demand response or ethics	Supply determination within demand constraints	Demand response to "quality" or other costly physician inputs
Illustrative paper	Dranove (1988)	Farley (1986)	Ma and McGuire (1997)
Information	Asymmetric	Complete	Complete
Physician's action in influencing use	Unobservable	N/A	Noncontractible
Physician's action in influencing outcome	No	N/A	Yes

The third mechanism posits that a physician supplies a noncontractible input in health production; this is additional to other measurable quantities of treatments such as days, visits, or tests. This noncontractible input can be regarded as quality, such as the time and effort with which the physician tailors the treatment to the patient's needs, or the physician's care in delivering services. For our purpose, the important aspect of quality is that it affects health outcomes as well as the patient's behavior. Knowing a physician's quality, a patient will respond by demanding more or less care, depending on the productive relationship between the two inputs. Ma and McGuire (1997) developed such a theory. That paper models patient–physician interaction symmetrically—one input (quality) is chosen by the physician, another (quantity) by the patient. Notably, effort, a productive input, affects health outcome.

In this section, we develop a general model which yields each of the above three mechanisms as special cases. Each special case generates an empirical hypothesis, which we will test with our data. Our model generalizes Ma and McGuire (1997) in several directions. First, the possibility of persuasion is added. Second, we explicitly model provider quantity setting by a limit on utilization. Third, the provider does not know patients' preferences with certainty.

We begin with the patient's utility function: $\theta U(q, e) - p_\mathrm{d} q$, where θU is the patient's benefit of receiving treatment quantity q when either the physician's amount of persuasion or quality is e. (So for the moment, e represents either persuasion or quality.) The parameter θ can be regarded as the patient's "type," its value being known only to the patient. The physician only knows that θ follows the distribution function F with density f. The fee that the patient pays per unit of q is p_d.

Given e, a patient responds by choosing the optimal quantity. A patient with parameter θ picks q to solve:

$$\max_q \theta U(q, e) - p_\mathrm{d} q. \tag{1}$$

The optimal quantity satisfies the first-order condition:

$$\theta \frac{\partial U(q, e)}{\partial q} = p_\mathrm{d}. \tag{2}$$

From (2), a patient's quantity choice is affected by θ and e. Let the optimal quantity be $q(\theta, e)$. The comparative statics with respect to q can be obtained by the total differentiation[1] of (2):

$$U_q(q, e)\,\mathrm{d}\theta + \theta[U_{qq}(q, e)\,\mathrm{d}q + U_{qe}(q, e)\,\mathrm{d}e] = \mathrm{d}p_\mathrm{d}.$$

Setting $\mathrm{d}p_\mathrm{d} = 0$, we obtain:

$$\frac{\partial q}{\partial \theta} = -\frac{U_q(q, e)}{\theta U_{qq}(q, e)} > 0, \qquad \frac{\partial q}{\partial e} = -\frac{U_{qe}(q, e)}{U_{qq}(q, e)}.$$

If $U_{qe}(q, e) > 0$ (input complementarity), then $\partial q/\partial e > 0$. This implies that a patient selects a higher treatment quantity when the physician uses the higher effort or persuasion level.

[1] Subscripts of functions denote their partial derivatives.

The physician is assumed to maximize her expected profit, the expectation being taken over the patient's preference parameter θ. The provider selects effort or persuasion to affect the patient's demand, along with a quantity ceiling that restricts the maximum quantity of care a patient can receive. The extensive form of the game is described as follows. In the first stage, knowing only the distribution function $F(\theta)$, the physician picks an effort or persuasion level e as well as a quantity ceiling Q. In the second stage, the patient observes Q, θ, and e, and picks quantity q, where $q \leq Q$.

Consider the second stage when the patient must choose a quantity. Given e, suppose the patient's choice is given by the maximization of (1). If that yields a solution higher than Q, then Q will be the actual quantity. Otherwise, it will be given by (2). From (2), given Q and e, there exists a Θ such that $q(\theta, e) \leq Q$ if and only if $\theta \leq \Theta$, or $q(\Theta, e) = Q$. Given e, a choice of a ceiling, Q, is equivalent to a choice of Θ.

Let p_s be the price that a physician receives per unit of quantity, and $C(q, e)$ the physician's cost function when q units of quantity and e units of effort or persuasion are used. The physician's expected profit is

$$\begin{aligned}\Pi(q, e, \Theta) \equiv \int_0^{\Theta} [p_s q(\theta, e) - C(q(\theta, e), e)]\, \mathrm{d}F(\theta) \\ + [1 - F(\Theta)][p_s q(\Theta, e) - C(q(\Theta, e), e)]. \end{aligned} \tag{3}$$

For those patients with θ below Θ, the quantity ceiling does not bind, and the profit is given by the integrand of the first term of (3); for θ above Θ, the quantity ceiling binds, and the last term in (3) is the profit. The physician picks e and Θ to maximize (3) subject to (2) for $\theta \leq \Theta$. The first-order conditions are

$$\frac{\partial \Pi}{\partial \Theta} = (p_s - C_q(Q, e))[1 - F(\Theta)]\frac{\partial q(\Theta, e)}{\partial \theta} = 0 \tag{4}$$

$$\begin{aligned}\frac{\partial \Pi}{\partial e} &= \int_0^{\Theta} (p_s q_e - C_e - C_q q_e)\, \mathrm{d}F(\theta) + [1 - F(\Theta)][p_s q_e(\Theta, e) - C_e - C_q q_e(\Theta, e)] \\ &= \int_0^{\Theta} (p_s q_e - C_e - C_q q_e)\, \mathrm{d}F(\theta) - [1 - F(\Theta)]C_e = 0, \end{aligned} \tag{5}$$

where the second equality of (5) follows from (4). The first-order condition (4) shows why a physician would set a quantity limit. When the value of θ is high, the patient demands a large quantity of care. The marginal cost function, however, is increasing and convex. For some given price p_s the physician will make a loss for those patients who have high values of θ. Installing a quantity limit avoids these losses. In fact (4) simply says the marginal cost at the quantity limit Q is exactly equal to the physician's price p_s.

In the general model, the equilibrium quantity is a result of the patient's choice or the physician's quantity limit. In each case, the equilibrium quantity is a function of the patient's and physician's prices, p_d and p_s, as are the equilibrium effort e and the quantity limit Q. Two special cases of the general model will correspond, respectively, to pure quantity setting and pure effort–persuasion theories of health care utilization.

For pure quantity setting, the utility function U is assumed to be independent of e so that the demand function q only depends on θ and price p_d. Then the equilibrium quantity limit will be $p_s = C_q(Q)$. The equilibrium quantity limit Q will only depend on p_s, not on p_d. Also, when the quantity limit does not bind, for those consumers with low values of θ, the equilibrium quantity will be given by $\theta U_q(q) = p_d$. We summarize the dichotomy by the following result.

Proposition 1 (Absent effort or persuasion). *In the pure quantity setting model, (i) when the quantity of treatment is less than the ceiling set by the physician, quantity is a function of the patient's price p_d but not the physician's price p_s; (ii) the physician's quantity ceiling is a function of p_s but not p_d.*

Next, we consider pure effort or persuasion. In the general model, the quantity ceiling is chosen to limit demands of patients with high values of θ. The pure effort–persuasion model can therefore be thought as a case in which Q is set at an arbitrarily high level: the physician must accept any quantity demanded by the patient. In this case, the second term in (3) vanishes and the equilibrium effort or persuasion will be given by the following first-order condition:

$$\int_0^\infty (p_s q_e - C_e - C_q q_e)\,\mathrm{d}F(\theta) = 0. \tag{6}$$

The equilibrium effort depends on p_s directly, and indirectly on p_d through the demand function (2). In turn, the equilibrium quantity depends on both p_s and p_d.

Proposition 2 (Absent quantity ceiling). *In the pure effort–persuasion model, quantity is a function of both patient's and physician's prices, p_s and p_d.*

The above propositions will be used for empirical tests of the pure quantity and pure effort–persuasion models. We now propose a method to distinguish between quality and persuasion. To accommodate the presence of both (unproductive) persuasion and (productive) effort, we expand the patient's subjective benefit function to $U(H(q, \hat{e}), \tilde{e})$, where H denotes a patient's health, and is a function of quantity of health care q and productive effort, $\hat{e}$. Subjective benefit also depends on a physician action, $\tilde{e}$, the persuasive activity that affects patient's demand but not health. We observe measures of health in our data, allowing us to see if the physician action influencing demand works through or independently of a health effect. We test the special case of our model in which H is a function only of q, not $\hat{e}$. This corresponds to a pure persuasion model. We now state:

Proposition 3 (Effort versus Persuasion). *In a pure persuasion model, health outcomes are unaffected by inducement, which depends on patient's and physician's prices, p_s and p_d. After effects from quantities are controlled for, health outcomes do not depend on prices in the pure persuasion model.*

3. Estimation strategy

In this section we use Propositions 1–3 for empirical tests. Our data contain several special features. First, as in many health care data sets, prices paid by clients are not necessarily equal to those received by providers, due to insurance and reimbursement practices. As a result, we can separately identify demand and supply prices. Second, our data contain information about whose decision it was to end treatment for alcohol abuse. The client may "quit," or the provider may stop the episode by calling it "complete." Using this information we say that the provider's quantity limit is binding, if and only if, an episode is complete. Finally, our data include information on health outcomes. This information enables us to test whether the provider's action is persuasion or effort.

3.1. The general specification

We begin with an empirical specification of the general model. Let q denote the quantity demanded by a client; Q, the quantity limit set by the provider. Let q^* be the observed quantity. Given that q^* is selectively observed, the reduced form model can be described by the disequilibrium model in the literature,[2] such as:

$$q = X_{\mathrm{d}}\beta_{\mathrm{d}} + \delta_{\mathrm{d}} p_{\mathrm{d}} + \gamma_{\mathrm{d}} p_{\mathrm{s}} + \mu_{\mathrm{d}}, \tag{7}$$

$$Q = X_{\mathrm{s}}\beta_{\mathrm{s}} + \delta_{\mathrm{s}} p_{\mathrm{d}} + \gamma_{\mathrm{s}} p_{\mathrm{s}} + \mu_{\mathrm{s}}, \tag{8}$$

$$q^* = jq + (1 - j)Q, \quad j = I(q \le Q), \tag{9}$$

where $I(\cdot)$ equals to one if the event in the bracket is true and zero otherwise. Eqs. (7) and (8) describe a client's demand and a provider's limit, respectively. Both a client's demand (q) and the provider's limit (Q) are functions of the supply price and the demand price in the general model. The vector X_{d} represents factors influencing a client's demand, such as income, health and demographic variables, and μ_{d} is an error term that denotes unobserved client characteristics. Likewise, the vector X_{s} represents factors affecting a provider's limit and μ_{s} is an error term that captures unmeasured provider features. Because the provider may set the limit based on a patient's observed characteristics, X_{s} may contain elements of X_{d}. In addition, μ_{d} and μ_{s} may be correlated if the provider's limit is chosen according to some unobserved demand characteristics. Finally, Eq. (9) says that the observed quantity q^* is either a client's demand or a provider's limit, depending on j; the indicator j equals to one if Q is no less than q and zero otherwise. Notice that the data contain information on the indicator j. Thus, for each client we know exactly whether the observed quantity is a provider's limit or a client's demand.

The system of Eqs. (7)–(9) assumes a static perspective on treatment: q and Q are both chosen independently at the beginning of a course of treatment. In practice, the assumption may be problematic since the two decisions could affect one another. For instance, a client may choose his demand according to the provider's quantity limit. To adapt the model for our setting, we allow q and Q to have some relations: the client chooses the doctor's limit Q when it is not much bigger than q. Specifically, we modify Eq. (9) to

$$q^* = jq + (1 - j)Q, \quad j = \begin{cases} 1 & \text{if } q \le Q - \alpha + \varepsilon \\ 0 & \text{otherwise} \end{cases}. \tag{10}$$

[2] For the detailed discussion on the disequilibrium model, see Maddala (1983) for a review.

The parameter α is positive, and ε a random error with zero mean. Eq. (10) says that on average whenever the client's desired demand is within α units of the provider's limit, he revises his demand upward and accepts the provider's limit.

Assume that μ_{d} and μ_{s} are jointly normal with zero means, standard deviations σ_{d} and σ_{s}, respectively, and correlation ρ. Assume also that ε is normally distributed with a zero mean, standard deviation σ_{ε}, but uncorrelated with either μ_{d} or μ_{s}. The likelihood function for a client whose observed health care use is q^*, and the indicator j can be expressed as

$$\begin{aligned}
&\ln f(q^*, j) \\
&\quad = j\Bigg[\ln\phi\left(\frac{q^* - X_{\mathrm{d}}\beta_{\mathrm{d}} - \delta_{\mathrm{d}}p_{\mathrm{d}} - \gamma_{\mathrm{d}}p_{\mathrm{s}}}{\sigma_{\mathrm{d}}}\right) - \ln\sigma_{\mathrm{d}} \\
&\quad + \ln\Phi\left(\frac{(\rho\sigma_{\mathrm{d}}/\sigma_{\mathrm{s}})(q^* - X_{\mathrm{d}}\beta_{\mathrm{d}} - \delta_{\mathrm{d}}p_{\mathrm{d}} - \gamma_{\mathrm{d}}p_{\mathrm{s}}) - \alpha - (q^* - X_{\mathrm{s}}\beta_{\mathrm{s}} - \delta_{\mathrm{s}}p_{\mathrm{d}} - \gamma_{\mathrm{s}}p_{\mathrm{s}})}{\sqrt{\sigma_{\mathrm{s}}^2(1-\rho^2) + \sigma_{\varepsilon}^2}}\right)\Bigg] \\
&\quad + (1-j)\Bigg[\ln\phi\left(\frac{q^* - X_{\mathrm{s}}\beta_{\mathrm{s}} - \delta_{\mathrm{s}}p_{\mathrm{d}} - \gamma_{\mathrm{s}}p_{\mathrm{s}}}{\sigma_{\mathrm{s}}}\right) - \ln\sigma_{\mathrm{s}} \\
&\quad + \ln\Phi\left(\frac{(\rho\sigma_{\mathrm{s}}/\sigma_{\mathrm{d}})(q^* - X_{\mathrm{s}}\beta_{\mathrm{s}} - \delta_{\mathrm{s}}p_{\mathrm{d}} - \gamma_{\mathrm{s}}p_{\mathrm{s}}) + \alpha - (q^* - X_{\mathrm{d}}\beta_{\mathrm{d}} - \delta_{\mathrm{d}}p_{\mathrm{d}} - \gamma_{\mathrm{d}}p_{\mathrm{s}}}{\sqrt{\sigma_{\mathrm{d}}^2(1-\rho^2) + \sigma_{\varepsilon}^2}}\right)\Bigg]
\end{aligned} \tag{11}$$

where ϕ and Φ are the density and cumulative density functions of the standard normal distribution, respectively. The exact variables used in the estimation are specified below.

3.2. *Do providers use effort or persuasion?*

We use Proposition 1 to test for the presence of effort or persuasion (for now we do not distinguish between them). Under a quantity ceiling restriction, Eqs. (7)–(9) reduce to

$$q = \beta_{\mathrm{d}}X_{\mathrm{d}} + \delta_{\mathrm{d}}p_{\mathrm{d}} + \mu_{\mathrm{d},} \tag{12}$$

$$Q = \beta_{\mathrm{s}}X_{\mathrm{s}} + \gamma_{\mathrm{s}}p_{\mathrm{s}} + \mu_{\mathrm{s}}, \tag{13}$$

$$q^* = jq + (1-j)Q, \tag{14}$$

where the indicator j is defined by (10). Recall that p_{s} and p_{d} are included in Eqs. (7) and (8) because of the provider's endogenous effort or persuasion. Without effort or persuasion, q does not depend on p_{s}; nor does Q on p_{d}. The joint restriction $\delta_{\mathrm{s}} = \gamma_{\mathrm{d}} = 0$ is therefore used for testing the presence of effort or persuasion.

3.3. *Do providers set a quantity ceiling?*

We use Proposition 2 to test for the presence of a quantity limit. Under pure effort–persuasion restriction, the quantity limit Q is set to be above of all patients' demands. As a result, the observed quantity is always a client's demand and can be fully captured by the regressors in (7). Formally, under pure effort or persuasion, Eqs. (7)–(9) simply have q^* $= q$, and reduce to

$$q^* = X_d\beta_d + \delta_d p_d + \gamma_d p_s + \mu_d. \tag{15}$$

There are a number of ways to test for the presence of a quantity limit. The simplest way is to use the indicator function. Proposition 2 restricts q^* to be q and j in (10) to 1. Hence, the presence of a quantity limit can be tested by checking the proportion of clients whose demands are constrained. As will show later in the data, about 40% of clients' demands are limited, strongly supporting the presence of a quantity limit.

The above method, though straightforward, relies heavily on the quality of separation indicator. A more conservative way is to check whether q^* can be fully explained by the demand equation. Rewrite (15) as

$$q^* = X_d\beta_d + \delta_d p_d + \gamma_d p_s + \tau\hat{q}_g + \mu_d,$$

where $\hat{q}_g$ denotes the predicted quantity from the general model. The coefficient τ is expected to be one under the general model. In pure effort–persuasion restriction, however, τ would equal zero since the regressors in (7) fully explain q^*.

Lastly, the quantity ceiling restriction can also be tested by comparing the likelihood value of the general model with that of the one with an arbitrary large Q. Because the likelihood function consists of both j and q^*, this method can be thought as a combination of the above two methods. Nevertheless, the statistical power of the test depends on the chosen value of Q. To be conservative, we choose the value of Q that minimizes the likelihood function of (11), and then use that likelihood value for empirical tests. Hence, Proposition 2 is tested by the constraint: $\beta_s = \delta_s = \gamma_s = 0$ (except for the constant term in X_s).

3.4. *Do providers use effort?*

We use Proposition 3 to test for the presence of effort. To derive a testable hypothesis, we need to specify the health production function. Let H_1 and H_0 denote a client's health at admission and discharge, respectively. A client's health outcome H_1 is specified as a linear combination of a client's health at admission (H_0), personal characteristics (X_d), health care use (q^*), and effort or persuasion (e), namely,[3]

$$H_1 = \alpha_0 H_0 + \alpha_{1d} X_d + \alpha_2 q^* + \alpha_3 e + \eta_h, \tag{16}$$

where η_h denotes a random error. The above specification allows e to affect H_1 directly or indirectly through q^*. Proposition 3 says that conditional on q^*, H_1 is unaffected by persuasion.

Neither effort or persuasion is observed. To complete the estimation, we use the supply price as a proxy for e. Under pure persuasion, Proposition 3 implies that $\alpha_3 = 0$.

[3] Our specification of treatment production function is similar to Lu and McGuire (2002), except that they excluded effort as an input in the production function.

4. The data

Our data source is the Maine Addiction Treatment System (MATS). The data describe people receiving substance abuse treatment at any publicly funded facility in Maine. About 1% of Maine's 1.2 millions residents have contact with a public substance abuse agency in a year (Lu and McGuire, 2002). A large portion of the service costs of substance abuse clients are funded by the state regulatory agency,[4] the Office of Substance Abuse (OSA).[5] For evaluation purposes, each agency receiving OSA funds is required to report information for every treated episode. In total, MATS collected records of over 30,000 treatment episodes from 82 agencies, spanning the time between October 1990 and September 1995.[6]

MATS collects treatment information through standardized forms completed during personal interviews at admission and discharge.[7] The admission form covers extensive questions about a client's background, such as demographics, income, employment and legal status, as well as an initial assessment of a client's health status. The discharge form contains comprehensive information on services provided in an episode, including providers, expected payment sources, delivered services (types of treatment, number of visits, etc.), along with an assessment of a client's health condition at that time. In particular, the discharge form records which party decided to terminate the treatment episode. We regard quantity as the provider's limit if and only if a provider says that a client has completed treatment. Because of the abundant information in the data set, MATS has been used by Lu and McGuire (2002), Machado (2001), and Ackerberg et al. (2001) for the study of treatment productivity, and by Shen (1999) for the study of strategic risk selection by providers.

Our analysis focuses on outpatient treatment for alcohol abuse, the most prevalent form of treatment in MATS.[8] We further restrict the sample in a number of ways. First, due to institutional reasons, some clients in the data could not choose quantity.[9] We ignore these clients because their treatment behaviors are different from our model. Second, our analysis relies on the client's payment source to construct proxies for prices. Clients with no recorded payment sources in the data are therefore dropped. Finally, in order to control

[4] The total budgeted expenditure of the Office of Substance Abuse was $10,085,716, or about $8 per capita in 1995. Around 60% of total budget is used to finance treatment costs for indigent clients. Source: Office of Substance Abuse 1995 Data Book; State of Maine Budget Document, 1994–1995.

[5] The Department of Human Service was the responsible agency prior to the creation of OSA. OSA was created in July 1990 as a branch of the State's Executive Department. After 1 July 1996, OSA was transferred to the Department of Mental Health, Mental Retardation, and Substance Abuse Service. OSA was responsible for allocating state and federal funds for substance abuse, and for contracting with agencies that provide substance abuse services.

[6] Our data actually contain MATS episodes over 7 years, from September 1989 to October 1996 (FY90–96). However, the data for FY90 and FY96 are truncated, resulting a much smaller number of episodes compared to other sample years. In addition, many episodes in these 2 years are incomplete due to truncation. The analysis therefore uses only data of treatment episodes between FY91 and FY95.

[7] OSA requires that a client to be interviewed either by his counselor, or by an experienced staff member in the agency.

[8] Of all alcohol abuse episodes in MATS, a half of them (54.5%) receive outpatient treatment for alcohol abuse.

[9] We explicitly exclude patients who are (1) enrolled in Driver's Education and Evaluation Program (DEEP) (2) currently in jail (3) referred to other agencies or programs (e.g. for inpatient service) (4) deceased or moved away. DEEP are programs designed to prevent future offenses caused by drivers with problems of substance abuse. Enrollees in these programs may be required to attend a certain number of visits to recover their licenses.

Table 2
Basic characteristics for clients at admission

	Mean percent	Standard deviation	Median
Age (years)	33.2	10.7	32.2
Education (years)	11.5	2.2	12.0
Income ($ per month)	928.3	875.8	500.0
Male (%)	69.9		
Married (%)	21.7		
Psych problem (%)	13.7		
Legal involvement (%)	40.8		
Employment			
Full-time (%)	28.6		
Part-time (%)	9.8		
Unemployed (%)	61.6		
Reported frequency			
Light (%)	47.7		
Moderate (%)	18.8		
Intermediate (%)	24.1		
Heavy (%)	9.4		
Clinical assessment			
Causal (%)	6.8		
Life-style involved (%)	23.5		
Life-style dependent (%)	47.5		
Dysfunctional (%)	22.2		
Total observations	7615		

for agency fixed effects in the estimation, we only keep those agencies having at least 20 MATS episodes over the sample years. The remaining sample consists of 7615 episodes. Because MATS is collected on an episode basis, there are only 6778 unique clients. In what follows, to simplify discussion, we disregard the distinction between episodes and clients.

Table 2 reports the characteristics of clients included in X_{d}. The first set of characteristics consists of demographics of clients. 69.9% of clients are male and 78.3% are single, although 67.8% of single clients were married before. The average client is around 30 years old, with a high school diploma and a monthly household income of around $900.[10] Among the clients, 28.6% have full-time jobs, 9.8% work part-time and the remaining 61.6% do not work at all, partly explaining the heterogeneity in the clients' monthly household incomes. Around 41% of clients have some legal involvement at admission, either being on probation or parole or waiting for a trial. Finally, 13.7% of clients have recognized psychiatric comorbidity at admission.

The second set of variables in X_{d} are related to a client's initial health. We employ two kinds of measures of a client's health at admission. First, we construct four dummies from the client's reported frequency of alcohol usage at admission, and denote the client as

[10] This is the average monthly income when income is known and stated by the patient and is less than $9999.

Table 3
Payer status and prices of demand and supply

Payer status	Demand price	Supply price	
OSA	Low	Low	21.9%
Medicaid	Low	High	29.4%
Self-Paid	High	Low	18.5%
Third Party	High	High	30.2%

"light" if the client has been abstained from alcohol for over a month, "moderate" if his usage is less than once per week, "intermediate" if the usage is less than once per day, and "heavy" if the usage is even greater. Second, based on MATS's terminology, we construct dummies from the clinician's assessment of the severity of the client's health condition, indicating whether the client is a "causal," "life-style involved," "life-style dependent," or "dysfunctional" user, as assessed by the clinician. The summary statistics of these health variables are displayed in Table 2. In our sample, 18.8% of clients are moderate, 24.1% are intermediate and 9.4% are heavy users. By contrast, 23.5% are life-style involved, 47.5% are life-style dependent, and 22.2% are dysfunctional.

According to the client's primary expected payment source, we construct four different dummies for payers: "OSA," "Medicaid," "Self-Paid," and "Third Party," and present its distribution in Table 3.[11] About 22% of clients are primarily supported by OSA funds (OSA); 29.4% are covered either by Medicaid or Medicare programs (Medicaid); 30.2% have private insurances (Third Party), and 18.5% pay treatment costs out of their own resources (Self-Paid). Many clients who are classified as Self-Paid are also partially supported by OSA.

The exact price paid by clients and recorded by providers are not available in MATS. Nevertheless, the price a client pays and the fee a provider gets for treatment are mainly determined by the client's payer status. We thus use a client's payer status as a proxy for demand and supply prices. The ranking of supply and demand prices is shown in Table 3. Medicaid and OSA clients in general face a lower (zero) demand price. The provider receives a more generous reimbursement when clients are covered by Medicaid or private insurances. Because we only know relative prices, the demand or supply price is coded as 1 when it is "high" and 0 when it is "low".

Using payers as proxies of prices may introduce a selection problem because clients may choose payers according to their initial health. Although half of the clients in the sample cannot choose their payers (Medicaid or OSA), the other half (Third Party or Self-Paid) may have some choice. In addition, the agency may attempt to match clients of worse initial health with generous payers.[12] For instance, an agency may seek to enroll regular clients in Medicaid, and these clients may have larger health needs. To see if there is such a correlation between a client's payer and his health, we display in Table 4 the reported alcohol usage by payers. The severity indexes in Table 4 are almost identical across payers, though Medicaid

[11] MATS reports a client's primary, secondary, and tertiary expected source of treatment payment. We use the primary source of payment to identify a client's payer group.

[12] Clients need to meet certain criteria to be eligible for Medicaid programs. For OSA clients, their incomes also have to be below a certain threshold.

Table 4
Payer status and reported frequency at admission and at discharge

	Light	Moderate	Intermediate	Heavy
Admission				
OSA (%)	48.0	18.6	24.2	9.2
Medicaid (%)	45.9	22.1	23.3	8.7
Self-Paid (%)	52.7	15.9	22.5	8.9
Third Party (%)	42.0	18.5	28.1	11.4
Discharge				
OSA (%)	73.2	12.5	10.7	3.6
Medicaid (%)	68.4	16.9	11.1	3.6
Self-Paid (%)	78.0	9.6	9.0	3.4
Third Party (%)	73.3	12.6	10.8	3.3

Table 5
Payer status and health care use (visits)

	Completer (Q)		Quitter (q)		Total	
	Mean	Median	Mean	Median	Mean	Median
OSA	17.9	12	6.6	4	10.2	6
Medicaid	27.4	18	9.5	5	15.2	7
Self-Paid	17.8	14	7.5	4	12.2	7
Third Party	20.1	15	7.4	4	13.4	8
Total	20.7	14	7.9	4	12.9	7

and Third Party clients have slightly worse severity. This implies that there is only a weak correlation between a client's health and his observed severity.[13]

Following the same usage groups, we measure a client's health outcome according to the reported frequency of alcohol usage at discharge. In MATS, an abstinent client falls in our "light" user category. As seen in Table 4, over 70% of clients are abstinent and less than 4% are still heavy users after treatment. We see at least a 20% increase in light users and a 6% decrease in heavy users, indicating an improvement in reported health.

We measure a client's health care use by the number of attended visits in an episode. Table 5 displays the mean and median of attended visits for complete (q) and incomplete episodes (Q). To show the relation between the health care use and demand or supply prices, we further separate them by the client's payer. The average number of attended visits is 20.7 for complete and 7.9 for incomplete episodes. Clearly, the number of attended visits is much higher for completers. Medicaid and Third Party clients, two groups with a high supply price, have more visits, especially for completed episodes. Also, the sample median is consistently lower than the average for quitters, completers and the entire sample: health care use is positively skewed.

[13] We also check if there is a correlation between a client's clinical assessment and his payer. We find a similar pattern: Medicaid and Third Party clients have health slightly worse than the average.

5. Empirical results

5.1. Basic specification

Our basic specification let a client's demand depend on his personal characteristics (including health at admission) and supply and demand prices.[14] Because a provider knows all the information in the data when setting the quantity limit, all the regressors in X_{d} are also included in X_{s}. Additionally, we include provider dummies in X_{s} allowing the specific practice style of each provider to influence its limit.[15] Finally, the client's health care use is measured by the log number of visits because of skewness in health care use.

5.1.1. General model

Table 6 presents the estimated results of the client's demand (7) and the provider's limit (8) in the general model. Our estimates suggest that aged, educated or working clients demand more health care, as do clients with higher incomes; these clients may be more willing to attend treatment. Clients with legal involvements or severe clinical assessment demand less health care. On the other hand, the provider sets lower limits for male or older clients, but higher limits for those who have psychological problems. In addition, the provider sets higher limits for clients with worse health conditions; the worse is the client's health, the higher the limit.

The coefficients of supply and demand prices in (8) are 0.138 and −0.186, respectively; both are significant. They suggest that a provider sets the limit 13.8% higher for clients having a high supply price, but 18.6% lower for clients having a high demand price. Given that the average number of visits for completers is 20, the estimates suggest that the limit is, relative to OSA clients, 2.7 visits higher for Medicaid clients; 1.0 lower for Third Party; 3.7 lower for Self-Paid. By contrast, the estimated coefficient of demand and supply price in (7) are 0.014 (insignificant) and 0.153 (significant), respectively. The positive coefficient of supply price shows that a provider may employ effort or persuasion.

The last set of variables in Table 6 include the standard deviations of all three random errors, σ_{d}, σ_{s}, and σ_{ε}, the correlation between μ_{d} and μ_{s}, ρ, as well as α in Eq. (10). The random errors μ_{d} and μ_{s} are quite correlated (0.380). The correlation indicates that unobserved characteristics (to us) are important determinants for clients to choose demands, and for providers to set limits. For instance, the provider may set lower limits for client showing little interests in participating treatment. The constant α is estimated to be 0.621 and significant; it implies that if the client's demand is within a few visits of the provider's limit, they are willing to complete the treatment. The standard deviation of ε is 0.778 but insignificant.

Our general model controls for a selection bias in the demand and supply models using the condition specified in (10). It is of interest to compare the estimated results of the general model with those in a model without a selection correction. To do this, we separately estimate (7) and (8) for complete and incomplete episodes, respectively, following the basic specification, and present its results in Table 7. A comparison between Tables 6 and 7 reveals

---

[14] A client's income is measured by the square root of his monthly income.

[15] In other specifications, we allow the provider dummies to influence a client's demand as well (see below).

Table 6
Estimated coefficients of the general model

	A client's demand[a]	A provider's supply[b]
Age (years)	0.010 (0.002)**	−0.003 (0.002)*
Male	−0.130 (0.036)**	−0.151 (0.034)**
Married	−0.143 (0.039)**	−0.052 (0.035)
Education (years)	0.026 (0.008)**	−0.005 (0.007)
No legal involvement	−0.254 (0.035)**	0.042 (0.036)
Full-time	0.140 (0.043)**	−0.085 (0.038)*
Part-time	0.161 (0.054)**	−0.052 (0.049)
Income (square root)	0.003 (0.001)*	−0.002 (0.001)
Psych problem	−0.024 (0.045)	0.111 (0.046)*
Life-style involved	−0.217 (0.043)**	0.092 (0.042)*
Life-style dependent	−0.346 (0.041)**	0.167 (0.047)**
Dysfunctional	−0.360 (0.057)**	0.276 (0.068)**
Moderate user	−0.040 (0.073)	0.313 (0.059)**
Intermediate user	−0.141 (0.074)	0.573 (0.069)**
Heavy user	−0.149 (0.076)	0.514 (0.070)**
Supply price	0.153 (0.034)**	0.138 (0.031)**
Demand price	0.014 (0.035)	−0.186 (0.036)**
Constant	1.712 (0.126)**	
σ_d	1.181	(0.081)**
σ_s	0.885	(0.140)**
ρ	0.380	(0.151)**
α	0.621	(0.052)**
σ_ε	0.778	(0.501)
Likelihood value	−15,442.3	
Observations	7615	

The number in the parenthesis denotes the standard deviation. The constant in the provider limit equation is omitted due to providers' dummies.

[a] Provider dummies: no.

[b] Provider dummies: yes.

* Significant at 5%.

** Significant at 1%.

significant differences in estimated coefficients, especially for variables relating to a client's health at admission. Before controlling for selection, estimates of health-related variables in (8) are smaller and not all significant. After controlling for selection, the estimates are significant and larger. By contrast, estimates of all the health-related variables in (7) are smaller when selection is controlled for. This suggests that there is a downward bias for health related variable estimates in the sample of completed episodes, but a corresponding upward bias in the sample of incomplete ones. Finally, the coefficients for demand and supply prices are quite different as well.

5.1.2. Proposition 1: Absent effort or persuasion?

The goal of this paper is to use Propositions 1–3 to test for the presence of each mechanism. A pure quantity limit (Proposition 1) imposes a joint constraint: $\delta_s = \gamma_d = 0$ on the general model. The values of the likelihood function is −15,442.3 for the unconstrained (general) model and −15,468.4 for constrained (pure quantity) models. Given that the LR value (52.2)

Table 7
Estimated coefficients of the model without controlling for selection

	A client's demand[a]	A provider's supply[b]
Age (years)	0.005 (0.002)**	0.002 (0.002)
Male	−0.106 (0.036)**	−0.173 (0.039)**
Married	−0.090 (0.041)*	−0.072 (0.039)
Education (years)	0.006 (0.008)	0.019 (0.008)*
No legal involvement	−0.110 (0.035)**	−0.052 (0.035)
Full-time	−0.001 (0.044)	0.021 (0.040)
Part-time	0.082 (0.056)	0.022 (0.052)
Income (square root)	0.001 (0.001)	0.001 (0.001)
Psych problem	0.030 (0.044)	0.088 (0.053)
Life-style involved	−0.115 (0.043)**	0.032 (0.042)
Life-style dependent	−0.199 (0.039)**	0.016 (0.042)
Dysfunctional	−0.175 (0.054)**	0.067 (0.069)
Moderate user	0.008 (0.077)	0.259 (0.060)**
Intermediate user	0.050 (0.074)	0.462 (0.060)**
Heavy user	0.037 (0.078)	0.368 (0.065)**
Supply price	0.132 (0.034)**	0.202 (0.035)**
Demand price	−0.088 (0.036)*	−0.134 (0.038)**
Constant	1.447 (0.115)**	
Observations	4622	2993

The number in the parenthesis denotes standard deviation. The constant in the provider limit equation is omitted due to providers' dummies.

[a] Provider dummies: no.

[b] Provider dummies: yes.

* Significant at 5%.

** Significant at 1%.

is well above the critical value at 1% (9.2), we reject the restriction and the hypothesis that a provider does not use effort or persuasion.

5.1.3. Proposition 2: Absent quantity ceiling?

To test Proposition 2, we first compare the likelihood value of the general model with that of the pure effort–persuasion model. The value of the pure effort–persuasion model is −15672.3, yielding an extremely large LR value (460.0). This is not surprising since about two-fifth of clients in the data have complete episodes, while the pure effort–persuasion model predicts all episodes are incomplete. Nevertheless, as noted earlier, the statistical power of this result may depend on how one interprets the separation indicator. It is therefore useful to compare with another test not directly relying on the information of j. Table 8 presents the results that check if the demand equation can fully explain the observed quantity. As quite evident from the Table, the coefficient τ is estimated to be quite different from 0 (the expected value of pure effort–persuasion model) and even significantly larger than 1 (the expected value of the general model). Again, we reject the hypothesis that a provider does not set a quantity limit.

At this point we have evidence that providers use a quantity ceiling, and that they use *either* effort or persuasion. We now employ data on outcomes to differentiate effort from persuasion.

Table 8
Estimated coefficients of the pure effort–persuasion model

	Coefficient	Standard deviation
Age (years)	−0.001	0.001
Male	0.020	0.028
Married	0.011	0.030
Education (years)	−0.006	0.006
No legal involvement	0.053	0.026*
Full-time	−0.031	0.031
Part-time	−0.034	0.040
Income (square root)	−0.001	0.001
Psych problem	0.021	0.035
Life-style involved	0.028	0.032
Life-style dependent	0.056	0.029
Dysfunctional	0.085	0.043
Moderate user	0.005	0.050
Intermediate user	0.028	0.048
Heavy user	0.025	0.052*
Supply price	−0.013	0.025
Demand price	−0.011	0.026
Predicted quantity (general model)	1.080	0.022**
Constant	−0.001	0.001
Observations	7615	

* Significant at 5%.
** Significant at 1%.

5.1.4. *Proposition 3: Effort versus persuasion*

Proposition 3 tests if a client's health outcome is affected by the supply price, a proxy for effort. Given the ordinal health outcome measure, we estimate the health production function by the ordered logit, where better outcome measures are assigned for smaller numbers.[16] Furthermore, we stratify (Eq. 16) according to the client's reported alcohol usages at admission because treatment may work differently for clients with different initial health status.[17] The estimated coefficients of the health production for each usage group are presented in Table 9.

As displayed in Table 9, light users have quite different treatment effects than other users.[18] Attending more visits, for instance, contributes much less to a client's health outcome for light users than other clients. Our coefficient of interest, supply price, is insignificant across all usage groups. Since our health outcome is actually a severity measure (negative health), the effort would imply that the coefficient on the supply price is negative. It is positive in all the models. The findings fail to reject the hypothesis that a provider's action is entirely persuasion.

An endogeneity problem may lead to a bias in the estimate of the supply price coefficient. Clients with worse health status are likely to choose more generous payers. Although, the

[16] Specifically, the health outcome is denoted as 4 if a client is a heavy user at discharge; 3 for intermediate user; 2 for moderate user, and 1 for light user.
[17] Lu and McGuire (2002) found that the treatment effects may depend on a client's severity at admission.
[18] Our results are consistent with the findings of Lu and McGuire (2002).

Table 9
Estimation of health outcomes with clinical assessment

	Light	Moderate	Intermediate	Heavy
Age (years)	0.001 (0.007)	−0.026 (0.007)**	0 (0.005)	−0.007 (0.008)
Male	−0.327 (0.166)*	−0.118 (0.132)	−0.186 (0.118)	0.041 (0.198)
Married	0.136 (0.182)	0.146 (0.151)	−0.056 (0.132)	−0.145 (0.207)
Education (years)	−0.13 (0.034)**	0.038 (0.031)	−0.034 (0.024)	−0.032 (0.037)
No legal involvement	−0.124 (0.161)	0.357 (0.128)**	0.353 (0.117)**	0.339 (0.192)
Full-time	−0.215 (0.202)	0.078 (0.154)	−0.176 (0.131)	−0.593 (0.219)**
Part-time	−0.157 (0.262)	0.247 (0.195)	−0.04 (0.181)	0.272 (0.335)
Income (square root)	−0.012 (0.006)*	−0.004 (0.004)	−0.003 (0.004)	−0.003 (0.005)
Life-style involved	0.365 (0.194)	0.499 (0.165)**	0.654 (0.149)**	0.066 (0.236)
Life-style dependent	0.168 (0.343)	0.201 (0.193)	−0.376 (0.302)	−1.071 (1.315)
Dysfunctional	0.475 (0.320)	0.427 (0.194)*	−0.409 (0.293)	−0.756 (1.281)
Psych problem	0.256 (0.341)	0.473 (0.217)*	−0.147 (0.307)	−0.628 (1.284)
Total visits (log)	−0.028 (0.063)	−0.722 (0.055)**	−1.054 (0.053)**	−1.075 (0.085)**
Supply price	0.232 (0.156)	0.208 (0.125)	0.135 (0.109)	0.3 (0.176)
Observations	3629	1436	1831	691

The number in the parenthesis denotes standard deviation. The health outcome is denoted as 4 if a client is a heavy user at discharge; 3 for intermediate user; 2 for moderate user; and 1 for light user. The cut-off values in the ordered logit are omitted.

* Significant at 5%.

** Significant at 1%.

coefficients for the supply price are quite similar across usage groups (ranging from 0.13 to 0.30), the supply price may still capture some unobserved health conditions of clients. To check if our results are robust to unobserved client health conditions, we re-estimate (Eq. 16) without client's clinical assessment. If clients choose their payers based on their health status, the coefficient of supply price should be sensitive to the exclusion of these variables (notice that these variables are all significant in Table 6). As Table 10 shows, the estimated coefficient of the supply price remains almost unchanged in all usage groups; our results are insensitive to this observed health measure.

5.2. *Robustness checks*

5.2.1. *Other specifications*

In this subsection, we present a number of robustness checks by varying elements in X_d, the sample composition, and variables in X_s. We first check if our results are robust to elements of X_s. In the basic specification (Table 8 or column (1) of Table 11), X_s is identical to X_d except for provider dummies. The financial conditions for a provider may also affect its quantity limit.[19] Because a provider's income source is mainly determined by the payer mix of its clients, we construct four dummies from income sources: "OSA Above," "Medicaid Above," "Third Party Above," and "Self-Paid Above," to represent financial condition. The dummy "OSA Above" is set at 1 if the proportion of OSA clients

[19] Lu and McGuire (2002) found that a provider's financial condition had a large impact on the client's health care use. Our dummies are constructed in the same way as in their paper.

Table 10
Estimation of health outcomes without clinical assessment

	Light	Moderate	Intermediate	Heavy
Age (years)	0.003 (0.007)	−0.022 (0.006)**	−0.001 (0.005)	−0.006 (0.008)
Male	−0.324 (0.166)	−0.115 (0.132)	−0.164 (0.117)	0.03 (0.197)
Married	0.151 (0.182)	0.154 (0.151)	−0.055 (0.132)	−0.173 (0.206)
Education (years)	−0.127 (0.033)**	0.039 (0.031)	−0.033 (0.024)	−0.036 (0.036)
No legal involvement	−0.105 (0.160)	0.338 (0.127)**	0.355 (0.117)**	0.33 (0.191)
Full-time	−0.211 (0.202)	0.061 (0.153)	−0.194 (0.130)	−0.625 (0.218)**
Part-time	−0.157 (0.262)	0.265 (0.195)	−0.035 (0.181)	0.263 (0.334)
Income (square root)	−0.013 (0.006)*	−0.005 (0.004)	−0.003 (0.004)	−0.003 (0.005)
Psych problem	0.382 (0.193)*	0.546 (0.163)**	0.66 (0.148)**	0.058 (0.235)
Total visits (log)	−0.026 (0.063)	−0.726 (0.055)**	−1.05 (0.053)**	−1.068 (0.085)**
Supply price	0.246 (0.156)	0.204 (0.125)	0.146 (0.109)	0.298 (0.175)
Observations	3629	1436	1831	691

The number in the parenthesis denotes standard deviation. The health outcome is denoted as 4 if a client is a heavy user at discharge; 3 for intermediate user; 2 for moderate user; and 1 for light user. The cut-off values in the ordered logit are omitted.

* Significant at 5%.

** Significant at 1%.

for a provider in a fiscal year is higher than the sample average; likewise for other dummies. In column (2) of Table 11, we estimate the model by adding dummies of agency income sources in X_s.

Next, in columns (3)–(5), we check if our results are sensitive to different sample compositions. For each column, we re-estimate the model under the specification of Column (2), with a more restricted sample. Clients who are legally involved may behave differently than those who are not. In column (3), we use only clients that are not legally involved at admission. As noted before, some clients in the sample have more than one episode. In column (4), we include only their first episodes.

In column (5), we estimate the model by restricting the sample to episodes with more than three attended visits. This sample restriction could be important for two reasons. First, for very short episodes, clients may "experiment" rather than "participate in" the treatment program. Such a restriction enables us to check whether our result is driven by short episodes. Second, our estimation relies on the assumption that errors are normally distributed. By excluding a number of short episodes (most of short episodes are incomplete), we can check if our results are sensitive to distributional assumptions.

Finally, we allow the practice style of each provider, captured by dummies of provider or agency income sources, to influence not only a provider's limit, but also the level of effort or persuasion. We re-estimate each specification (columns (1)–(5)) by adding provider dummies or agency income dummies to X_d. The estimated results are presented in columns (6)–(10) of Table 11.

5.2.2. *Propositions 1–3*

Table 11 presents the likelihood function values of the general model, as well as that for the pure effort or persuasion and for pure quantity models. Our results strongly reject the pure quantity ceiling model in every case, supporting the presence of quantity limits.

					involved		three visits			involved	episode	three visits
Specification	Client's demand (q)	Income dummies							x	x	x	x
		Provider dummies						x	x	x	x	x
	Provider's limit (Q)	Income dummies		x	x	x	x		x	x	x	x
		Provider dummies	x	x	x	x	x	x	x	x	x	x
Likelihood value		General	−15,442.3	−15,439.4	−8,920.2	−13,024.8	−9,298.9	−15,058.3	−15,036.2	−8,753.9	−12,674.5	−9,038.9
		Effort/persuasion	−15,468.4	−15,465.0	−8,939.4	−13,049.8	−9,301.6	−15,115.5	−15,098.0	−8,798.3	−12,729.5	−9,040.7
		Quantity setting	−15,672.3	−15,672.3	−9,068.4	−13,247.0	−9,526.6	−15,485.5	−15,468.9	−8,985.3	−13,077.0	−9,404.0
LR value		Proposition 1	52.2**	51.2**	38.4**	50.0**	5.4	114.4*	123.6**	88.8**	110.0**	3.4
		Proposition 2	460.0**	465.8**	296.4**	444.4**	456.6**	854.4**	864.6**	463.2**	805.0**	730.2**
t value		Proposition 2	49.14**	49.24**	48.28**	49.20**	44.63**	43.87**	43.53**	45.52**	42.75**	45.10**
		Proposition 3 (light)	1.48	1.48	0.95	1.67	1.10	1.90	1.84	1.16	1.96	1.62
		Proposition 3 (moderate)	1.66	1.66	1.59	1.23	1.66	0.85	0.88	0.73	0.86	0.93
		Proposition 3 (intermediate)	1.24	1.24	0.82	1.59	1.84	0.54	0.34	0.29	0.93	1.09
		Proposition 3 (heavy)	1.70	1.70	1.88	1.48	0.86	1.16	1.40	2.04*	1.29	0.74
Total observations			7,615	7,615	4,356	6,778	5,372	7,615	7,615	4,356	6,778	5,372

* Significant at 5%.
** Significant at 1%.

The results also reject the pure effort or persuasion model in almost every case, suggesting the presence of effort or persuasion. Lastly, the *t* value of the supply price is insignificant in almost all cases. Where it is not, the coefficient is positive and significant, and this is consistent with the presence of persuasion but not effort in the provider's action.

6. Conclusions

Health care use is the result of an interaction between a client's health demand and a provider's health supply. To affect the utilization, a provider can either change his own supply, or influence a client's demand for health care. Three types of provider–patient interactions are discussed in the paper: rationing sets a quantity limit; effort and persuasion influence a client's health demand by productive (but noncontractible) and unproductive inputs, respectively. Each mechanism is distinct and has different implications.

While previous research has emphasized the distinctions among these mechanisms in theory, there has been very little work that studies them empirically.[20] Instead, most empirical studies regard persuasion as the only available mechanism for providers, and interpret all the utilization change affected by supply factors as PID. In fact, utilization change may be affected by other mechanisms such as provider rationing and effort; each of these has welfare implications different from persuasion. For instance, while persuasion is not productive, effort is.

The distinction is particularly important in markets with competing managed care plans, where quantity limits and persuasion have very different implications for normative interpretation of a quantity effect of managed care. If plans ration by quantity limits, consumers' demand remains valid, and their choices among plans could be regarded as being in their best interest.[21] By contrast, to the extent that consumers' demand is manipulated by providers, consumers accede to the rationing, persuaded by the provider that quantities are in their best interest. Under this mechanism, consumers' choices among plans will not reliably signal their true self-interest.

In this paper, we construct a theoretical model that incorporates all three interactions as special cases. When the general model is specialized, a set of empirical implications emerges. We employ these implications to test for the presence of each mechanism, using data on substance abuse treatment in the Maine Addiction Treatment System. We find evidence for rationing, confirming the view that health providers ration services to prevent high use of some clients. The presence of persuasion is also quite evident; it says that PID exists, even when accounting for mechanisms of effort and rationing. We do not find evidence supporting the presence of effort.

We now discuss some limits of our analysis. First, as we said earlier, using payers as proxies of prices of supply and demand may lead to some bias, since clients with high

[20] See, however, Grytten et al. (2001) who distinguish an availability effect (rationing) from inducement effect (persuasion). Their work uses incomes of physicians who practice under various levels of competitions (high or low physician density) to separate these two effects. They do not consider effort as a mechanism affecting utilization.

[21] The choices do not necessarily lead to an efficient outcome in the market due to adverse selection and other potential market failures.

severity or treatment values are inclined to choose generous insurances. Although the sample shows a weak correlation between payer and severity, it is still possible that clients who benefit from treatment more are in plans with high supply prices. In that case, the effects of rationing and persuasion are likely to be overestimated.

Second, we use a static model of provider–client interaction. We have assumed that the quantities demanded by clients and supplied by providers are determined at the start of an episode. In practice, a provider's limit and a client's demand may be determined sequentially. It is likely that providers and clients modify their decisions using the information gathered during a course of treatment. A richer, sequential model could extend and refine our results.

The main contribution of this paper is to develop a framework for distinguishing mechanisms of health care provider–client interactions. While our findings may be specific to substance abuse treatment, our methods can be generalized to other health services. We expect other conclusions to be drawn, based on the features of each health service. The majority of health data sets do not contain information about a client's completion status, a key variable in the estimation. Our paper has demonstrated the importance of this key variable, particularly when there is concern about how providers might take actions to alter health care use.

Acknowledgements

This research received support from NIDA-P50-DA10233 and AHRQ-PO1-HS10803. We thank one anonymous referee and one editor for their help on the paper and Victor Aguirregabiria, Chungrong Ai, Richard Frank, Kevin Lang, Mingshan Lu, and Joseph Newhouse for comments on earlier versions of this paper. We also are grateful to the Maine Office of Substance Abuse for cooperation in providing the data for the study. Any remaining errors are ours.

References

Ackerberg, D., Machado, M.P., Riordan, M.H., 2001. Measuring the Relative Performance of Providers of a Health Service. NBER Working Paper No. W8385.

Cromwell, J., Mitchell, J.B., 1986. Physician-induced demand for surgery. Journal of Health Economics 5, 293–313.

Cuyler, A.J., 1989. The normative economics of health care finance and provision. Oxford Review of Economic Policy 5, 34–58.

De Jaegher, K., Jegers, M., 2000. A model of physician behavior with demand inducement. Journal of Health Economics 19, 231–258.

Dranove, D., 1988. Demand inducement and the physician/patient relationship. Economic Inquiry 26, 251–298.

Eisenberg, J.M., 1986. Doctors' Decisions and the Cost of Medical Care. Health Administration Press, Ann Arbor, MI.

Evans, R.C., 1974. Supplier-induced demand: some empirical evidence and implications. In: Perlmann, M. (Ed.), The Economics of Health and Medical Care. Macmillan, London, pp. 162–173.

Farley, P.J., 1986. Theories of the price and quantity of physician services. Journal of Health Economics 5, 315–333.

Fuchs, V.R., 1978. The supply of surgeons and the demand for operations. Journal of Human Resources 13, 32–56.

Gaynor, M., 1994. Issues in the industrial organization of the market for physician services. The Journal of Economics and Management Strategy 39 (1), 211–255.

Gaynor, M., Gertler, P., 1995. Moral hazard and risk spreading in partnerships. RAND Journal of Economics 26, 591–614.

Grytten, J., et al., 2001. The income effect and supplier induced demand. Evidence from primary physician services in Norway. Applied Economics 33 (11), 1455–1467.

Gruber, J., Owings, M., 1996. Physician financial incentives and cesarean section delivery. RAND Journal of Economics 27, 99–123.

Hillman, A., et al., 1992. Physicians' utilization and charges for outpatient diagnostic imaging in a medicare population. Journal of the American Medical Association 268, 2050–2054.

Kessler, D.P., McClellan, M., 1996. Do doctors practice defensive medicine? Quarterly Journal of Economics 111, 353–390.

Lu, M., McGuire, T.G., 2002. The productivity of outpatient treatment for substance abuse. Journal of Human Resources 38 (2), 309–335.

Ma, C.A., McGuire, T.G., 1997. Optimal health insurance and provider payment. American Economic Review 87 (4), 685–704.

Maddala, G.S., 1983. Limited-Dependent and Qualitative Variables in Econometrics. Economic Society Monographs in Quantitative Economics.

Machado, M.P., 2001. Dollars and performance: treating alcohol misuse in Maine. Journal of Health Economics 20 (4), 639–667.

McGuire, T.G., Pauly, M.V., 1991. Physician response to fee changes with multiple payers. Journal of Health Economics 4 (10), 385–410.

McGuire, T.G., 2000. Physician agency. In: Culyer, A.J., Newhouse, J.P. (Eds.), Handbook of Health Economics. North Holland.

Pauly, M.V., 1980. Doctors and Their Workshops: Economic Models of Physician Behavior. University of Chicago Press.

Rice, T., 1983. The impact of changing medicare reimbursement rates on physician-induced demand. Medical Care 21, 803–815.

Rizzo, J.A., Blumenthal, D., 1994. Physician labor supply: do income effects matter. Journal of Health Economics 13, 433–453.

Shen, Y., 1999. Selection Incentives in a Performance-Based Contracting System. Unpublished dissertation, Boston University.

Stano, M., 1987. A further analysis of the physician inducement controversy. Journal of Health Economics 6, 229–238.

Williams, A., 1998. Medicine, economics, ethics, and the NHS: a clash of cultures? Health Economics 7, 565–568.

Yip, W., 1998. Physician responses to medical fee reductions: changes in the volume and intensity of supply of coronary, artery bypass graft (CABG) surgeries in the medicare and private sectors. Journal of Health Economics 6 (17), 675–699.