Ramsey pricing and supply-side incentives in physician markets

Gerard J. Wedig

Health Care Systems Department, Wharton School, University of Pennsylvania, Philadelphia, PA, USA

Received July 1992, final version received July 1993

In this paper, I develop a theory of optimal prices using a relative value scale where there are fixed costs of medical practice. I consider a regulator constrained to set prices in excess of marginal cost, in markets where physicians can create demand at the margin. Under these conditions, basing prices on physician costs alone is shown to be suboptimal. Instead, prices should anticipate the behavior of physicians by setting profit margins highest for services least susceptible to demand creation. This is equivalent to a form of Ramsey pricing, used in this case to minimize the deadweight loss of oversupply.

Key words: Physician markets; Price regulation; Ramsey pricing

JEL classification: I18; L51

1. Introduction

It is widely believed that unregulated prices are a flawed mechanism for rationing physician services. There is less agreement, however, on the design of a regulatory alternative. The use of mandatory deductibles and copays mitigates moral hazard, but also exposes the consumer to additional financial risk [Zeckhauser, (1970)]. Partly for this reason, recent policy initiatives and reforms have turned to supply-side incentives designed to influence physician behavior. The supply-side approach is evident, for example, in Medicare’s resource-based relative value scale (RBRVS). The RBRVS uses resource-based prices to influence physician decisions such as service mix and specialty choice [Hsiao, (1988), (1991), Pauly, (1991)].

Unfortunately, designing supply-side incentives into a fee-for-service (FFS) payment structure poses special problems. This owes to two features of the...
medical market place: (1) the agency relationship between physician and patient which reduces consumer autonomy and gives rise to induced demand and, (2) the need to set FFS prices at levels greater than the physician's marginal costs, in order to pay for certain fixed costs such as practice expense and returns to medical education. Given profit-seeking physicians, prices which exceed marginal costs may encourage physicians to exploit the agency relationship by creating demand for unneeded services.

Seen in this light, Medicare's RBRVS attempts a 'second best' solution to the problem of service oversupply. Supporters of the RBRVS argue that by setting prices in proportion to marginal costs, the system achieves 'incentive neutrality' whereby physicians are financially indifferent to supplying different services. Under incentive neutrality, physicians with even a minor interest in patient welfare are motivated to supply the best mix if not the right level of services. In the present context, this means that physicians are free to substitute evaluation and management for procedural therapies, enter primary care markets, etc., without suffering a financial penalty.

Given the problems associated with cost-plus pricing, optimizing the mix of services conditional on program expenditures is a worthy goal. However, negotiations over average price levels invariably involve groups of physicians primarily concerned with their net incomes rather than overall program expenditures. It is arguably physician income, not programmatic expenditures, which sets the binding constraint on regulatory behavior. In this sense, the physician pricing problem is analogous to the classic regulated monopoly problem: prices should be set to maximize social welfare given a (physician level) profitability constraint.

This way of framing of the problem is significant, because the accepted economic theory of optimal price regulation subject to a profits constraint differs from the principle of incentive neutrality. According to the theory of Ramsey pricing, the markups of different services over marginal costs should be in inverse proportion to their measured demand elasticities in order to minimize deviations from the first best solution [Ramsey (1927), Baumol and Bradford (1970)]. As pointed out by Newhouse, however, these results apply in neoclassical markets which differ in important respects from the problem considered here (Newhouse, 1991).¹ Most importantly, the goal in the physician problem is to use prices to ration the supply rather than the demand side of the market. This raises the following important questions. Can Ramsey prices be adapted to the supply side pricing problem and, if so, do they retain their optimal properties?

These questions are the subject of this paper. I am able to show that under

¹Newhouse (1991), raises the possibility of using Ramsey pricing to allocate practice overhead, but notes that 'the Ramsey rule cannot be straightforwardly applied to physician fee schedules' [Newhouse (1991), p. 367]. This comment may be viewed as the point of departure for my analysis.
a particular (but not unreasonable) specification of physician utility, Ramsey prices are most preferred among the set of breakeven prices, including incentive-neutral prices. Intuitively, if one views the allocation of fixed costs as distortive of behavior, then the question is whether there is a way to allocate these costs in a way which distorts incentives the least. Ramsey pricing implies that if demand inducement for some services is less responsive to price than for others, then a greater share of fixed costs should be loaded onto these services. This results in the smallest amount of excess quantity and deadweight loss.

I use a model of physician behavior which is characterized by interdependent utility between physician and patient to derive these results. The model is quite general and allows me to consider extensions which include income effects, substitute and complementary services as well as the role of insurance. In each case, the robustness of the Ramsey pricing approach is confirmed. I conclude that Ramsey pricing is a promising approach for setting fees which deserves consideration in the ongoing debate over physician price reform. For example, if one views a major component of practice expense as fixed and joint to the various services the physician offers, the results here suggest a mechanism for allocating these expenses across the various services.

There are three important assumptions which must be maintained to arrive at these findings. First, I assume that physicians engage in demand inducement in the sense that they prescribe marginal services which the patient would not demand under full information. However, my model is consistent with a variety of mechanisms which limit the amount of demand creation, including the physician's conscience, or some model of patient search. Thus, the model I employ is consistent with both the 'traditional' demand inducement literature [e.g., Evans (1974)] as well as more recent economic models of the agency relationship [e.g., Dranove (1988)].

Second, I consider only fee-for-service arrangements and rule out capitation or two-part pricing as a regulatory option. Previously, Ellis and McGuire, (1990), showed that if physicians and patients bargain over quantities, a two-part pricing arrangement could be used to achieve the first best. I rule out two-part pricing strategies here, however, because they may be difficult to implement or may be politically inviable.2 My results should be viewed in the context of a regulator forced to manipulate a single pricing parameter.

Finally, I assume that fixed costs are a non-trivial part of the physician's costs and significant enough to force prices above marginal costs. There is some evidence to support this idea. It has been estimated that practice

---

2One of the problems with the two-part pricing scheme is deciding how the fixed payment should vary by practice size and scope of services offered. For example, a simple 'pass through' of all joint, fixed overhead may encourage inefficient expenditures on such overhead.
expenses are equal to almost one half of physician revenues [Gonzales (1990)]. Moreover, overhead may account for up to two thirds of this total. Similarly, the costs of physician training, including the time needed to complete medical school and specialty training generate a large fixed cost which must be amortized over the physician's working life [Burstein and Cromwell (1985), Marder and Willke (1991)].

The remainder of the paper is organized as follows. Section 2 develops the model and the basic application of Ramsey pricing to the supply-side problem. Three important extensions of the model are subsequently developed to demonstrate its robustness and generalizability. Section 3 offers some remarks on the applicability of the approach to current policy. The paper concludes in Section 4.

2. The model

Consider a single physician of a given specialty (e.g., general internist). The physician operates a multi-service practice ('firm') in which S different services are produced. The physician's long run economic costs consist of both joint, fixed costs, (e.g., medical training, rent, utilities, nonservice specific capital equipment) and variable costs, (his/her time, supplies, ancillary labor, etc.). Costs for the physician practice are written,

\[ C = F + C(q^s, q^f, \ldots), \quad C_q > 0, C_{q,q} = 0, \]

where \( q^s \) is the physician's production of service \( s \) and \( F \) is the joint, fixed cost. Variable costs are assumed to be constant over the relevant range of labor–leisure decisions considered here (this assumption is easily relaxed).

Consumers are insured for a constant fraction, \( \alpha \), of their medical expenses and a regulatory agency sets the maximum price, \( P^s \), physicians can charge for each service, \( s \). The marginal benefits of each service are initially assumed to be independent of the quantities of other services consumed (substitutes and complements are modelled later).

Physicians select quantities to maximize a utility function in two arguments: physician net income and patient benefits net of their out-of-pocket expenses (i.e., \((1 - \alpha)\) of billed charges). The physician's utility function is,

\[ U(q^s, q^f, \ldots) = U \left( \sum_s P^s q^s - C, B(q^s, q^f, \ldots) - (1 - \alpha) \sum_s P^s q^s \right), \]

where \( B(\cdot) \) are the aggregate benefits the physician's patients receive from treatment (assumed to be strictly concave in each argument, \( q^s \)). Without loss of generality, assume that \( U_1, U_2 > 0 \). Initially, also assume that \( U_1/U_2 \), the marginal rate of substitution between income and net patient benefits, is
constant across various income levels. Given that marginal costs are also assumed to be constant, this rules out income effects on the physician’s supply decision. Income effects are added to the model below and this assumption is relaxed at that point.

Maximization of this function with respect to \( q^1 \) gives the following first-order condition for the physician’s optimal choice of \( q^1 \):

\[
U_1(P^s - C_{q}^s) + U_2(B_{q}^s - (1 - x)P^s) = 0.
\] (2a)

Note that no explicit market constraints are placed on the physician. However, the inclusion of patient benefits in the utility function is consistent with either internal or market-driven constraints on behavior. For this reason, I argue that the model is consistent with the variety of demand inducement and agency models which have appeared in the literature. In general, the model predicts that quantities are a weighted average of desired supply and demand, similar to Ellis and McGuire (1990). No restrictions are placed on the relative magnitude of \( U_1 \) to \( U_2 \) and the model’s results are shown to be robust to this ratio.

The social planner’s problem is to maximize net economic benefits produced in the market. This is done by setting the relative prices for services, subject to the constraint that average prices allow the physician to recover some targeted level of his fixed costs. The specific fraction of fixed costs which is to be recouped may be dictated by a second (government-initiated) budget constraint.

2.1. Ramsey pricing without moral hazard

Consider the regulator’s problem in the absence of insurance (moral hazard) effects (i.e., assume that \( x = 0 \) in Equation (2)). This is a special case in which physicians behave as if patients pay the full price of the treatments they prescribe. Two possible justifications are that (1) physicians are ‘socially motivated’ or (2) they recognize that eventually all payouts are reflected in the insurance premia of their patients. It turns out that this special case is instructive and provides a useful starting point.

The regulator’s problem is similar to the usual Ramsey pricing problem, in that relative prices must be set to achieve a second best optimum while meeting a profitability objective. The key difference is in the way in which quantity is determined at the margin. Price has an ambiguous relationship with marginal benefit in the agency model. Marginal benefit is determined at the margin by the physician’s internal maximization problem and the relative weights assigned to profits versus patient welfare. Quantities generally differ from those which would result in a well informed market setting. Depending on the strength of the agency relationship, quantities may also be an
increasing or decreasing function of price, although the plausible case of imperfect agency implies a positive price elasticity. ³

The robustness of the Ramsey pricing formula can be demonstrated in spite of these anomalies. The social planner’s problem is to select prices, \( P^s \), to:

\[
\max \Phi = B(q^s, q^f, \ldots) - C(q^s, q^f, \ldots) - F,
\]

subject to

\[
\sum_s P^s q^s(P^s) - C(q^s, q^f, \ldots) = F',
\]

as well as (2a). Here, \( F' \) is interpreted as targeted reimbursement above marginal costs.

The first-order condition from maximizing (3) with respect to the price of any service, \( s \), is

\[
(P^s - C^s) = \frac{B^s_q - C^s_q}{\Omega} - \frac{q^s}{q^f},
\]

where \( \Omega \) is the La Grange multiplier associated with the profitability constraint.

Note that when \( \alpha = 0 \), according to (2a),

\[
(1 - MRS)(P^s - C^s) = (B^s_q - C^s_q),
\]

where \( MRS = U_1/U_2 \). ⁴

Using (6), substitute for \( (B^s_q - C^s_q) \) in (5) and solve to get,

\[
\frac{P^s - C^s}{P^s} = \frac{\Omega}{1 - MRS} - \frac{1}{\epsilon^s} = \frac{K}{\epsilon^s},
\]

where \( K \) is a constant which does not vary across services and \( \epsilon^s \) is the observed own price elasticity of service \( s \).

Note that neither \( \Omega \) nor \( MRS \) vary across services so the proposed optimum is consistent with the usual definition of Ramsey prices. In the

³The uncertain sign of the price elasticity is apparent upon totally differentiating Equation (2) with respect to price and quantity. In the absence of insurance and with the assumed restriction on income effects, the partial derivative is positive provided \( U_1 > U_2 \), where the physician values net income more than patient welfare at the margin. The opposite holds in the less plausible case of \( U_2 > U_1 \), where the physician values patient welfare more than net income at the margin.

⁴The designation \( MRS \) refers to the marginal rate of substitution between income and patient benefits in the physician’s utility function. When \( MRS \) exceeds 1, the physician acts as an imperfect agent in the sense of valuing his own income at a rate greater than patient welfare. The converse holds if \( MRS < 1 \). Finally, when \( MRS = 1 \) a ‘perfect’ agency relationship exists in the sense that patient welfare and physician income are weighted equally.
appendix I verify that second-order conditions to this problem hold as well. As part of that proof it is shown that sign \((1 - MRS - \Omega) = \text{opp sign (e)}\), so that the constant \(K\) is of the same sign as the measured elasticity.

These results imply that optimal price-cost margins are (1) always positive and (2) inversely proportional to the absolute value of the measured price elasticity. In contrast to the conventional Ramsey problem, \(\epsilon^2\) will, under the plausible assumption of imperfect agency, be positive. According to these findings, the Ramsey rule is still applicable and is used in this case to minimize the welfare loss associated with oversupply. Simple Ramsey prices are thus robust to the agency problem.

Equation (6) helps to provide the intuition for the general result. Ordinary Ramsey prices are based on the principle that price-cost margins measure the difference between marginal benefit and marginal cost (this follows from the fact that price equals marginal benefit in neoclassical markets). As a result, the price-cost margin alone contains all necessary information about the welfare loss of a marginal quantity increase or decrease. In the ordinary Ramsey formula, when price elasticities are small, the price-cost margin (marginal welfare loss) is permitted to be high, since the welfare loss occurs over fewer marginal units of quantity (due to the low elasticity).

In the agency model, physicians do not replicate the ordinary market outcome due to the arbitrary weight attached to patient utility versus physician income. However, a proportional relationship is maintained between price-cost margins and marginal welfare loss since the physician wishes to minimize the loss in patient welfare at any income level (see Equation 6). Price-cost margins still indicate the marginal welfare loss of incremental quantity (up to a proportionality constant) and the simple Ramsey formula is preserved. The Ramsey formula also works regardless of the sign of the measured price elasticity. The simple intuition here is that deadweight loss is weighted the same whether it occurs from excessive or deficient levels of quantity.

This intuition is illustrated graphically in Fig. 1. The top half of the figure (Fig. 1a) illustrates the intuition behind traditional Ramsey pricing. Here, I consider two services with different demand (marginal benefits) elasticities but identical, constant, marginal costs. For a given mark up over marginal cost (up to price \(P_o\)), deadweight loss for service \(a\) (curve \(D_a\)) is given by the area between the demand and marginal cost curves, the triangle bounded by \(BCE\). Conversely, service \(b\) (\(D_b\)), with a smaller demand elasticity suffers a smaller deadweight loss given by the shaded triangle \(ADE\). As noted in

\[5\text{In the special case in which } MRS = 1 \text{ (perfect agency) the physician always supplies pareto efficient quantities of each service no matter what their price (see Equation 6). This is because marginal increases in price have no net effect on physician utility since increases in income are exactly balanced by the perceived loss of patient utility. Here, quantities are fixed with respect to prices and any vector of prices which covers costs will also yield a first best outcome.}\]
Fig. 1. a. Traditional Ramsey pricing. b. Ramsey pricing in physician markets. Supply (demand inducement) curve assumes $MRS = 2$.

Baumol and Bradford (1970), this illustrates the rationale for the Ramsey rule of setting margins highest for services with the smallest demand elasticities.

The extension of Ramsey pricing to physician markets, as specified by the model, is illustrated below. Fig. 1b adds the ‘supply’ (demand inducement) curve of the physician for the plausible case where the physician is an imperfect agent for the patient, leading to a positive price elasticity. In this case, the intersections of regulated price with the supply curves $S_a$ and $S_b$ actually determine quantities in the physician market, $Q_a'$ and $Q_b'$, respecti-
G.J. Wedig, Ramsey pricing and supply-side incentives in physician markets 373

vely. Here, for convenience, I have assumed that $MRS=2$, so that the supply curve has the opposite slope and is thus a mirror image of the underlying marginal benefits curve (i.e., $S_a$ corresponds to $D_a$ and $S_b$ to $D_b$).\textsuperscript{6}

The figure illustrates the fundamental intuition behind the Ramsey approach to physician pricing. Namely, services with relatively inelastic marginal benefits curves will also have inelastic supply (demand inducement) curves. This is because physicians minimize net patient welfare loss at the margin through their relative quantity decisions. As a result, for a given increment of price above marginal cost, the service with the smallest price elasticity will also suffer the smallest deadweight loss as measured by the difference between marginal benefit and marginal cost. In Fig. 1b, the deadweight loss for service $b$ is given by the shaded triangle $A'D'E$ which is smaller than the deadweight loss for service $a$, triangle $B'C'E$.\textsuperscript{7} For this reason, mark ups over marginal cost should be higher for service $b$ than for service $a$.

This is the essential intuition for why the observable price elasticities from the physician's agency relationship with the patient contain meaningful information; because even in the case of imperfect agency, the elasticities maintain a link with the more relevant marginal benefits curve.

2.2. Ramsey pricing with cross-price elasticities: substitution and income effects

By assuming away income effects and interdependent benefits functions, the model used above rules out the possibility of cross-price elasticities in physician supply. This is done as an analytic convenience in order to keep the model simple and develop its basic intuition. These assumptions are now relaxed in order to achieve two other objectives: (1) to show that the basic Ramsey finding is robust to these extensions and (2) to show how proper measurement of the relevant elasticity changes in each case.

2.2.1. Income effects

Income effects are modelled by assuming that the agency relationship is a function of the physician's net income. In a reasonable model of physician–patient agency, the physician weights patient utility more heavily as his

\textsuperscript{6}Note that when $MRS=2$, Equation (6) describing quantity can be written as

\[ P = 2C_v - B_v \]

Thus, the supply curve has the opposite slope as the ordinary demand curve (defined by the simple equation $P = B_v$) and intersects the ordinary demand curve when $P = C_v$, as depicted in Fig. 1b.

\textsuperscript{7}Note that deadweight loss is still measured as the area between the marginal benefits and marginal cost curves.
As before, let $MRS$ denote the marginal rate of substitution between net income and net patient benefits, $(U_1/U_2)$. Then,

$$MRS = MRS(Y), MRS_Y < 0,$$

where $Y = \sum_i P_i q_i(P_i) - C(q^*, q', \ldots)$.

This assumption creates cross-price elasticities between services since increases in $P_i$ increase income and alter the agency basis for all service quantity decisions. Equation(s) (2a) now define an interdependent system of supply equations:

$$q^* = f^*(P^*, P' \ldots), \quad (9a)$$

$$q' = f'(P^*, P' \ldots), \quad (9b)$$

Changes in $q^*$ which result from a change in $P^*$ are now solved by totally differentiating the entire system of equations, and inverting the Jacobian matrix. While the algebra is tedious, essential intuition can be captured by considering the simple case where physicians supply only two goods. It is easily shown that for two services, s,t:

$$\frac{\partial(q^*)}{\partial(P^*)} = \frac{(1 - MRS)\Phi + B_{q_t,q_t}(1 - MRS)}{DET(J)}$$

$$\frac{\partial(q^*)}{\partial(P^*)} = \frac{(1 - MRS)\Theta}{DET(J)}$$

where $\Phi = -(P^* - C_q)^2(1 - MRS) + B_{q_t,q_t}(P^* - C_q)q_t^2$, $\Theta = (P^* - C_q)(P^* - C_q)(1 - MRS) + B_{q_s,q_s}q_s^2$, and $DET(J)$ is the determinant of the Jacobian. The own price elasticity contains both an income effect (the first term in the numerator of (10a)), as well as a substitution effect. The cross-price elasticity contains an income effect only.

Equation (3) can again be maximized subject to (4) as well as (10a) and (10b). After simplification, optimal prices are given by,

$$\frac{P^* - C_q}{P^*} = \frac{K}{(\epsilon^*/Y)}$$

where $K = \Omega/(1 - MRS - \Omega) - \sum_{n=1}^N (1 - MRS_Y)(P^* - C_q)^2B_{q_{n,q_{n}}}/DET(J)$ and $\epsilon^*/Y = B_{q_{t,q_{t}}}(1 - MRS)/DET(J)$.

The result has the following interpretation: optimal price–cost margins are
still inversely proportional to the (absolute value of) the own price elasticity, provided one uses the income compensated price elasticity. This is the elasticity which results if changes in price are coupled with a lump sum change in income which leaves overall income unchanged.

The result accords with intuition. Because the regulator's problem is to select from a set of relative prices, all of which leave income unchanged, income effects should not affect the basic form of the solution. It is important to note, however, that the income compensated elasticity differs in important ways from the ordinary price elasticity which is given in (10a). Specifically, the income compensated elasticity is not proportional to the ordinary elasticity. This means that the relative elasticities of two services differ depending upon whether one is measuring their income compensated or ordinary elasticities. As Equation (11) indicates, only the income compensated elasticities should be used to measure relative elasticities and set relative prices.

2.2.2. Substitutes and complements

Service substitutes and complements can be modelled by assuming that the marginal benefit of any service depends on the quantities of other services:

\[ B^s_q = B(q^s, q^t, \ldots). \]

Cross-price elasticities between services ensue because a price induced increase in one service affects the marginal benefits of other services and thereby affects their supply as well. For example, a price increase for a specific type of test will lead to increases in the supply of that test but will also lead to decreases in the supply of other, substitute tests, which are no longer needed.

As before, supply curves (Equations (2a)) are interdependent and take the same form as Equation(s) (9). The full effects of a price increase must again be solved by totally differentiating the vector of supply curves and inverting the Jacobian.

The intuition of how this affects optimal pricing can be captured with a simple three service good example. In this example, services s and t are regarded as substitutes whereas a third service, service u, does not substitute for either service s or service t. Consider an increase in the price of service s. It can easily be shown that,

\[
\frac{d(q^s)}{d(P^p)} = \frac{B_{q^s,q^t}(1 - MRS)}{DET(J)},
\]

\[
\frac{d(q^t)}{d(P^p)} = \frac{-B_{q^s,q^t}(1 - MRS)}{DET(J)},
\]

(13a) (13b)
The own-price elasticity is positive, and the cross-price elasticity of the
substitutable service is negative (assuming that $B_{s,t,s}$ is negative). Finally,
service $u$, which does not substitute or complement service $s$ has a zero cross-
price elasticity.

Using (13), (4) is again maximized and optimal prices are recomputed. Define $\varepsilon^{st}$ as the cross-price elasticity of service $s$ with respect to the price of service $t$. Optimal prices can then be written as,

$$\frac{P_s^* - C_q^*}{P^*} = K \cdot \frac{1 - (\varepsilon^{st}/\varepsilon^t)}{\varepsilon^s (1 - (\varepsilon^{st}/\varepsilon^t))} \cdot \frac{(1 - (\varepsilon^{st}/\varepsilon^t))}{(1 - (\varepsilon^{st}/\varepsilon^t))},$$

where $K = (\Omega/(1 - MRS - \Omega))$.

If we apply the 'symmetric' assumption that $(\varepsilon^{st}/\varepsilon^t) = (\varepsilon^{st}/\varepsilon^s)^8$, then (14) can be simplified to,

$$\frac{P_s^* - C_q^*}{P^*} = K \cdot \frac{1}{\varepsilon^s + \varepsilon^{st}} \cdot \frac{(1 - (\varepsilon^{st}/\varepsilon^t))}{(1 - (\varepsilon^{st}/\varepsilon^t))}.$$

(14a)

This form of the solution is most amenable to interpretation. According to (14a), the price increase's cross-elasticity effect is incorporated into an overall elasticity which is used to set the optimal mark up. This is to be expected, since the full effect of any price increase should account for its effects on the utilization of other services.$^9$

The case of perfect complements illustrates this most clearly. If two goods are perfect complements, then clearly they should be treated as a composite good and priced as a single good with a single elasticity. However, the own-price elasticity of either component of the composite good will underestimate the composite good's true elasticity. When the cross-price elasticity is added onto the own-price elasticity, the composite good's actual elasticity is properly measured. By using this aggregate elasticity to price each individual service, the composite good is properly priced and optimal quantities result.

2.3. Ramsey pricing with moral hazard

As a third extension to the basic model, consider the effects of insurance

$^8$This would hold if for example, the two services had (1) the same own-price elasticity, (2) the same marginal costs and (3) the same quantity level at the point where marginal benefit equals marginal cost.

$^9$The obvious exception is the case of income effects considered above. In that case, the effects of own price on other quantities is ignored because they are canceled out once average prices are adjusted to bring income back to its previous level.
and moral hazard on the optimal solution (i.e., assume that consumers are insured for a fraction, $\alpha$, of their expenditures). If physicians account for the effects of insurance on their patients' out-of-pocket expenditures, quantities are set such that,

$$B_q^s - C_q^s = (1 - \text{MRS} - \alpha)(P^s - C_q^s) - \alpha C_q^s.$$  \hspace{1cm} (2b)

Most notably, the welfare loss from marginal consumption is no longer proportional to the price-cost margin, but also reflects an additional term, $\alpha C_q^s$, which is due to the effects of moral hazard on consumption (marginal benefit is driven further below marginal cost by the effects of moral hazard). This is reflected in the solution to the regulator's problem, (4), which now becomes,

$$\frac{P^s - C_q^s}{P^s} = \frac{\alpha}{(1 - \text{MRS}) - \Omega} + \frac{\Omega}{(1 - \text{MRS}) - \Omega \frac{1}{1 - \text{MRS}}}.$$ \hspace{1cm} (15)

Price-cost margins are, as before, decreasing in the own price elasticity. However, the simple proportional relationship between the inverse price elasticity and the price-cost margin (the Ramsey rule) is replaced by a linear relationship.

The (linear) form of (15) is directly related to the (linear) form of (2b) and can be explained as follows. As in the original problem, the optimal solution equates the welfare loss of the marginal unit of quantity with a multiple of the price elasticity. However, the welfare loss of the marginal unit of quantity is now a positive function of both the moral hazard effect as well as the price-cost margin. For this reason, the inverse elasticity is set proportional to the sum of the price-cost margin plus the moral hazard effect. The intercept in (15), a negative term, represents the moral hazard effect subtracted from both sides of the resulting equation.

The fact that the intercept is negative in (15) is noteworthy. This indicates that optimal price-cost margins are greater than relative inverse price elasticities. Compared to the case of no insurance, a larger portion of the overhead cost is loaded upon those services with the smallest elasticities. Again, this can be explained algebraically. Reductions in measured elasticities must be met by a proportional increase in the sum of the price-cost margin plus the moral hazard effect. Because the moral hazard effect is constant and independent of the elasticity, this means that price-cost margins must respond more than proportionally to the reduction in the measured elasticity to preserve Equation (15). Thus, optimal price-cost margins respond even more sharply to measured elasticities where there is moral hazard.

An example provides some useful intuition. Consider a world characterized by a great deal of moral hazard and 'flat of the curve' medicine. In this
example, increases in the price–cost margin have a small relative effect on the marginal benefit of the last service consumed (the marginal benefit is already small). Therefore, in order to achieve a profitability objective while minimizing social welfare losses, the optimal strategy is to load virtually all of the profit margin onto those services with the smallest price elasticities in order to minimize the number of services supplied with little or no medical value. That is, optimal margins are set even higher for those services with small price elasticities compared to the case of no moral hazard. This is consistent with (15).

The regulator’s problem is more difficult in this case because elasticities and the profits constraint do not, in and of themselves identify the optimal vector of prices. The intercept in (15) must be identified as well. In general, this requires knowledge of both the coinsurance rate as well as the agency parameter, MRS.

It proves useful to have a robust pricing alternative to the formal set of prices implied in (15). Since optimal price–cost margins exceed relative inverse elasticities in Equation (15), one promising alternative is to consider a ‘third best’ solution in which relative prices are, as before, equal to relative inverse elasticities (the traditional Ramsey formula once again). Intuition suggests that such prices may be superior to the set of all relative prices which are less than proportional to relative inverse elasticities (are less sensitive to inverse elasticities than is implied by traditional Ramsey pricing).

This intuition can be investigated formally by considering changes in the relative prices of two services along the constraint path of breakeven prices. By showing that movements towards the traditional Ramsey solution always result in welfare increases, the regulator is given a robust alternative to prices set according to (15). The results of this inquiry are summarized in the following proposition:

Proposition I: Suppose ε>0 and α>0 with optimal prices described by (15), and that MRS>(1+αε) for all ε. Then for any two services, s, t, social

\[ \frac{p_s - c_s}{p_s} = \frac{p_t - c_t}{p_t} = K(1/\epsilon_s - 1/\epsilon_t), \]

\[ \frac{p_s - C_s}{p_s} = \frac{p_t - C_t}{p_t} = K(1/\epsilon_s - 1/\epsilon_t), \]

plus the breakeven constraint. Here, however, an additional unknown must be estimated, the parameter K. In general, K depends on both ε and the MRS, so that both α and the MRS are needed to identify this parameter and prices generally.

Weaker versions of the proposition are also easy to prove without assuming that MRS>(1+αε) for all ε. Most notably, in the absence of this assumption, the proposition still holds so long as the price–cost margin of service s is greater or equal than that of service t. Since equal price–cost margins is the incentive-neutral scheme, this means that movements towards Ramsey prices are always welfare improving if we start from incentive-neutral prices.
welfare is increasing for increases in \( P^a \) and decreases in \( P^t \) along the constraint path provided,

\[
\frac{(P^a - C^a_q)/(P^a)}{(P^t - C^t_q)/(P^t)} < \frac{\epsilon^t}{\epsilon^a} \quad \text{and} \quad \epsilon^a < \epsilon^t.
\]

Proof: See Appendix.

Proposition 1 indicates that, under the stated conditions, the Ramsey pricing rule is strictly preferred to any alternative set of prices on the constraint path in which the relative margins of services are strictly less than their relative inverse price elasticities. For example, Ramsey pricing would be strictly preferred to equal price-cost margins as reflected in the incentive-neutral approach. Moreover, continuous improvements in social welfare are realized along the constraint path provided services with smaller price elasticities have ever increasing relative margins, up until Ramsey prices are reached.\(^1\)

The policy significance of the result is two-fold. First, if the intercept in (15) is unknown, the regulator can use the robust strategy of traditional Ramsey pricing, confident that this approach is superior to a general class of alternatives. Second, if for any reason the regulator is constrained away from using the optimal result, the regulator can be sure that movements toward Ramsey pricing result in continuous improvements in social welfare. The social planner should therefore strive to increase the relative prices of less elastic services, even if full Ramsey pricing proves infeasible.

3. An application: Medicare's RBRVS and practice costs

The Ramsey pricing model can be applied in any FFS context in which physicians are paid unit fees which exceed marginal costs. Single payer national healthcare systems, as well as private insurers may benefit from this pricing strategy. Even HMO's which use risk pools and some degree of capitation may benefit from Ramsey pricing, provided they also use FFS arrangements with rates which exceed marginal costs.

Medicare's RBRVS provides a current example of how Ramsey pricing could be used to optimize physician price reform. Application of Ramsey pricing to these reforms raises a number of logistic issues which bear on the model's workability. These are discussed here in order to illustrate both how the model can be applied, as well as its potential limitations.

The RBRVS divides the physician's fees into three components, physician work, practice costs and medical malpractice. Currently, relative practice costs

\(^1\)Of course, increases in the relative prices of less elastic services beyond the Ramsey point have an uncertain effect on welfare, depending on the magnitude of the intercept in Equation (15).
expense payments are based on historical charge levels, although a legislative mandate exists to change this method. Ramsey pricing techniques could be used to make the allocation of practice expenses into relative prices. In this case, practice expenses would be divided between fixed, undifferentiated overhead and marginal costs. Subsequently, marginal practice costs would be added directly to the physician labor costs of the responsible services. Undifferentiated overhead, on the other hand, would be allocated based upon relative inverse price elasticities, using the Ramsey pricing formula.

This application of Ramsey pricing, as well as its cost and utilization implications have been simulated by the author and his colleagues [Pauly et al. (1993)]. The simulations yield some interesting and intuitive findings, such as the fact that patient-initiated contacts are paid relatively higher prices than envisioned under the incentive-neutral method of pricing. The major logistic issues which arise in this application concern the measurement of two parameters: (1) the marginal and overhead components of practice costs and (2) the income compensated price elasticity of service supply. Both of these measurements are needed to implement the Ramsey formula. In the case of practice costs, estimates of the marginal costs of each service have been taken from accounting estimates done by the Physician Payment Review Commission (PPRC, 1992). Admittedly, regression based parameters would be more reliable, but these are not currently available. Similarly, there are no reliable, existing estimates of income compensated price elasticities of supply, and so these simulations have used assumed values which are to some extent informed by existing research. The lesson is that the Ramsey pricing approach is information intensive and relies upon parameters which have not been precisely measured to date. Its application presupposes a certain amount of empirical work to reliably estimate the relevant parameters.

As a practical matter it may be impossible to obtain exact or even reasonably precise estimates of cost and price elasticity parameters. The resulting measurement error will cause prices to diverge from their theoretical Ramsey optimum. However, the use of incentive-neutral pricing (equal proportional markups) does not follow. As Proposition 1 indicates, prices do not have to mirror the Ramsey formula exactly in order to yield welfare improvements. Even with measurement error, welfare improvements are still guaranteed (relative to equal proportional markups) provided actual prices fall somewhere in between the Ramsey and incentive-neutral price levels (see Proposition 1).

This suggests the following pricing strategy. Where costs and price elasticities are estimated with error, confidence intervals about the estimates can be obtained. Partial Ramsey prices (an average of the Ramsey and incentive neutral price levels) can then be constructed, such that there is a high level of statistical confidence that prices fall between the 'true' Ramsey and incentive-neutral price levels. This 'do no harm' pricing strategy would
still result in welfare improvements. Even small differences in relative margins are preferred to equal mark ups over marginal cost so long as we are sure that there is some difference in the price elastic responses of different services to changes in price.

4. Conclusion

The use of supply-side price incentives to control the problem of overutilization has gained increasing visibility as reflected, for example, in Medicare's recently adopted resource-based relative value scale. Reliance upon supply-side policies to allocate resources, however, is based on a critical assumption which is not fully appreciated by policy makers: that physicians control marginal utilization decisions, not patients. Absent this assumption, supply-side pricing is an exercise in reallocating physician rents, with little effect on economic resource allocation (Pauly, 1991).

Conversely, if one adopts the assumption that physicians can influence marginal utilization, then the specific behavioral model of supply behavior is central to the model of pricing which is adopted. Supply-side prices must go beyond notions of fairness and equity and adopt a theory of physician behavior. This goes to the heart of the problem with incentive-neutral or cost-based pricing solutions. Basing prices solely on costs makes economic sense only if physicians behave as neoclassical profit maximizers and only if marginal-cost pricing proves to be feasible.

My analysis is based on an alternative, perhaps more plausible set of assumptions. First, in many physician markets, prices exceed marginal costs with the result that physicians would like to supply more services at the going price. Second, physicians control utilization at the margin. Together, these assumptions imply infinite supply unless one adopts a model of agency or demand inducement to explain how resources are rationed. The key finding, of course, is that a plausible model of demand inducement leads to a Ramsey rather than cost-based pricing solution.

While reasonable, the assumptions which underlie the results given here are somewhat controversial. The structure of physician costs and, more importantly, the model of demand inducement are key foundations of the theory which need to be measured and supported empirically. This underscores the need for additional empirical work in these areas. In this regard, tests of the model of quantity allocation as a weighted average of desired supply and demand (Equation (2a)) are especially relevant.

Appendix

1. Second-order conditions for case of \( \alpha = 0 \)

The following is a verification that the second-order conditions of the
Ramsey pricing solution hold in the absence of moral hazard. Since the marginal benefits of each service $s$, are independent of other quantities $t$, and since net income and the MRS are constant along the constraint path, there are no cross price elasticities along the constraint path. The matrix of second derivatives has nonzero elements along the diagonal only, and a necessary and sufficient condition is $\Phi_{p,p} < 0$ along the constraint path for all services $s$ (i.e., it is sufficient to show that all diagonal elements of the matrix of second derivatives of $\Phi$ are negative along the constraint path).

Compute $\Phi_{p,p}$ at the optimum (with the help of Equation (6)) as,

$$\Phi_{p,p} = q_p P \frac{d}{dP} \left[ (1 - MRS - \Omega) \frac{(P - C_q)}{P} - \frac{\Omega}{\epsilon} \right],$$  

(A.1)

where $PCM_p$ is the partial derivative of the price cost margin with respect to $P$ ($>0$) and where I suppress the superscript $s$ for convenience. Note that $\epsilon$ is constant and does not affect the second derivative.  

From inspection of (A.1), second-order conditions are satisfied (i.e., $\Phi_{p,p} < 0$) if $\text{sign} \left( 1 - MRS - \Omega \right) \Omega = \text{opp sign} \left( q_p \right)$. Also, according to (7) and the fact that $\Omega < 0$, $\text{sign} \left( K \right) = \text{sign} \left( \epsilon \right)$ under the same conditions.

In order to complete the proof, I verify that $\text{sign} \left( 1 - MRS - \Omega \right) = \text{opp sign} \left( q_p \right)$ as required. Since $\Omega$ is the shadow price of the constraint in this problem, start with,

$$\Omega = \frac{d\Phi}{dF} = (B_q - C_q) \frac{dq}{dF},$$  

(A.2)

where both terms on the right-hand side are vector quantities and $'$ indicates transposed vector.

In order to satisfy the profits constraint for any increase in $F$, the requirement is,

$$(P - C_q) \left( \frac{dq}{dF} \right) + q^' \left( \frac{dP}{dF} \right) = 1,$$  

(A.3)

which can be rewritten,

$$(P - C_q) \left( \frac{dq}{dF} \right) = (1 - \mu),$$  

(A.3.a)

where $\mu$ is some positive value, since increases in $F$ are always associated with increases in prices.

Recall that the first-order condition for physician behavior is,

$$(B_q - C_q) = (1 - MRS)(P - C_q).$$  

(A.4)

Substitute this into (A.2) to get,
\[ \Omega = \frac{d\Phi}{dF} = (1 - MRS)(P - C_q)q/dF, \] 
\[ \text{(A.2.a)} \]
and substitute in (A.3.a) to get,
\[ \Omega = \frac{d\Phi}{dF} = (1 - MRS)(1 - \mu). \]
\[ \text{(A.5)} \]
Finally, this gives,
\[ (1 - MRS - \Omega) = (1 - MRS)(\mu). \]
\[ \text{(A.6)} \]
where \( \mu \) is a positive value. Since sign \( (q_p) = \text{opp sign} \ (1 - MRS) \), it follows that \( (1 - MRS - \Omega) = \text{opp sign} \ (q_p) \) as proposed. QED

2. Proof of proposition 1

Proposition 1: Suppose \( \epsilon > 0 \) and \( \alpha > 0 \) with optimal prices described by (8) and that \( MRS > (1 + \alpha \epsilon) \) for all \( \epsilon^s \). Then for any two services, \( s, t \), social welfare is increasing for increases in \( P^s \) and decreases in \( P^t \) along the constraint path provided,
\[ \left( \frac{P^s - C_{q^s}}{P^s} \right) < \epsilon^t \text{ and } \epsilon^s < \epsilon^t. \]
\[ \text{(A.8)} \]
Proof: The derivative of welfare along the constraint path for increases in price \( s \) and decreases in price \( t \) is given by,
\[ \frac{d\Phi}{dP^s} = (B_q^s - C_q^s)q_p^s + (B_q^t - C_q^t)q_p^t(dP^t/dP^s). \]
\[ \text{(A.7)} \]
Using Equation (2b) this can be rewritten,
\[ \frac{d\Phi}{dP^s} = [(1 - MRS) \left( \frac{P^s - C_q^s}{P^s} \right) - \alpha]q^s \]
\[ + [(1 - MRS) \left( \frac{P^t - C_q^t}{P^t} \right) - \alpha]q^t(dP^t/dP^s). \]
\[ \text{(A.8)} \]
The profits constraint can be differentiated with respect to \( P^s \) and \( P^t \) to find that along the constraint path,
\[ (dP^t/dP^s)(q^t/q^s) = \frac{[(P^s - C_q^s)/(P^s)\epsilon^s + 1]}{[(P^t - C_q^t)/(P^t)\epsilon^t + 1]} \]
\[ \text{(A.9)} \]
Making this substitution into (A.9) and rearranging, it follows that \( d\Phi/dP^s > 0 \) if and only if,
G.J. Wedig, Ramsey pricing and supply-side incentives in physician markets

\[
\frac{[(P^s - C_q^c)/(P^s) - (\alpha/(1 - MRS))]\epsilon^s}{[(P^s - C_q^c)/(P^s)\epsilon^s + 1]} < \frac{[(P^t - C_q^c)/(P^t) - (\alpha/(1 - MRS))]\epsilon^t}{[(P^t - C_q^c)/(P^t)\epsilon^t + 1]}.
\]

(A.10)

Given that \(\epsilon > 0\), and \((1 - MRS) < 0\) (since second-order conditions for (8) hold), (A.10) is of the general form,

\[
\epsilon^s \frac{PCM^s + C}{PCM^s \epsilon^s + 1} < \epsilon^t \frac{PCM^t + C}{PCM^t \epsilon^t + 1}
\]

where \(PCM\) denotes price-cost margin and \(C\) is a positive constant. Upon inspection, the inequality holds if \(\epsilon^s PCM^s = \epsilon^t PCM^t\). To show it also holds when \(\epsilon^s PCM^s < \epsilon^t PCM^t\), note that since \(\epsilon^s C < 1\) (which follows from \(MRS > (1 + \alpha)\)), the expressions on both sides of (A.11) are increasing in the quantity \(\epsilon PCM\). This fact, together with the fact that \(\epsilon^s C < \epsilon^t C\) ensures that the inequality also holds if \(\epsilon^s PCM^s < \epsilon^t PCM^t\). QED

References