A model of physician behaviour with demand inducement

Kris De Jaegher *, Marc Jegers

Vakgroep Micro-economie van de Profit- en Non-Profit-Sector, Vrije Universiteit Brussel, Pleinlaan 2, B-1050 Brussel, Belgium

Received 31 October 1997; received in revised form 24 February 1999; accepted 28 July 1999

Abstract

We present a model of the physician–patient relationship extending on the model by Farley [Farley, P.J., 1986. Theories of the price and quantity of physician services. Journal of Health Economics 5, 315–333] of supplier-induced demand (SID). First, we make a case for the way this model specifies professional ethics, physician competition, and SID itself. Second, we derive predictions from this model, and confront them with the neoclassical model. Finally, we stress the importance of considering how SID affects patient welfare in providing an example where physicians’ ability to induce makes patients better off. To evaluate patient welfare, we derive approximations of the patients’ welfare loss due to physician market power in both the neoclassical model and the inducement model. © 2000 Elsevier Science B.V. All rights reserved.

JEL classification: I11
Keywords: Supplier-induced demand; Utility and profit maximization; Monopolistic competition; Welfare economics

1. Introduction

Patients generally do not know which health services they want, and this is why they consult physicians. But because they do not know which health services they

* Corresponding author. Tel.: +32-2-629-2214; fax: +32-2-629-2282; e-mail: kris.de.jaegher@vub.ac.be

0167-6296/00/$ - see front matter © 2000 Elsevier Science B.V. All rights reserved.
PII: S0167-6296(99)00029-6
Table 1
Profits, reduction in consumer surplus, and deadweight loss as a proportion of revenue

<table>
<thead>
<tr>
<th></th>
<th>Profits</th>
<th>Loss in CS</th>
<th>Deadweight Loss</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neoclassical</td>
<td>$</td>
<td>\sigma_{o}</td>
<td>^{-1}$</td>
</tr>
<tr>
<td>Neoclassical + altruism</td>
<td>$</td>
<td>\sigma_{o}</td>
<td>^{-1}$</td>
</tr>
<tr>
<td>Inducement</td>
<td>$</td>
<td>\sigma_{o}</td>
<td>^{-1}$</td>
</tr>
<tr>
<td>Inducement + altruism</td>
<td>$</td>
<td>\sigma_{o}</td>
<td>^{-1}$</td>
</tr>
</tbody>
</table>

want, upon treatment, patients cannot know either to which extent health services contributed to their health. Indeed, it is possible that patients would have found themselves cured with a lower level of health services, or in case their disease is self-curing, with no health services at all. Therefore, as long as the price physicians receive for health services exceeds marginal cost, they may prescribe levels of care to the right of the demand curve according to which perfectly informed patients would behave. This is the supplier-induced demand (SID) hypothesis (Evans, 1974).¹

Testing for absolute SID involves the impossible task of observing a perfectly informed patient (Mooney, 1994). Therefore, it has been marginal SID, i.e., inducement upon entry or upon a change in fee, which has been tested for,² involving the additional assumption that physicians induce more as their income gets under pressure. However, a recurring problem with this version of the SID hypothesis has been its failure to make predictions that distinguish it from neoclassical theory. The first focus of this paper is to try and derive such distinguishing predictions. For this purpose, we base ourselves on and we make a case for the model of SID by Farley (1986). But as argued by Labelle et al. (1994), what ultimately matters is not so much whether or not physicians induce demand, but how this affects patient welfare. And therefore, the second focus of this paper is the welfare implications of SID.

Section 2 provides a justification for the model we employ, which is based on the model of Farley (1986). In Section 2.1 we make a case for an intuitive form of professional ethics, and show how the form of these professional ethics determine which form SID takes. In Section 2.2, we show how competition could constrain physicians in their ability to induce demand. Before deriving predictions from the model of Farley (1986), we first treat the neoclassical model, and the neoclassical model extended with altruism (Section 3). This is for expositional reasons, as it allows us to stepwise increase complexity. It also allows us to confront the predictions and the welfare implications of the neoclassical model with those of

¹ The same problem is addressed in the literature of credence goods, a category of goods for which both search and experience by consumers are expensive (Darby and Karni, 1973).
² The terms marginal and absolute SID are due to Birch (1988).
Table 2

| Model                        | $|\xi_{QA}|$ | $|\xi_{PA}|$ | $|\xi_{AQ}|$ | $|\xi_{AQ}|$ | $\xi_{PA}$ | Effect of entry on $\phi$
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Neoclassical</td>
<td>&gt; 1</td>
<td>&gt; 0</td>
<td>&gt; 0</td>
<td>&gt; 0</td>
<td>&gt; 0</td>
<td>no increase</td>
</tr>
<tr>
<td>Neoclassical + altruism</td>
<td>&gt; 0</td>
<td>&gt; 0</td>
<td>&gt; 0</td>
<td>&gt; 0</td>
<td>&gt; 0</td>
<td>ambiguous</td>
</tr>
<tr>
<td>SID</td>
<td>&gt; 2</td>
<td>&gt; 1</td>
<td>&gt; 1</td>
<td>= -1</td>
<td>= -1</td>
<td>no increase</td>
</tr>
<tr>
<td>SID + altruism</td>
<td>&gt; 1</td>
<td>&gt; 0</td>
<td>&gt; 1</td>
<td>&lt;</td>
<td>$\in [-1, 0]$</td>
<td>ambiguous</td>
</tr>
</tbody>
</table>

the inducement model. Section 4 then treats the welfare implications (Section 4.1) and the predictions (Section 4.2) of the SID model we have made a case for, and compares these results to the ones of the neoclassical model in Tables 1 and 2. Section 5 lists several caveats to our results. We end with some conclusions in Section 6.

2. Justification for the model used

2.1. Altruism and demand-setting

Most SID models resemble simple models of persuasive advertising (Stano, 1987). Both the advertising model and the inducement model assume that firms shift demand, but are constrained by the costs involved with demand-shifting. In the advertising model, the costs of shifting demand are pecuniary, whereas in the SID model they are non-pecuniary, moral costs. We now argue that treating inducement as a variant of advertising leads to an unintuitive model of SID, and we make a case for the model of Farley (1986), which is based on demand-setting and altruism.

Let us first look at the simple model of persuasive advertising, which describes the following maximization problem for the representative firm (Comanor and Wilson, 1974): $^3$

$$\max_{\phi, A} pQ(\phi, A) - C(Q(\phi, A), A)$$

where the firm sets both price $\phi$ and advertising level $A$. $Q$ is function of both $\phi$ and $A$, with $Q_{A} > 0$ and $Q_{\phi} < 0$, and where the second-order conditions are assumed to be met. Advertising persuades consumers that they want to buy more, resulting in Fig. 1 in a shift in demand from $D_1$ to $D_2$. The extent to which the firm can shift demand is constrained because advertising is costly ($C_A > 0$). The effect of advertising on consumer welfare cannot be evaluated, as advertising involves a change in consumer preferences (Schmalensee, 1972).

$^3$ Appendix C lists the symbols used throughout the paper.
Let us now look at the SID model. In its extreme form, it assumes that there is no limit to the physician’s ability to induce demand. But this leads to the absurd result that as long as the price of health services exceeds marginal cost, physicians will continue to induce demand until the patients’ full income is extracted (Wilensky and Rossiter, 1981). The model is then rescued by assuming that physicians do not exploit their full ability to induce demand because they are constrained by their professional ethics (Reinhardt, 1985). To formally model SID, the advertising model now seems readily available, and one can simply again take maximization problem (1), with $A$ inducement instead of advertising, and where costs as a function of $A$ should now be interpreted as moral costs expressed in pecuniary terms.

However, the advertising analogy masks fundamental differences between the inducing physician and the advertising firm. The advertising firm, if it does not advertise, faces demand $D_1$ in Fig. 1. By advertising, it persuades consumers to buy more at each price, and it thereby shifts demand to $D_2$. The physician on the contrary does not face demand $D_1$, indeed does not face any demand whatsoever. This is because as stated in Section 1, patients do not know what they want, and determining demand is exactly what the physician is supposed to do (Labelle et al., 1994). Therefore, the only rationale for continuing to model SID as demand-shifting is modelling professional ethics such that the physician faces zero moral costs as long as she \footnote{Patients are male in this paper, physicians female.} sets price–quantity combinations along $D_1$, the demand she would face if the patient would be perfectly informed, but faces higher moral costs as she shifts demand further away from $D_1$.

As we now argue, this is an unintuitive way of modelling professional ethics. To show this, we will evaluate consumer welfare, which unlike in the advertising
Consider in Fig. 1 point \( b \) on \( D_1 \) and point \( f \) on \( D_2 \). The model of professional ethics described in the previous paragraph says that at point \( b \), moral costs are zero, whereas at point \( f \), they are non-zero. However, interpreting Fig. 1 as referring to a representative patient, at point \( b \) consumer surplus \((abc)\) is smaller than the consumer surplus at point \( f \) \((aed-efg)\). The reason of this of course is that in this example, we have let the shift in demand from \( D_1 \) to \( D_2 \) coincide with a decrease in price. Except for quantity, to reflect the physician’s moral costs of shifting demand, one may therefore include price as an argument in the physician’s utility from professional ethics.\(^5\) But this brings us very close to the way Farley models professional ethics, namely, as utility interdependence:

\[
w(Y,U)
\]

where the physician’s utility \( w \) depends on the representative patient’s utility \( U \), besides on own income \( Y \). As altruism is defined in economics as one agent’s utility depending on another agent’s utility (Becker, 1981, p. 173), we will use the term altruism instead of professional ethics in the rest of the paper. As well, we will consider inducement as demand-setting, rather than as demand-shifting, as setting demand for a certain level of patient utility is what the physician is now in fact doing.

But only modelling the physician as valuing the utility of the representative patient, as in Farley (1986), is incomplete. Indeed, a physician valuing only the utility of the representative patient would prefer to treat one patient only, and to provide this one patient with a lot of utility, as this would allow her to produce high levels of representative utility without losing much profits.\(^6\) Again, this is not a very intuitive model of altruism. An alternative modelling strategy is then to let each individual patient’s utility be an argument in the physician’s utility function. But in this case, it is not clear how the utility of a new patient enters into the physician’s utility function, as a new patient is also a new good, and a utility function with an extra good is a different utility function from the original one. \( w_1 \) and \( w_2 \) in Eq. (3) are two different utility functions, where the first, besides income, has \( P \) utility arguments, and the second \( P + 1 \) utility arguments:

\[
\begin{align*}
& w_1(Y,U_1, \ldots, U_P), \\
& w_2(Y,U_1, \ldots, U_P, U_{P+1}).
\end{align*}
\]

As pointed out by Lancaster (1966), this problem of how to deal with new goods is a general one in economic theory. One solution to this problem is to include all possible goods (all patients in the market) in the physician’s utility

\(^5\) This was suggested to us by one of the referees.

\(^6\) We are indebted for this point to Bruno Heyndels.
function. But this means stretching the physician’s power of imagination very far (Lancaster, 1966). Additionally, with such a specification altruism could break down, due to the individual physician’s free-riding incentives of letting other physicians take care of patients, while obtaining the same non-pecuniary benefits as when she would take care of these patients herself (Sugden, 1982). We therefore suggest that the physician’s utility, besides on the average utility of her current patients, also depends on the number of patients \( P \) in her clientele.\(^7\) Indeed, as we will assume that all patients are identical (see Section 2.2), the utility of one patient and the number of patients are the only goods that could produce the characteristic “altruism” for the physician.\(^8\)

\[
W(Y, U, P)
\]

A similar model of altruism is found in the models of Becker and Lewis (1973) and Becker and Tomes (1976) of the demand for children, and in the model of Pauly (1973) of fiscal federalism. Intuitively, altruism is a function of the quality of helping (utility of the representative patient), and of the quantity of helping (the number of patients who are being helped).

2.2. Competitive constraint

In assuming that patients have no information whatsoever, SID models as they arose in the 1970s can be seen as a reaction against models which treated patients as perfectly informed, sovereign consumers (Van Doorslaer, 1987). Since the 1980s, however, several authors (Pauly, 1980; Dranove, 1988; Rochaix, 1989) have described compromise models, in which physicians induce demand, but are constrained in doing so by the patient’s information. Additionally, models that have direct relevance for the SID hypothesis are found in the literature of credence goods (see \(^1\)) and expert services (e.g., Pitchik and Schotter, 1987; Wolinsky, 1993; Taylor, 1995). These models assume that patients are Bayesians. Physicians are then constrained in inducing demand, in that in response to certain diagnoses, patients may refuse treatment (Pitchik and Schotter, 1987), or look for a second opinion (Wolinsky, 1993; Rochaix, 1989). Additionally, if patients have imperfect information of their true health status, they may reject the diagnosis if it contradicts their information (Dranove, 1988), or may not initiate demand if their information tells them that their health status is not below a certain level (Taylor, 1995).

However, as Shaked (1995) points out, models that assume Bayesian patients fail to explain how patients acquire their information. Indeed, in these models, patients are assumed to be as good epidemiologists as physicians, whereas epidemiological information would seem to be exactly what patients lack. Again,

---

\(^7\) This corresponds to Rochaix (1989) (p. 65), who does again not include average utility.

\(^8\) For heterogeneous patients, one could include some measure of inequality, e.g., the variance in utility.
as formulated in Section 1, given that patients buy treatments, they have no way of finding out how these treatments contributed to their utility. Physician competition now provides a rationale for how patients may acquire information. Though it is expensive for patients to evaluate the extent to which health services contribute to their utility (the marginal utility of health services), they can experience the level of utility they end up with upon treatment (absolute utility). If the patients of one physician in the market on average obtain higher utility than other patients in the market, then some of these other patients may find out about this through word-by-mouth advertising, and think it unlikely that this higher utility is due to this physician having a less severe caseload. There are now two ways in which patients may use these observed differences in utility. First, they may use them to derive epidemiological information, and directly interfere in the treatment course by evaluating physician’s diagnoses, as described in the previous paragraph. Second, patients may not interfere in the treatment itself, but evaluate physicians upon treatment, and then decide to stop consulting the physicians who provide less utility, and start consulting the physician who provides more utility. As the individual physician’s number of patients now depends on the utility she provides to her patients, as long as price exceeds marginal cost, she has an incentive to attract more patients by increasing the utility she provides. This is how Farley (1986) models physicians as being constrained by patient information in inducing demand. We concentrate on Farley’s model, because, even though containing a competitive constraint, it fits the SID hypothesis’ intuition that physicians determine the content of treatment.

For simplicity, we follow Farley in treating a symmetric model where it is assumed that the market consists of \( Z \) perfectly identical patients and \( N \) perfectly identical physicians. The number of patients \( P \) of one of the identical physicians in the market, physician \( i \), is a function of the utility \( U_i \) provided to one of her identical patients, and of a vector of utilities \( \overline{U}_{j \neq i} \) provided by the other physicians in the market.

\[
P(U_i, \overline{U}_{j \neq i})
\]

with \( P_{U_i} > 0, P_{\overline{U}} < 0 \). The assumption that perfectly identical patients do not all respond in the same way to an increase in the utility provided by one of the physicians deserves some explanation. Identical patients can have different information about the utilities provided by individual physicians, because they acquire their information randomly. In particular, one may assume that patients pass on information to each other, but randomly meet each other (Pauly and Satterthwaite, 1981). Therefore, if one physician provides a higher level of utility than all other physicians in the market, then some patients will find out about this, and others will not.

It is additionally assumed that physicians set the utility they provide to their patients considering the other physicians’ decisions as given. This sets out a model of monopolistic competition based on imperfect information (cf., Stiglitz, 1989, p.
The way competition is modelled is similar to Bertrand–Nash competition, with the difference that physicians set utilities instead of prices. Because we have assumed perfectly identical physicians, as long as all of the $N$ physicians provide the same utility, they will each get a share $Z/N$ of the fixed number of identical patients $Z$ in the market. Such outcomes where all physicians provide the same utility are described by the $DD'$ curve in Fig. 2, where it is additionally assumed that all patients in the market consult a physician as soon as some reservation utility $RU$ is provided. We draw the dependent variable on the $X$-axis because of the analogy with demand curves.

Let us now start from a situation where each physician provides $RU$. If this is optimal to her given that $RU$ is provided by all other physicians, the representative physician will raise the utility she provides above $RU$. The number of patients she can attract for a fixed level of utility provided by other physicians is described by $dd'$. But because of the assumption of identical physicians, all physicians will always behave in the same way, and the outcome must always lie on the $DD'$ curve. A Nash equilibrium $EU$ is reached where the $dd'$ curve and the $DD'$ curve cross. In this Nash equilibrium, the physicians’ conjecture that other physicians will not change their utility in response to the utility they set is met. We cannot formulate the conditions for the existence of such an equilibrium without describing the model completely (see Section 4), so we simply assume at this point that there exists such an equilibrium.

3. Neoclassical model

Before describing Farley’s model completely, and deriving additional results from it, for expositional reasons, we first describe the neoclassical model of a
physician firm with market power, and then add altruism to this model. This will allow us to compare the welfare implications and predictions of the model of Farley (1986) to those of the models treated in the current section.

The maximization problem of the representative neoclassical physician firm with market power is just the same as Eq. (1), but without demand-shifting $A$:

$$\max_{\phi} \phi Q(\phi) - C(Q(\phi)).$$

(6)

The solution of maximization problem (6) yields the Lerner index of market power

$$\frac{\phi - C_{Q}}{\phi} = \frac{1}{|\varepsilon_{Q_s}|}$$

(7)

with $C_{Q}$ marginal cost, $|\varepsilon_{Q_s}|$ the absolute value of the elasticity of the total demand facing the physician.

Let us first look at the welfare implications of this model. Though this neglects the fact that marginal costs are probably increasing, $\phi Q/|\varepsilon_{Q_s}|$ (derived using Eq. (7)) could be argued to be a good approximation of the firm’s profits (abed in Fig. 3). Because market power involves deadweight loss (as marginal utility $= \phi > C_{Q}$), the loss in consumer surplus to the patient because of the absence of perfect competition will be larger than the gain in profits to the physician. The model does not allow to derive an approximation of this loss in consumer surplus, nor of the deadweight loss. The best one can do is to provide an upper bound for these losses. In particular, one can assume that the demand facing the physician is linear, allowing one to interpret it as the demand function of a representative consumer (Varian, 1992, p. 153). The upper bounds on the losses are now found by treating the most extreme case of market power, namely, the one of monopoly, and can be derived algebraically, or simply graphically. As is clear from the linear demand function in Fig. 3, the upper bound on the loss in consumer surplus is $3/2$ of the profits (abce), i.e., $(3/2)(\phi Q/|\varepsilon_{Q_s}|)$, and the upper bound on the dead-

---

**Fig. 3. Monopolistic firm.**
weight loss is \((1/2)(\phi Q/|\epsilon_{\phi}|)(bcd)\), or half of the profits (cf., Cowling and Mueller, 1978). These results are summarized in the first row of Table 1.

Let us now look at some empirical implications of the neoclassical model. Rewriting Eq. (7), we find

\[
\phi \left(1 - \frac{1}{|\epsilon_{\phi}|} \right) = C_Q. \tag{8}
\]

It can be seen from Eq. (8) that \(|\epsilon_{\phi}|\) has to be larger than one in this model — otherwise, the right-hand side in Eq. (8) would be negative. As well, firm entry is conventionally assumed to increase elasticity. From this it follows that price will increase upon entry. Finally, as we will be comparing the neoclassical model to a model where patients decide whether or not to consult a physician, and physicians then decide on how much each patient consume, we rewrite maximization problem (6) in terms of \(Q = P\delta\), with \(P\) the number of patients consulting the physician, and \(\delta\) the number of services bought by each patient. For simplicity, we will assume that costs \(C\) depend on \(P\) multiplied by \(\delta\), i.e., as long as \(Q\) is constant, costs are the same no matter if the physician treats few patients intensively, or many patients extensively.

We now rewrite maximization problem (6) in terms of \(P\) and \(\delta\):

\[
\max_{\phi} \phi \delta(\phi) P(\phi) - C(\delta(\phi) P(\phi)). \tag{9}
\]

It is clear that \(|\epsilon_{\phi}| = |\epsilon_{P\delta}| + |\epsilon_{\delta\delta}|\). However, the neoclassical model does not imply anything about the magnitude of these constituent elasticities, only that their sum is larger than one. In view of the inducement model we treat in Section 4, it also turns out important to look at the expression \(\epsilon_{P\delta}\). This expression is not at fixed prices, as in the neoclassical model one level of \(P\) and one level of \(\delta\) is associated with each price (in other words, correlation between \(P\) and \(\delta\) is spurious correlation). Physicians setting lower prices will both have more patients, and sell more to each of these patients, therefore \(\epsilon_{P\delta} > 0\). These empirical implications are summarized in the first row of Table 2, where \(\epsilon_{P\delta}, \epsilon_{\delta\delta}\) refers to the relation between these two elasticities, \(\rho_{\psi}\) to the fact that the neoclassical model cannot predict anything about the relation of these elasticities.

Let us now add altruism, as described in Section 2.1, to the neoclassical model. Because we have argued that \(P\) should be an argument in the utility-maximizing physician’s maximand, we again need to express \(P\) and \(\delta\) separately as functions of \(\phi\): \(^9\)

\[
\max_{\phi} u(\phi \delta(\phi) P(\phi) - C(\delta(\phi) P(\phi)), P(\phi), U). \tag{10}
\]

\(^9\)Note that price reductions are more efficient in producing utility from altruism than lump-sum transfers, as a price reduction will lead to the reduction of deadweight loss. This is counter to the claim of Becker (1981) that, to an altruistic firm lump-sum transfers are more efficient than price reductions. This is due to the fact that in Becker, prices can be reduced below marginal cost.
Manipulating the first-order condition (FOC) obtained from Eq. (10), we now obtain the following Lerner index:

$$\frac{\phi - C_0 + \gamma}{\phi} = \frac{1}{|\varepsilon_{Q_0}|}$$

(11)

where $\gamma$ is the physician’s altruistic marginal benefit of providing an extra $Q$, expressed in monetary terms. 10 Though this neglects the fact that the marginal benefit $\gamma$ is likely to decrease as $Q$ increases, $\phi Q/|\varepsilon_{Q_0}|$ still provides an approximation of the physician’s profits, which this time includes a non-pecuniary part. Because of the presence of this non-pecuniary part in the physician’s profits, higher physician profits do not necessarily mean lower consumer surplus. Therefore, this neoclassical model with altruism added to it does not allow us to approximate the reduction in consumer surplus caused by market power, nor the deadweight loss. Indeed, if altruism is strong enough, the physician will reduce price until it equals marginal cost, and the approximation of profits may then fully reflect altruistic profits. The welfare implications of the neoclassical model extended with altruism are summarized in the second row of Table 1.

We will now show that the predictions that price elasticity of demand is always higher than one, and that price always decreases upon entry, cannot be maintained in this model. This can be seen by rewriting Eq. (11):

$$\phi \left(1 - \frac{1}{|\varepsilon_{Q_0}|}\right) + \gamma = C_0.$$  

(12)

It is clear that because of the presence of an altruistic marginal benefit, $|\varepsilon_{Q_0}| < 1$ no longer implies that the right-hand side is negative. Several authors have noted the low price elasticities observed in health care, which are counter to the predictions of the market power model. Zweifel (1982) termed this the “riddle of elasticity”. 11 Altruism could now provide an explanation for these low price elasticities (Philips, 1982). However, the price elasticities referred to in the “riddle of elasticity” are market elasticities, whereas the predictions made are for the firm-level. Market elasticity and firm-level elasticity are only equal if each firm is a monopolist. For firms that have market power but are not monopolists, market-level elasticities lower than one are a very normal result. In any case, physician–firm elasticities have mostly been estimated to be larger than one. 12 Therefore, Philips’ explanation of the “riddle of elasticity” seems to be based on confusion of market- and firm-level units.

10 $\gamma = (w_p/w_f)(P_0/Q_0) + (w_U/w_f)(u_p/Q_0)$.

11 “Elastizitätsrätsel”.

12 We are indebted for this point to one of the referees.
Concerning the magnitude of the constituent elasticities $|e_{p\delta}|$ and $|e_{\delta \phi}|$, this model again cannot make any predictions. The same prediction about $e_{p\delta}$ as in the pure neoclassical model applies. Rewriting Eq. (12), we find:

$$
\phi = \left(1 - \frac{1}{|e_{Q\phi}|}\right)^{-1} \left(C_Q - \gamma\right)
$$

(13)

From Eq. (13), it can be seen that entry does not necessarily lead to a price decrease. If altruism is a normal good to the representative physician, i.e., if the income effect on the demand for altruism is positive, then entry may lead $\gamma$ to decrease enough to compensate for the increase in elasticity. If income effects on the demand for altruism are extremely strong, then we get satisficing behaviour rather than maximizing behaviour (Farley, 1986). The physician then targets a certain profit level, rather than maximizing profits. This is known as the target income hypothesis (Newhouse, 1970). Upon entry, the physician may now increase price such that their pre-entry income is maintained.

Positive cross-sectional correlations between price and physician density have been claimed to provide indication for the SID hypothesis in the literature. Indeed, as Reinhardt (1985) points out, this is the only prediction that distinguishes this hypothesis from the neoclassical model. However, assuming that physicians are altruistic, and that altruism is a normal good, suffices to explain this observation. The predictions for the neoclassical model with altruism added to it are summarized in the second row of Table 2.

4. Supplier-induced demand

We now describe Farley’s model and derive additional results from it.

4.1. Welfare economics

The utility $U$ of each of the identical patients is a function $u$ of consumption $k$ and consumption of health services $\delta$.

$$
U = u(k, \delta).
$$

(14)

The patient has income $y$, the price of health services is $\phi$, normalized such that the price of consumption is one. The budget constraint is then

$$
y = k + \phi \delta
$$

(15)

$k$ is substituted by using the budget constraint:

$$
U = u(y - \phi \delta, \delta).
$$

(16)

It is assumed that each consumer can only be a patient one time during the studied period. How the number of patients of a physician depends on the utilities
provided to patients is described above in Eq. (5). Filling in Eqs. (16) and (5) in Eq. (4), we obtain the following maximization problem, to be maximized not only w.r.t. \( \phi \), but in the demand-setting model of SID we have made a case for in Section 2.1, also w.r.t. \( d \):

\[
\max_{\phi, d} w \{ \phi \delta P - C(\delta P), P, u(\gamma - \phi \delta, \delta) \}.
\]

(17)

The FOC are

\[
\frac{\partial w}{\partial \phi} = \left[ u_\delta - \phi u_1 \right] \left[ \delta P_U (\phi - C_\phi) w_\gamma + w_p P_U + w_U \right] + P (\phi - C_\phi) w_\gamma = 0,
\]

(18)

\[
\frac{\partial w}{\partial \delta} = -\delta u_1 \left[ \delta P_U (\phi - C_\phi) w_\gamma + w_p P_U + w_U \right] + w_\delta \delta P = 0
\]

(19)

where it is assumed that the second-order conditions are met,\(^{13}\) and that the representative physician’s expectations about \( P_U \) are met, i.e., that there is a Nash equilibrium.\(^{14}\)

As we will now show, there are now two margins instead of the single margin derived in Section 3.\(^{15}\) We derive the first margin from Eq. (19)

\[
\frac{\phi - C_\phi + \gamma}{\phi} = \frac{1}{|\varepsilon_{P\delta}|}
\]

(20)

where \( |\varepsilon_{P\delta}| = P_\gamma, u_\delta \delta \phi / P \), and \( \gamma = (w_\gamma + w_p P_U) / w_\gamma \). \( \delta \) is shown in Appendix A to be the altruistic marginal benefit of treating an extra patient, expressed in pecuniary terms (i.e., the “shadow benefit”). From Eq. (18), it follows that

\[
\frac{\phi - C_\phi + \gamma}{\phi - C_\phi} = \frac{1}{|\varepsilon_{P\delta}|}
\]

(21)

where \( |\varepsilon_{P\delta}| = -\delta P_U [u_\delta - \phi u_1] / P \), i.e., the absolute value of the elasticity of \( P \) w.r.t. \( \delta \) for fixed \( \phi \). Eq. (21) is not a very familiar expression, but then again one does not usually assume firms to set quantity as well as price. Combining Eqs.

\(^{13}\) As Green (1978) points out, this does not automatically follow from concavity of \( w \) in \( Y, P \) and \( U \). So concavity of \( w \) in \( \phi \) and \( \delta \) is an additional condition.

\(^{14}\) The equilibrium conditions for a Bertrand–Nash equilibrium are more complex then is normally the case, as they involve two variables (\( \phi \) and \( \delta \) set by one physician) depending on two other variables (\( \phi \) and \( \delta \) set by another physician). We do not state them here as they are not essential to our argument.

\(^{15}\) To situate our contribution, Farley (1986) concentrates on two extreme situations: one where the physician is a pure profit maximizer (which leads her to derive Eq. (20), with \( \gamma = 0 \), and one where the physician is altruistic but does not face any competition (\( P_\gamma = 0 \). As can be seen from Eq. (19), the latter results in the physician equating marginal utility from own income to marginal utility from patient income: \( w_\gamma u_\delta = w_\gamma \).
and applying the implicit function theorem, we find the second margin, and a more familiar expression:

\[
\frac{\phi - C_Q}{\phi} = \frac{|\varepsilon_{PB}|}{|\varepsilon_{PB}|} = 1
\]

where \(|\varepsilon_{PB}|\) is the elasticity of \(\delta\) w.r.t. \(\phi\), for constant \(P\), and therefore for constant utility \(u\) of the representative patient.

Which of the two margins should we now employ to derive an approximation for the physician’s profits? As one would want this approximation to take into account that the physician also earns non-pecuniary profits, Eq. (20) is the best basis for deriving profits. With the same imperfection as above that \(\gamma\) is in fact marginal, and not average benefit, the profits per service \(\delta\) sold are \(\phi Q / |\varepsilon_{PB}|\). As a proportion of expenditure, profits therefore are \(|\varepsilon_{PB}|^{-1}\).

In the neoclassical model treated in Section 3, a positive margin implies Pareto inefficiency. As shown by Farley (1986), this is not the case in the inducement model, where the outcome is Pareto efficiency, whether or not the margin is positive. This is because in the inducement model, given \(\delta\), the physician is not obliged to set price such that it equals the patient’s marginal utility from this \(\delta\). Formally, combining FOC Eqs. (18) and (19) to eliminate \(\delta P \cdot (\phi - C_Q)w + w_k P + w_Y\), one obtains

\[
\frac{u_k}{u_k} = C_Q.
\]

This result is intuitively clear. As such, a profit-maximizing physician never has an incentive to waste consumer surplus by creating deadweight loss, as this prevents her from extracting this wasted consumer surplus. But the consumer’s power to select the quantity optimal to him, given the price set by the firm, forces the firm to create deadweight loss (see Section 3). If a physician has the power to set both price and quantity, however, she first sets a Pareto efficient quantity, such that total welfare is maximized, and then sets a price to distribute this total welfare between the patient and herself. In redistributing welfare from the patient to herself, she will take into account how this affects her number of patients and/or her altruism. Both the number of patients and altruism in turn have been argued in Section 2 to depend on the utility of the representative patient. Therefore, inevitably, the physician maximizes her utility given a certain utility level of the representative patient. But maximizing one’s utility given the utility of somebody else is nothing else than the definition of Pareto efficiency.

It is insightful to illustrate the result of Pareto efficiency graphically, and in particular with respect to the demand function of a perfectly informed representative patient. In Fig. 4, \(D\) is a representative patient’s perfect-information demand curve for health services, MR is marginal revenue, MC is marginal cost, and AC is average cost. In addition to these familiar curves, we construct iso-profit and iso-utility curves in the \(\phi\delta\)-space. One iso-profit curve is familiar, and is the
average cost curve (AC), i.e., the curve of $\phi \delta$-combinations yielding zero profits. Iso-profit curves for non-zero profit levels ($\pi_n$, $\pi_p$ and $\pi_{max}$) have the same shape as the average cost curve, and their minimum also coincides with marginal cost. Iso-utility curves deserve some clarification, and we show in Appendix B how they can be constructed for the example of a linear demand function. It is clear that lower iso-utility curves present higher utility levels, as along them, the same $d$ can be obtained for a lower price. Additionally, the maximum of every iso-utility curve is where it crosses the perfect-information demand curve $D$. This is because the perfect information demand curve is the locus of the optimal $d$ to the patient given each $\phi$. To keep utility constant, the patient has to be compensated by a lower price both if he is forced to consume less or forced to consume more than these optimal quantities given by the demand function.

All the tangency points of iso-profit and iso-utility curves, corresponding to different levels of competition and $r$ or altruism, evidently are at the level of $\delta$ where marginal benefit equals marginal cost, otherwise, they would not be Pareto efficient outcomes. In Fig. 4, the Pareto efficient level of $\delta$ is the same for different levels of utility, and therefore for different numbers of patients. Fig. 4 thereby seems to neglect that a higher number of patients implies a shift in marginal cost per patient. However, once one assumes that each of the tangency points corresponds to a different general equilibrium (corresponding to different levels of competition and/or altruism), the problem of shifting marginal cost is solved, as in equilibrium, the fixed number of patients in the market $Z$ will always be proportionally divided over physicians.

We can now derive the effect on patient welfare of the physician’s ability to induce. This leads us to the following proposition.

![Fig. 4. Welfare implications of SID.](image-url)
Proposition 1. If a monopolistic physician sets both price and quantity, and if she targets a level of income close to the income she obtains in the non-inducement case, then SID makes the representative patient better off.

We prove Proposition 1 with an example, graphically shown in Fig. 4. Let us assume that the representative physician cannot induce demand, and is a monopolist. This leads to the familiar result that the physician sets quantity such that marginal revenue (MR) equals marginal cost (MC), and is obliged to set the price corresponding to this quantity along the demand curve \( D \). This monopoly outcome is associated with physician profits \( \pi_m \) and patient utility \( u_m \). As monopoly involves deadweight loss, \( \pi_m \) and \( u_m \) are not tangent, and the shaded lens below \( \pi_m \) and above \( u_m \) shows price–quantity combinations which make both the representative patient and the physician better off. When the physician is given the ability to induce, she may now be choosing exactly one of these price–quantity combinations in the shaded lens. This occurs, e.g., when the physician happens to be targeting the income she would obtain if she would be not able to induce demand. In this case, the physician will spend the extra total welfare obtained by the ability to induce demand only for creating extra patient welfare, not for creating extra pecuniary profits.

Proposition 1 illustrates the importance of considering the effects of SID on patient welfare, in showing a case where SID makes the representative patient better off. But of course, if altruism does not take the form conditional for Proposition 1, a completely different result may apply, and the physician’s ability to induce may cause the representative patient’s full consumer surplus from health care to be extracted. Though one cannot derive an approximation for the patient’s loss in consumer surplus due to the physician’s ability to induce demand, one can in Farley’s inducement model derive an approximation for the patient’s loss due to market power. This is because contrary to the neoclassical model, as there is Pareto efficiency in Farley’s SID model, any pecuniary profits to physicians correspond to a loss in consumer surplus to the representative patient. Additionally, contrary to the neoclassical model, the two different margins allow us to distinguish pecuniary from total (pecuniary + non-pecuniary) profits. Margin (22) involves a change in the level of services \( \delta \) for a fixed number of patients \( P \), and therefore for fixed patient utility \( U \). Fixed \( P \) and \( U \) involves fixed non-pecuniary profits, however, and margin (22) therefore reflects only pecuniary profits. Therefore, besides an approximation for the proportion of revenue consisting of total profits \( |\varepsilon_{\pi_d}|^{-1} \), see above), in \( |\varepsilon_{\pi_d}|^{-1} \) we have an approximation of the loss to the representative patient of the physicians having market power, expressed as a proportion of revenue. These welfare implications of the SID model are summarized in row four of Table 1. We also provide there the results for the inducement model without altruism (row three).

We now go on to compare the welfare implications of the different models summarized in Table 1. Comparing the inducement model with the inducement
model extended with altruism, from putting $\gamma = 0$ in Eq. (20) and comparing to Eq. (22), it is clear that in the pure inducement model, $|\epsilon_{P0}| = |\epsilon_{\delta0}|$. The expression $|\epsilon_{P0}|$ as such does not have any economic meaning in the inducement model. It is not demand elasticity, but just the sum of $|\epsilon_{P0}|$ and $|\epsilon_{\delta0}|$, where $|\epsilon_{\delta0}|$ does not reflect demand elasticity, but reflects the elasticity of the iso-utility curve (see Fig. 4) going through the equilibrium $\phi \delta$-combination. Yet for making comparisons with the neoclassical model, it is useful to consider $|\epsilon_{\delta0}|$ in the inducement model anyway. We now see that in the pure inducement model, $|\epsilon_{P0}| = 1/2|\epsilon_{\delta0}|$. Comparisons between the inducement model and the neoclassical model can now be readily made for the profit values, as values are available for each of the models. Comparisons between the two models concerning loss in consumer surplus and deadweight loss are only possible in the case without altruism. Additionally, one should be careful in making comparisons, as in the inducement case, the values for the losses are approximations, whereas in the neoclassical case they are upper bounds on the losses (see Section 3), in that monopoly is assumed. The results from comparing the welfare implications of the different models are summarized in Proposition 2.

**Proposition 2.** Both with and without physician altruism, employing the neoclassical approximation of profits when the SID model is the true model means underestimating profits. The true profits could be up to twice as big. Concerning the loss in consumer surplus and the deadweight loss due to market power, only the non-altruistic variants of the neoclassical model and the SID model can be compared. Employing neoclassical upper bounds for these losses when the SID model is the true model means underestimating the loss in consumer surplus due to market power, and overestimating deadweight loss. The true loss in consumer surplus could be up to twice as big, whereas the true deadweight loss is zero.

The results in Proposition 2 concerning profits can be restated in terms of the Lerner index of market power. Of the two margins in the inducement model (20) and (22), it is clear that Eq. (20) provides the true Lerner index, as $P$ is the true unit of demand (Farley, 1986, p. 331). Comparing Eq. (20) to the Lerner-indices in the neoclassical case with and without altruism (Eqs. (7) and (11) in Section 3), the following proposition now follows.

**Proposition 3.** Employing the neoclassical Lerner index when the SID model is the true model means underestimating market power. Real market power could be twice as big.

### 4.2. Empirical implications

Section 4.1 provides approximations of profits, consumer surplus extracted, and deadweight loss for the different models treated, but does not provide any
predictions to distinguish between these models. We attempt to provide such predictions by deriving the magnitudes and relation of several elasticities in the different models, and by looking at the effects of entry on price.

Rearranging Eqs. (20) and (22), we find the expressions Eqs. (24) and (25) equating marginal benefit to marginal cost, for \( P \) holding \( \delta \) constant, and vice versa:

\[
\phi \left( 1 - \frac{1}{|e_{p\delta}|} \right) + \gamma = C_Q \tag{24}
\]

\[
\phi \left( 1 - \frac{1}{|e_{\delta\delta}|} \right) = C_Q. \tag{25}
\]

It follows from Eq. (24) that in the inducement model extended with altruism, \( |e_{p\delta}| \) could be smaller than one, whereas \( |e_{\delta\delta}| \) must be larger than one. Therefore, the result that firm-level elasticities smaller than one provide indication for altruism (see Section 3) still applies in the present inducement model, but only applies to \( |e_{p\delta}| \). In other words, \( |e_{\delta\delta}| \), which again is only relevant for making a comparison with the neoclassical model, must be larger than one in the inducement model. Additionally, \( |e_{p\delta}| \) must be smaller than \( |e_{\delta\delta}| \). Finally, the inducement model extended with altruism allows the possibility that price increases upon entry. These predictions are summarized in row four of Table 2.

In the inducement model without altruism, neither \( |e_{\delta\delta}| \) nor \( |e_{p\delta}| \) can be smaller than one. Intuitively, the marginal return from treating an extra patient given a fixed level of servicing is equated to the marginal return from providing each of a fixed number of patients with an extra service. As well, as both \( |e_{\delta\delta}| \) and \( |e_{p\delta}| \) are larger than one, \( |e_{\delta\delta}| \) must be larger than 2. Further, \( |e_{\delta\delta}| \) and \( |e_{p\delta}| \) are equal, and each equal \( \frac{1}{2} |e_{\delta\delta}| \). Finally, the pure inducement model does not allow increasing prices upon entry. These predictions are summarized in row three of Table 2.

We additionally derive results for \( |e_{p\delta}| \) in the pure inducement model, and in the inducement model extended with altruism. The pure inducement model involves a positive margin (Eqs. (20) and (22) are equal), and it is clear that a profit-maximizing physician wants to sell as much as possible given this positive margin. In the pure inducement model, for given \( \phi \), the physician therefore maximizes \( Q = P\delta \):

\[
\max_{\delta} P\delta
\]

\[
P\delta \delta + P = 0 \iff e_{p\delta} = -1.
\]

In fact, the inducement model, \( |e_{p\delta}| \) can be argued to provide a measure for the extent to which the physician is altruistic. Looking again at the approximations of
profits and of consumer surplus extracted given in the fourth row of Table 1, it is seen that 1/|ε_{Pd}| represents total profits (both pecuniary and non-pecuniary, and as a proportion of revenue), whereas |ε_{Pf}|/|ε_{Pd}| represents the loss in consumer surplus due to market power, and at the same time the pecuniary profits of the physician. It follows that (1 − |ε_{Pd}|)/|ε_{Pd}| represents the physician’s non-pecuniary profits. Therefore, ε_{Pd} provides a weight for the proportion of the physician’s profits constituting of pecuniary profits. With ε_{Pd} = −1, the inducing physician is a pure profit maximizer, whereas with ε_{Pd} = 0, her profits only consist of non-pecuniary profits (again, these are approximations neglecting the fact that marginal cost and altruistic marginal benefit could be non-constant).

Importantly, ε_{Pd} has a different magnitude in the neoclassical model (see Section 3). Correlation between P and δ is spurious correlation there, as P and δ cannot be simultaneously changed without changing price. As lower prices both involves the physician having more patients P and each patient buying more services δ, this spurious correlation is positive. In the inducement model, on the contrary ε_{Pd} is negative, reflecting the fact that the physician can set δ for given φ and that it is optimal to her to set δ such that additional δ makes the number of patients decrease. The results for ε_{Pd} are summarized in Table 2, and the comparison between the inducement model and the neoclassical model leads us to Proposition 4.

**Proposition 4.** A negative correlation firm-level patient-initiated and physician-initiated demand gives indication for SID. A positive correlation gives indication for the neoclassical model.

5. Caveats

We now state several caveats to the results presented in Section 4. First, the model ignores both time costs and insurance. If patients only co-pay a portion of the fee that physicians receive, or if each δ causes the patient a time cost, it can be checked that there will no longer be Pareto efficiency. In case of a coinsurance rate, there will be Pareto too much (Farley, 1986), and in case of a time cost, there will be Pareto too little. Second, in considering only one type of patients, the model does not treat the issue of accessibility. Indeed, in the logic of the model capitation and fee-for-service remuneration of physicians leads to the same outcomes. However, when there are several types of patients, the results apply only when physicians are able to third-degree price discriminate, i.e., charge different prices to different patients. Third, our patients react to higher levels of utility provided by physicians. This implies that a raise in utility by a physician will attract the same number of new patients no matter if it consists of high φ and low δ, or vice versa. However, it is likely that in the short run, differences in fees
are more easily discernible to patients than differences in treatment intensity (cf., Dranove and Satterthwaite, 1992). Fourth, the alternative SID models with Bayesian patients shortly reviewed in Section 2.2 should not be rejected a priori. Rather, these models and Farley’s model could apply for different levels of competition, and for different types of health care. For example, if the costs of switching to another physician are high, and if acquiring information about the optimal level of treatment is relatively cheap, then a patient who finds out that a new physician in the market is providing her patients with more utility may use this information to induce his current physician to provide him with more utility, rather than to switch to the new physician. Fifth, and importantly, both our welfare implications and our empirical predictions hinge upon the physician’s ability to set prices. This reflects the reality in only very little countries. This caveat is particularly important for the relevance of Proposition 1. Yet, our results are rescued if one assumes that \( \phi \) is implicit price, with \( \phi = f/h \), i.e., the implicit fee \( \phi \) equals real fee \( f \) divided by quality \( h \), where the physician is then in fact setting quality (cf., Bradford and Martin, 1995), and if one additionally assumes that \( \delta = dh \), i.e., real quantity \( d \) and quality \( h \) are perfect substitutes to the patient. Sixth, individual consumers may make a trade-off between not initiating demand, but running the risk of eventual high expenditures, and initiating demand, but running less risk of high expenditures. A negative \( \varepsilon_{p0} \) may then be explained by some physicians facing relatively less risk-averse consumers, and other physicians facing relatively more risk-averse consumers, and it does not reject the neoclassical model (see Table 2). Seventh, because of the costs involved with new patients, for constant \( Q \), treating more patients less intensively may be more costly to the physician than treating less patients more intensively (Hoerger, 1989). The cost function should then be \( C(P, \delta) \) instead of \( C(P0) \). In the SID model, margins (20) and (22) will then involve marginal costs \( C_p \) and \( C_\delta \), respectively. Looking at Table 2 now, \( |\varepsilon_{p0}| \neq |\varepsilon_{p0}| \) no longer rejects the pure inducement model. Eighth, the physician, rather than being an altruist, may be paternalistically valuing patient functioning instead of patient utility, or may be valuing applying her technical expertise. Indeed, one may wonder how the physician can find out the patient’s preferences in the first place. For the physician as a technician, who derives utility from “doing a good job”, one may imagine her utility function to be \( w(Y, \delta) \). This completely reverses the results in the inducement case. It can be checked that \( |\varepsilon_{p0}|^{-1} \) rather than \( |\varepsilon_{p0}|^{-1} \) will now be the best approximation of physician profits, as only the former involves non-pecuniary profits. Similarly, \( |\varepsilon_{p0}|^{-1} \) is now the approximation of the patient’s loss in consumer surplus. Furthermore, there no longer is Pareto efficiency. 16 Finally, it is \( |\varepsilon_{p0}| \) rather than \( |\varepsilon_{p0}| \) which

\[ \text{16 Unless one wants to include the physician’s non-pecuniary benefit of providing each patient with an extra health service in the Pareto criterion.} \]
6. Conclusion

A recurring problem in the SID literature is the failure to derive predictions that distinguish the SID hypothesis from alternative theories, mainly, the neoclassical model. Our approach has been to start from the premises of the SID hypothesis, namely, that physicians determine consumption once patients have initiated demand, but are constrained in doing so by their professional ethics, and to construct a plausible model given these premises. Our next step was then to derive several predictions from this model. Evidently, there are alternative explanations for such observations, were one to make such observations. For instance, rising prices upon entry, which have often been interpreted as an indication for SID, can perfectly be explained by the neoclassical model with altruism (see Section 3), and by models from the economics of information (e.g., Pauly and Satterthwaite, 1981). We, therefore, suggest that the predictions in Table 2 be used in the following way: when these predictions are not met, the SID model presented in this paper is rejected. This reasoning can be applied to each of the four models treated Table 2: 

- $|e_{sd}^d| < 1$, $|e_{sd}^d| < 1$, $|e_{sd}^d| > |e_{sd}^d|$ or $e_{ps} > 0$ reject the SID model. 
- $|e_{sd}^d| < 2$, $|e_{sd}^d| < 1$, $|e_{sd}^d| < 1$, $|e_{sd}^d| = |e_{sd}^d|$, $e_{ps} = -1$ or rising prices upon entry reject the SID model without altruism. 
- $|e_{sd}^d| < 1$, $|e_{sd}^d| < 1$, $|e_{sd}^d| = |e_{sd}^d|$ or $e_{ps} > 0$ reject the SID model with altruism. 
- $e_{ps} < 0$ rejects the neoclassical model. 

Our treatment suggests testing the SID hypothesis with firm-level data, for fixed physician density, rather than with aggregate household utilization data, for varying physician density. We argue that basing one’s tests on physician density increases the complexity of one’s underlying theory, as one additionally needs a model of entry. Still, our firm-level model also has implications for the literature which uses household data and changes in physician density to test for the SID hypothesis, and which treats the demand for health care in a two-part model (e.g., Grytten et al., 1995). In the two-part model, two equations are estimated, one reflecting the decision on whether or not to buy a treatment, and the second equation reflecting the physician’s decision on how much care the patient buys. A positive sign of physician density in the second equation is then argued to provide indication for SID. The two decisions are assumed to be independent, i.e., the patient is myopic and when deciding on whether or not to buy care, considers only the price per service, not the expected total expenditure on health care. Technically speaking, the two-part model assumes that the error terms of the two equations are independent. Maddala (1985) criticizes the use of two-part models, and suggests as a first alternative that one uses selectivity models, to account for any correlation.

can be smaller than one, and $e_{ps}$ must be smaller than minus one. It should be stressed, however, that inducement in this case is only constrained by competition.
between the error terms. But he criticizes this approach in that it still does not give any insight in the patient’s decision process, and therefore makes a case for estimating structural equations, where the first equation would include the patient’s expected expenditure as an explanatory variable. This approach fits the model treated in this paper, where a patient’s decision to consult a physician depends on the utility the patient expects the physician to provide.

Along with the question of deriving falsifiable predictions from the SID model, together with Labelle et al. (1994), we have emphasized the importance of focusing on patient welfare. To illustrate this, we have provided an example where physicians’ ability to induce demand actually makes the patients better off. To evaluate patients’ welfare, we have derived approximations of the patients’ welfare loss due to the presence of physician market power both in the neoclassical and in the inducement case. Even if the data would not allow us to distinguish between the SID hypothesis and the neoclassical model, the derived approximations can still tell us within which bounds patients’ welfare losses lie. The model we use to derive these results is highly stylized, and counter to reality where it assumes that physicians can set prices. However, this allows us to start from familiar demand functions, and the results generalize for fixed fees if one assumes that physicians in fact set implicit prices when they are able to set quality. The intuition we want to provide is that physicians determine the content of treatment, as this is exactly what patients expect them to do. Because of their ability to determine the content of treatment, as opposed to firms who are at arms’ length with patients, physicians are able to distribute income from patients to themselves without too much loss in efficiency. The extent to which physicians distribute income from patients to themselves can be constrained by physicians’ professional ethics, but is as we have argued additionally constrained by physician competition.

Acknowledgements

Support of the Cultureel Steunfonds van de VUB is gratefully acknowledged (OZR63). We would like to thank Bruno Heyndels, Katrien Kesteloot, two anonymous referees, one of the editors, and participants of the 1997 Congress of the European Association for Research in Industrial Economics and the 1997 Dag van het Wetenschappelijk Economisch Onderzoek in Vlaanderen, for their helpful comments on earlier versions of this paper. Any remaining errors are our own.

Appendix A. Derivation of the shadow benefit $\gamma$

Deriving the altruistic marginal benefit of treating an extra patient is done in the following way (this is due to Formby and Millner (1985)). A compensation
function $O$ is substituted for the physician’s income $Y$. For fixed physician utility $\bar{W}$ and for fixed patient utility $\bar{U}$, this compensation function tells us by how much income $Y$ the physician should be compensated to keep her utility constant at $\bar{W}$, after her number of patients has exogenously been decreased by one. Clearly, this monetary compensation is a function of $P$, of $\bar{U}$, and of $\bar{W}$, the level at which we are keeping physician utility fixed.

\[ w(O(P, \bar{U}, \bar{W}), P, \bar{U}) = \bar{W}. \]  

(A1)

The implicit function theorem tells us that although we do not know this implicit function $O$, we can obtain its derivative w.r.t. $P$:

\[ \frac{\partial O}{\partial P} \bigg|_{\bar{U}} - w_p = 0 \Leftrightarrow \left( \frac{\partial O}{\partial P} \right)_{\bar{U}} = \frac{w_p}{w_Y} \]  

(A2)

where subscript $\bar{U}$ means keeping patient utility fixed. This says that to keep physician utility constant upon a decrease of her patients by one, one must increase income her income by $w_p / w_Y$. This is therefore the altruistic marginal benefit of treating an extra patient, expressed in pecuniary terms, or the shadow benefit of treating an extra patient. In a similar way, the shadow benefit of giving the typical patient an extra unit of utility (for constant number of patients) is $w_U / w_Y$. However, what shows up in expression (20) is the marginal shadow benefit of treating an extra patient, not of giving an extra unit of utility to the typical patient. But $w_U / w_Y$ can transformed in the following way, again using the implicit function theorem:

\[ \left( \frac{\partial O}{\partial U} \right)_{P} = \left( \frac{\partial O}{\partial P} \right)_{P} \frac{\partial P}{\partial U} = \frac{w_U}{w_Y} \Leftrightarrow \left( \frac{\partial O}{\partial P} \right)_{P} = \frac{w_U}{w_Y} \frac{1}{P}. \]  

(A3)

Expression (A3) gives us the marginal shadow benefit of treating an extra patient, strictly through its “effect” upon the utility of the typical patient. We now add expressions (A2) and (A3). It is clear that, to obtain $\gamma$, this sum still needs to be divided by $\delta$, once it is realized that Eq. (20) is the per-health-service margin on treating an extra patient, and Eq. (22) is the margin on providing an extra health service per patient.

Appendix B. Construction of iso-utility curves

The top part of Fig. 5 contains two patient indifference curves $u_o$ and $u_t$, and the typical patient’s perfect information demand curve for health services within a treatment $D$, in the $k\delta$-space. $u_o$ is the indifference curve where the perfectly informed representative patient is indifferent between buying and not buying...
Fig. 5. Construction of iso-utility curves.

health services. It therefore crosses the Y-axis where \( k = y \). To find the corresponding curves in the \( \phi \delta \)-space, first it should be realized that in any case there will be no waste, i.e., the full budget \( y \) will be spent, and any relevant point will lie on a budget curve going through the point \( y \), and second, each price \( \phi \)
corresponds to the slope of a budget curve going through the point \( y \). In Fig. 5, the corresponding points in the \( k\phi \)- and \( \phi\delta \)-spaces are shown for three different price levels, \( \phi_0 \), \( \phi_1 \), and \( \phi_2 \). In this way, the iso-utility curves in the \( \phi\delta \)-space corresponding to the indifference curves \( u_0 \) and \( u_1 \) are obtained.

The highest (lowest utility) iso-utility curve is a straight line because the demand curve has been drawn linearly, implying an additive separable utility function for the representative patient. For such a utility function, the consumer surplus is a correct measure for the representative patient’s utility (in particular, consumer surplus is a monotonic transformation of utility, where the transformation consists of subtracting patient’s income \( y \) from utility). This is triangle \( abc \) for price level \( b \) in Fig. 6. At price \( b \), how much should demand be increased for the patient’s full consumer surplus to be extracted? It is seen in Fig. 6 that the consumer surplus is extracted completely when demand is doubled (the triangle \( abc \) equals the triangle \( cde \)). Making a similar construction for each price level, we thus obtain the linear iso-utility curve \( u_0 \).

Algebraically, if the demand function of the perfectly informed representative patient is of the linear form \( \phi = a - b\delta \), then the patient’s utility function is \( u = k + \delta(a - (b/2)\delta) \). Substituting for \( k \) using the constraint \( y = k + \phi\delta \), this becomes \( u = y - \phi\delta + \delta(a - (b/2)\delta) \). The function describing the points where \( u = y \) can then be seen to be the function \( \phi = a - (b/2)\delta \). \( u_0 \) is therefore a linear curve with half the slope of the demand curve.

Fig. 6. Linearity of highest iso-utility function.
Appendix C. List of symbols used

Roman letters
\(d\) Number of health services
\(D\) Perfect-information demand curve
\(EU\) Equilibrium utility
\(f\) Nominal price of health services
\(h\) Quality
\(k\) Consumption
\(MC\) Marginal cost
\(MR\) Marginal revenue
\(N\) Number of physicians in market
\(O\) Compensation function
\(P\) Number of patients of the physician
\(Q\) Total amount of health care sold by physician
\(RU\) The patients’ reservation utility for a treatment
\(U\) Patient utility
\(u\) Patient utility function
\(W\) Physician utility
\(w\) Physician utility function
\(y\) Patient income
\(Y\) Physician income
\(Z\) Fixed number of identical patients in market

Greek letters
\(\gamma\) Non-pecuniary return
\(\delta\) Nominal number of health services
\(\varepsilon\) Elasticity
\(\phi\) Implicit price of health services

References


Maddala, G.S., 1985. A survey of the literature on selectivity bias as it pertains to health care markets.


Wilensky, G. R., Rossiter, L.F., 1981. The magnitude and determinants of physician-initiated visits in
