WORKING PAPER

Multitasking, Competition and Provider Payment

Karen Eggleston
Department of Economics
Tufts University

Discussion Paper 2001-01

Department of Economics
Tufts University
Medford, MA 02155
(617) 627-3560

Multitasking, Competition and Provider Payment

Karen Eggleston*

Many important dimensions of quality health care are difficult to observe, monitor, and motivate. This paper examines how competition among providers interacts with payment system incentives when the allocation of provider effort among multiple such dimensions or 'tasks' is noncontractible. The framework highlights that an optimal provider payment system, including optimal risk adjustment, should take account of provider multitasking.

JEL: I10
Keywords: payment incentives, competition, multitasking, capitation, managed care, rationing, risk adjustment

---

* Tufts University Department of Economics, Medford, MA 02155, and Research Associate, Kennedy School of Government, Harvard University; phone 617-627-5948; fax (617) 627-3917; karen.eggleston@tufts.edu. I am grateful to the Kennedy School and an NBER Predoctoral Fellowship in Aging and Health Economics for financial support and to William Encinosa, Albert Ma, Joseph Newhouse, Richard Zeckhauser, participants in the Seventh Northeast Regional Health Economics Research Symposium and anonymous referees for helpful comments. All remaining errors are my own.

I. Introduction

Many countries have introduced market forces into previously highly regulated and public-sector-dominated health care sectors. The goal is to improve their efficiency and responsiveness. It is hoped that competitive forces will inspire insurers and providers to strive for more cost-effective treatment options and forms of health care organization. Countries introducing such reforms include, for example, Belgium, Colombia, the Czech Republic, Germany, Ireland, Israel, the Netherlands, Poland, Russia, and Switzerland (van de Ven and Ellis 1999; Saltman and Figueras 1997). A stronger role of competition has also been debated in the US, particularly regarding the Medicare program.

Most nations also recognize, however, the many limitations of free-market competition in the health sector. These limitations arise predominantly from asymmetry of information, in particular the fact that many important dimensions of quality health care are difficult to observe, monitor, and motivate. Many socially important efforts cannot be written into health care contracts, either to assure adequate levels of desired efforts or to proscribe undesired efforts. In the tradition of Arrow (1963), we can say that there is an important missing market, the market for effort-contingent payments to health care providers. Provider competition to attract consumers can motivate some efforts, but consumers frequently lack sufficient information and market power to be discriminating purchasers of high quality, low cost services. In this "second best" situation, competition can lead to socially undesirable results, such as quality distortions to attract profitable and deter unprofitable patients. Concern about such effects has led many policymakers to doubt the effectiveness and desirability of increasing competition in health insurance and delivery markets.

Recent developments in the theory of optimal health insurance-provider payment systems, such as Ma and McGuire (1997) [hereafter MM], take the important step of considering the incentive effects of competition. But they do not highlight the implications of multiple dimensions of provider effort and how this interacts with competitive pressures and financial incentives. This paper seeks to fill this gap in the literature.[1] The theory builds upon MM and Frank, Glazer and McGuire (2000) by acknowledging that health care provider input to the health production process has many dimensions, and illustrating that this multi-dimensionality can have important policy implications. Since there is more than one kind of effort or "task" that the provider/agent can engage in for the patient/principal (or the payer/principal), the paper introduces a "multitask" principal-agent model into the provider payment framework.

The seminal work on multitask principal-agent theory is due to Holmstrom and Milgrom (1991).[2] This theory is particularly appropriate for analyzing incentives for agents called upon to perform many different tasks that are difficult to quantify, such as health care providers making intricate treatment decisions. Ma (1994) has applied the multitasking framework to provider payment systems, but abstracts from imperfect substitutability of efforts or possible effort complementarity and its implications, as highlighted here.[3] Provider competition also falls outside the scope of Ma's analysis.[4] Others have modeled payment incentives and policy instruments to combat risk selection (e.g., Lewis and Sappington 1999) and the effects of provider competition (e.g., Ellis 1998). But to the best of our knowledge, no one previously has focused on how competition among providers interacts with payment system incentives when the allocation of provider effort among multiple 'tasks' is noncontractible.

"Multitasking" for health care providers can be understood from several points of view. Medical professionals often "wear many hats". In addition to clinical care, there may be administrative and managerial responsibilities within a group practice or hospital, teaching and supervision tasks, as well as research and ongoing upgrading of technical skills. Doctors serve as "double agents", beholden both to patients on one hand and to their own group practice, hospital, or health plan employer on the other.

---

[1] This model illuminates and formalizes the central message of "Lesson 3" in Cutler and Zeckhauser (2000): "When consumer identity affects costs, competition is a mixed blessing. Allowing individuals to choose among competing health insurance plans can allocate people to appropriate plans and provide incentives for efficient provision. But it can also bring with it adverse selection…and gives plans incentives to distort their offerings to be less generous on care for the sick" (Table 10).

[2] For a recent overview of work on multitasking, see Prendergast (1999), pp. 21-29.

[3] Ma assumes that cost control and quality enhancement efforts are perfect substitutes in the provider's cost-of-effort function.

The multiple tasks that a provider engages in have often been modeled as one of two broad categories of "effort," such as cost control and quality-improvement effort (e.g., Ma 1994; Hart Shleifer and Vishny 1997). But even treating a specific condition for a specific patient can involve multiple tasks. A surgeon's work on a case might include reviewing patient records, having a preoperative office visit (to make diagnoses and discuss the benefits and risks of surgery), writing operative notes, performing the surgery, and then following up with postoperative care to treat complications and/or to monitor recovery. The efforts involved in different tasks may differ in their monitorability and in how they relate to each other and to formal treatment (i.e., either through a substitute or complementary relationship). For example, there may be economies of scale and scope involved with reviewing several patients' records and preparing staff for a set of particular services, or communicating with other professionals about certain kinds of cases. Alternatively, time and attention devoted to one patient may directly reduce the time available for treating a different case.

In research supporting the adoption of the Resource-Based Relative Value Scale (RBRVS) for physician payment, Hsiao et al. (1988a,b,c) analyzed what constitutes physician "work" (common parlance for "disutility of effort") through a survey of physicians in 18 specialties based on carefully crafted clinical vignettes. Since physician work includes "diverse tasks" (Hsiao et al. 1988b: 2350) and aspects of work are "often fragmented and intermingled with other activities" (Hsiao et al. 1988a: 882), the researchers asked surveyed physicians to evaluate work involved in separate phases of preservice, intraservice, and postservice care.

The RBRVS researchers found physician intraservice work to be a function of four dimensions: time; mental effort and judgment; technical skill and physical effort; and stress (arising from uncertainty regarding diagnosis and treatment and risk of iatrogenic harm to the patient). All four components were important for all specialties, and seemed to be partial substitutes in terms of providers' disutility cost-of-effort.[5] In other words, a

[5] The RBRVS researchers estimated total work as a Cobb-Douglas function of these four effort dimensions, deriving specialty-specific elasticities of work with respect to each dimension. In all cases the estimated elasticities were positive and significant.

physician may exert the same level of total work (disutility) for a short but intense and stressful procedure as for a longer but less mentally and physically demanding procedure.[6] Similarly, two procedures requiring about equal time may involve very different amounts of "work." For example, although orthopedic surgeons estimated that a "comprehensive office visit for initial evaluation of 48-year-old man with recurrent low-back pain radiating to the leg" and "decompresssion of carpal tunnel in 48-year-old woman, unilateral, ambulatory surgery unit" would both require about half an hour, the latter calls for significantly more mental and physical effort and risk, and therefore embodies almost twice the amount of "work" (Hsiao et al. 1988c: 2365).

The RBRVS study underscores that important aspects of physician effort are fundamentally noncontractible (e.g., mental effort) and interact with other efforts a provider undertakes. The estimated relationships suggest that the marginal effort cost of mental effort (i.e., the derivative of work with respect to mental effort) increases with the time devoted to the case. Quite intuitively, sustaining high mental effort for longer periods is more difficult than for shorter periods. Similarly, the marginal effort cost of physical effort is increasing in the level of mental effort.

Clearly, then, quality care involves multiple aspects of effort or multiple tasks, and certain aspects of physician effort, such as mental effort and judgment, are fundamentally noncontractible. The multitasking approach warns that "generally, the desirability of providing incentives for any one activity decreases with the difficulty of measuring performance in any other activities that make competing demands on the agent's time and attention" (Holmstrom and Milgrom 1991: 26). For example, to the extent that many important dimensions of quality cannot be measured or contracted upon, it is less desirable to give strong incentives for measurable quality indicators that compete for the provider's attention. If appropriate allocation of physician effort across different

[6] For example, specialists in internal medicine rated the work involved in an "initial office evaluation of a 37-year-old generally well man" to be very similar to that of a "hospital visit, day 3 following admission, 65-year-old man with uncomplicated myocardial infarction, established patient"; the hospital visit would be much shorter (18 vs. 34 minutes) but involve significantly more mental effort and risk. An additional illustration comes from pediatrics, where although a "comprehensive well-child visit at 5 years, new patient" would take 24 minutes compared to 14 minutes for "lumber puncture of 5-year-old for suspected meningitis," the latter calls for more mental and physical effort (and involves greater risk), so the estimated "work" is slightly more (Hsiao et al. 1988c: 2364-2365).

tasks is important for quality health care, payment systems should take this multitasking into account.

The paper is organized as follows. The next section presents a simple two-effort (two-task) model of health care provider multitasking. The remainder of the paper analyzes the interaction of payment incentives and competition in light of provider multi-tasking. First, a simple extension of the single-treatment model highlights how competition reinforces provider sensitivity to financial incentives when choosing efforts. Then we incorporate multitasking, professional ethics and supply-side cost sharing into the Frank, Glazer and McGuire (2000) model of managed care and discuss the implications for optimal provider payment, including optimal risk adjustment. Section IV concludes.

II. Provider Multitasking: A Two-Effort Model

To illustrate the implications of health care providers engaging in more than one type of effort (multitasking), consider a simple two-effort case. A provider may engage in two tasks--or equivalently, exert two kinds of effort--e_1 and e_2. The "work" associated with these efforts causes disutility of $G(e_1, e_2)$, increasing and convex in each individual effort level.[7]

The two efforts may complement or substitute for each other in the cost-of-effort function. If e_1 and e_2 compete for the provider's energy and attention, then a higher level of one effort will increase the marginal cost of the other effort: when $G_{12} > 0$, e_1 and e_2 are cost-of-effort substitutes. Alternatively, a higher level of e_1 may make increasing e_2 easier (i.e., $G_{12} < 0$), in which case e_1 and e_2 are cost-of-effort complements. For example, the RBRVS study (Hsiao et al. 1988a,b,c) found that physicians considered mental and physical effort to be cost-of-effort substitutes. The marginal effort cost of physical effort was increasing in the level of mental effort. Although this was generally true across all specialties, the elasticity of work with respect to different efforts differed across specialties. For example, the elasticity of work with respect to mental effort was estimated to be 0.414 for radiology, compared to 0.236 for general surgery; whereas the elasticity of work with respect to physical effort and technical skill was 0.417 for general surgery and only 0.141 for radiology.[8]

Let t represent medical treatment. Higher levels of t denote a higher intensity or number of services. A patient's benefits from treatment depend upon both the level of treatment, t, and the physician's efforts, e_1 and e_2. Let $F(t, e_1, e_2)$ denote the (monetary equivalent) of treatment benefits. As emphasized by MM, provider effort may complement or substitute for intensity of medical treatment. To capture some patient role in deciding on a course of treatment, assume, following MM, that patients choose treatment quantity t. The patient may choose, for example, whether or not to return for a follow-up visit or undergo a recommended procedure. The patient's choice is informed and influenced by provider efforts and by the required co-payment, b. For example, provider effort in improving amenities of care will likely induce the patient to choose more treatment (for any given co-payment). Effort to improve technical quality may also lead to increased demand for services. Patient responsiveness to these two different kinds of effort may differ. The provider may also engage in various cost-control tasks; such efforts plausibly would substitute for formal treatment. Better coordination of care, for example, could decrease duplicative tests or procedures. Provider allocation of effort between amenities, technical quality, and cost control may be influenced by financial incentives.

The timing in the model is as follows: The payer (e.g., health plan) determines the provider payment system and patient co-payment level. Providers then choose effort levels, and patients choose treatments in response. Finally, expenses are incurred and payments made. For simplicity we follow MM in assuming full patient information when choosing treatment quantity.

---

[7] To facilitate comparison with MM's results, we follow MM's notation.

[8] Interestingly, for several specialties (e.g., general surgery, pathology) the estimated work equation exhibits (slightly) increasing returns to scale: doubling all of the efforts--mental, physical, stress, time--would more than double the total amount of work (disutility).

Consider first the patient's choice of treatment. A patient must pay co-payment b per unit of treatment, so the marginal cost of t for the patient is b. A patient weighs this cost against the marginal benefit from treatment, $F_t(t, e_1, e_2)$. The first order condition for the patient choice of treatment quantity will define treatment as a function of provider efforts:

$$F_t(t, e_1, e_2) = b \Rightarrow t(e_1, e_2). \tag{1}$$

The function $t(e_1, e_2)$ embodies the patient's reaction to the provider's choice of efforts. The provider, foreseeing patient responsiveness to quality improvement, cost control, and other tasks, will in turn take this reaction function into account when choosing how much effort to allocate to each task.

The provider is assumed to choose effort levels primarily to maximize net revenue. (Professional ethics may also influence choice of effort levels, either through the shape of the cost-of-effort function G(.) or through modification of the objective function, as discussed below). The insurer or health plan pays the provider a prospective payment r (net of any fixed costs), plus a reimbursement per service of d (positive or negative) above variable cost **c**. In other words, the fee per service is $d + c$. A fee-for-service (FFS) payment system with a positive profit margin per service would be represented by r =0 and $d + c > c$ (i.e., $d > 0$). Any payment system with $d + c < c$ (i.e., $d < 0$) denotes supply-side cost sharing. Flat prospective payment corresponds to r >0 and $d + c = 0$ (i.e., $d = -c$).

The provider would like to maximize revenue, which may depend upon treatment, less the cost of treatment and the cost of effort. Thus, the provider seeks to maximize $r + (d + c)t - ct - G(e_1, e_2)$. For simplicity, assume the provider reports the number of services (i.e., files claims) truthfully.[9] Then net reimbursement per service will be $(d + c)t - ct$, or simply dt. The payment margin d captures a provider's treatment-based financial incentives. If the payment margin is positive (negative), net revenues increase (decrease) in treatment.

[9] See MM for a relaxation of this assumption in the single-task context.

The provider chooses efforts in light of how they influence patient choice of treatment: $t = t(e_1, e_2)$. The provider therefore seeks to maximize net revenue according to the following program:

$$\underset{<e_1,e_2>}{Max}\{r + dt(e_1,e_2) - G(e_1,e_2)\}. \qquad (2)$$

This objective function leads to symmetric first order conditions, where subscript i refers to the partial derivative with respect to effort i [i=1,2]:

$$dt_1 = G_1, \text{ and} \qquad (3)$$

$$dt_2 = G_2. \qquad (4)$$

The provider invests in each effort up to the point where the marginal gain in reimbursement equals the marginal effort cost.[10] The marginal reimbursement benefit will be either additional fee-for-service revenue or avoided cost-sharing liability.[11] More specifically, the left hand sides of the first order conditions will be positive in two cases: the payment margin is positive ($d > 0$) and the effort induces more treatment ($t_e > 0$), or the payment margin is negative ($d < 0$) and the effort substitutes for treatment ($t_e < 0$). For example, if effort to increase quality of care encourages patients to utilize more services, fee-for-service payment financially rewards providers for investing in quality-improving effort. In contrast, if cost control reduces demand for formal treatment, then supply-side cost sharing (such as capitation) creates incentive to control cost.[12] If both quality improvement and cost control are desirable, however, there will be problems setting the single payment parameter to induce optimal levels of both kinds of effort. Fee-for-service may induce quality investments but encourage little or no attention to cost control; supply-side cost sharing rewards cost control but not quality improvements.

---

[10] Equations (3) and (4) are comparable to MM equation (7).

[11] This simple model abstracts from other factors influencing choice of efforts, such as malpractice costs.

[12] In many cases the incentives facing a salaried provider will parallel those of a capitated provider. A salaried provider is "at risk" for effort associated with treatment and sacrifices savings of on-the-job leisure, much as a capitated provider is at risk for the costs of treatment and sacrifices residual monetary savings.

These simple first order conditions therefore illustrate that when providers may engage in various kinds of noncontractible efforts, reliance on a single policy instrument--treatment-based payment--cannot in general give incentive for appropriate levels of each effort. This is a health economics example of a more general phenomenon: to reach multiple targets, the number of policy instruments should be no less than the number of targets.[13] If policy targets include appropriate levels of both quality enhancement and cost control, then the single policy instrument of net reimbursement per service needs to be supplemented with an additional policy instrument. Perhaps separate payment arrangements are possible for different dimensions of treatment or for different tasks that health care providers must perform. Other potential policy instruments include minimum quality standards, selective contracting, competition, and reputational effects.[14] The range of instruments will be constrained by the fundamental noncontractibility of many aspects of quality health care. Payment rewarding measurable aspects of quality such as through "quality report cards," for example, gives incentive to allocate effort to those measured aspects of quality at the expense of other dimensions of quality.

The comparative statics of the two-effort model help to clarify when noncontractibility of effort may prevent a single treatment-based payment instrument from simultaneously achieving target levels of two provider efforts. Assume that the first order conditions are necessary and sufficient to solve the provider’s maximization problem. Then it can be shown that the comparative statics with respect to the payment margin d are

$$\frac{\partial e_1}{\partial d} = \frac{-t_1\left[dt_{22} - G_{22}\right] + t_2\left(dt_{12} - G_{12}\right)}{H}, \text{ and} \tag{5}$$

---

[13] See, for example, Tinbergen (1991), who notes that "to solve a [quantitative] problem the number of equations has to be equal to the number of unknowns. It follows that the number of instruments must be equal to the number of targets. Exceptions to this rule are possible only if the structure of the equations is abnormal" (p. 35).

[14] Some other contributions to the payment theory literature can also be understood in terms of the instruments-targets approach. For example, MM find that competition can expand the set of implementable efforts beyond that achievable by using the single payment margin instrument alone. Encinosa (2000) investigates payment systems when provider firms can choose three different organizational forms (pooling, specialization, and segmentation); this introduces an additional efficiency target, the socially desired organizational form. He finds that additional policy instruments, such as a minimum quality standard or

$$\frac{\partial e_2}{\partial d} = \frac{-t_2[dt_{11} - G_{11}] + t_1(dt_{12} - G_{12})}{H}. \qquad (6)$$

H is the Hessian, positive by the second-order sufficient condition for maximization.[15] The terms in brackets are also signed by the second-order conditions (and are negative).

The signs of the comparative statics depend upon several factors, including whether each effort is a complement or substitute to treatment; how they interact when the patient chooses treatment; and whether the efforts are complements or substitutes in the provider's cost-of-effort function. The possible combinations are summarized in Table 1 as six basic cases (defined by assumptions on t_1, t_1, and G_{12}), with five sub-cases for each (depending upon the signs of t_{12}, the payment margin, and $(dt_{12} - G_{12})$).

The impact of a change in the payment margin on effort can be decomposed into two effects, the "direct" effect and the "multitasking effect". These correspond to the first and second terms in the comparative statics, respectively. The direct effect captures the direct impact of the payment margin change on the marginal benefit of exerting effort. The multitasking effect then takes into account how the change in one effort will impact the marginal benefit and marginal cost of the other effort (through G_{12} and t_{12}).

Consider first the "direct effect". When treatment and effort are complements (substitutes), the direct effect is positive (negative): increasing payment above variable cost will encourage (discourage) efforts that complement (substitute for) treatment. For example, the direct effect of a higher fee per service on quality enhancement effort would be positive if quality enhancement encourages patients to demand more treatment. A provider who exerts quality enhancement effort will benefit from increased demand for services, and that higher demand will translate into higher revenues when the profit margin per service increases. In contrast, the direct effect of a higher profit margin per service will be to discourage provider efforts that substitute for treatment, such as cost control.

---

risk-adjusted cost-based payments (i.e., separate payment margins for high and low risks), are necessary to achieve the multiple efficiency targets.

[15] The expression for the Hessian is $H = [dt_{11} - G_{11}][dt_{22} - G_{22}] - (dt_{12} - G_{12})^2$.

When the two efforts do not otherwise interact, either in the provider cost-of-effort function or in the patient's treatment-responsiveness function (i.e., $G_{12} = 0$ and $t_{12} = 0$), then there is no additional "multitasking effect". The second term in (5) or (6) is 0. The sign of the comparative static will depend entirely on whether the effort is a complement or substitute to treatment. This single-task setting is summarized in MM Proposition 1: when treatment and effort are complements, effort is above zero when the payment system is fee-for-service (FFS); when effort and treatment are substitutes, effort is above zero under supply-side cost sharing.

When providers engage in multiple tasks and the model recognizes that these tasks may interact, the effect of a payment margin change on provider behavior is more subtle and complex. In the multitasking model summarized in Table 1, the comparative statics, when determinate, follow the sign of the "direct effect." The "multitasking effect" in some cases reinforces, and in others offsets, the direct effect. In an extreme case, the multitasking effect could outweigh the direct effect, leading to a result contradicting standard theory (e.g., countering MM Proposition 1).

In some special cases, when the multiple tasks relate to each other in specific ways, there is a straightforward extension of the single-task results to the multitasking setting. If both efforts either complement or substitute for treatment (i.e., t_1 and t_2 are of the same sign) and are complements in provider effort (G_{12}<0), then multitasking effects will generally reinforce the direct ("single task") effects. When treatment and both efforts are complements, effort levels are above zero when the payment system is FFS (Case 1.1); when both efforts and treatment are substitutes, effort levels are above zero under supply-side cost sharing (Case 3.1).

If a provider has motivations beyond pure net-revenue-maximization, the results are less stark. A physician who cares about patients may exert effort to improve quality even when financial incentives, such as from capitation payment, would suggest that the quality effort level should be zero. The following section describes the impact of professional ethics or "agency" for patients more formally. Here, however, we already make use of these results by assuming that providers may exert positive levels of efforts even when not in their financial best interest to do so.[16] Changes in payment incentives will affect efforts on the margin by encouraging a higher or lower level compared to the "professional norm" (rather than zero).

To illustrate the channels of impact from an increase in the payment margin, Table 2 traces out the direct and multitasking effects under different scenarios. We will here summarize only two illustrative cases highlighted in Table 2. First, consider two tasks or efforts that both complement treatment (Case 1). Perhaps e_1 represents technical quality improvement and e_2 represents better amenities for consumers. Assume that these two kinds of effort are complementary to treatment in the sense that higher technical quality and better amenities will each individually increase utilization. Then the direct effect of increasing the payment margin will be to induce positive levels of both kinds of effort, or effort above and beyond some minimum professional standard.[17] If this direct effect dominates, fee-for-service payment will reward improvement in both technical quality and amenities.[18] The fact that e_1 and e_2 are complements in provider effort serves to reinforce the direct effect, because a higher level of one effort reduces the marginal cost of the other. A second multitasking channel of impact, through patient treatment responsiveness to the other effort, may either reinforce or offset the direct effect (see Table 2).

In many circumstances, providers engage in tasks with differing relationships to treatment: some tasks complement, and others substitute for, billable treatments (Cases 2 and 5). For example, let e_1 continue to represent technical quality improvement, but e_2 represent cost control effort that substitutes for treatment. In this case, the direct effect of

---

[16] For example, the cost-of-effort function G(.) could be negative over some initial range of effort, so that the provider gains utility from exerting a certain "norm" of effort on behalf of patients. Normalizing these effort levels and their associated G(.) to zero, (3) and (4) would describe the marginal benefit and marginal cost of deviating from the professional "norm."

[17] Consider the effect of increasing d in (3) and (4): the marginal benefit (LHS) of both increases, so that the chosen effort levels both increase.

[18] Note, however, that the provider will choose how much effort to devote to improving amenities versus technical quality, according to patient responsiveness and reimbursement benefit; a higher fee per service could induce larger investments in amenities but little extra effort in technical quality improvement. More

increasing the payment margin is to encourage quality improvement and discourage cost control. Multitasking effects could dampen or even potentially overturn these direct effects.

Suppose that quality improvement and cost control are substitutes in the provider's effort function (Case 2). Then one multitasking effect arises because the two efforts compete for provider attention, and therefore a higher level of one makes the other more onerous. A higher level of quality effort will increase the marginal "work" disutility of cost reduction, reinforcing the direct effect of discouraging cost control effort.

A second multitasking effect arises if $t_{12} \neq 0$, and its direction will depend on whether the payment margin is positive (FFS) or negative (e.g., capitation). Suppose that $t_{12} > 0$, which means that the marginal treatment response to quality is higher for higher levels of cost control, or, alternatively, the marginal treatment reduction associated with cost control is less for higher levels of quality. Consider FFS (Case 2.2). Then a higher level of quality improvement will elicit a smaller demand response to cost control (a smaller negative number for $t_2 < 0$), which in turn boosts the marginal benefit of cost control effort and encourages cost control. This latter effect offsets the direct effect, which was to discourage cost control effort. If in contrast the payment margin is negative (Case 2.3), the original marginal financial benefit of quality effort is negative, because enhanced quality leads to more treatment expense coming out of the fixed (capitation) payment. Increasing the payment margin in this case reduces the degree of supply-side cost sharing. The direct effect is still to encourage quality enhancement and discourage cost control. But now the higher level of quality effort will also reduce the marginal benefit of cost control effort by eliciting a smaller demand response to cost control. This treatment-responsiveness multitasking effect now reinforces, rather than offsets, the direct effect and the cost-of-effort multitasking effect, which both discourage cost control.

In short, the overall impact of increasing the payment margin will depend not only on direct financial incentives but also on how the provider and patient respond to the incentive to reallocate efforts among the diverse tasks associated with quality health care.

---

generally, the actual levels of e_1 and e_2 will depend on which effort the provider finds least costly over what range of treatment and effort.

The single policy instrument of treatment-based payment cannot in general achieve multiple targets. Given the complexity of modern medicine, almost inevitably health care professionals will engage in a greater diversity of tasks than there are viable instruments for individually rewarding each task.

The policy implication is a simple plea for modesty: the search for the most sophisticated, "fine-tuned" incentive system, allowing use of high-powered incentives for all dimensions of wanted performance, is akin to the search for the Holy Grail. Ultimately we must rely to a greater or lesser extent on the professional discretion of providers to allocate their time and efforts according to the needs of each patient. Lower-powered treatment-based incentives (such as mixed payment or partial capitation) avoid extreme financial rewards for distorted effort allocations.

Additional instruments--paying separately for certain aspects of treatment, imposing minimum quality standards, fostering competition, etc.--may make achieving multiple targets easier. But each instrument or policy will also potentially introduce additional complications.[19] In the instruments-targets framework, professional ethics or provider fidelity to patient interests can be viewed as a potent instrument, in the sense that an appropriate "level" can help to achieve multiple targets simultaneously, with few adverse side-effects.[20]

The remainder of the paper will focus on one increasingly popular policy instrument in many health systems, fostering competition among providers. Such competition can help to induce provider efforts, but not all such rewarded efforts are desirable. For example, given some amount of supply-side cost sharing, providers benefit financially from competing to attract the healthy and avoid the sick, known as *risk selection*. The next section examines more directly the issue of how provider multitasking impacts the design of optimal payment systems (including optimal risk adjustment) when

[19] For example, see Finkelstein (2000) for empirical analysis of the adverse effects of minimum benefit regulation, focusing on the case of private health insurance for US elderly (Medigap).

[20] A provider who is a good "agent" for a patient may tend to over-indulge moral hazard, but this is counterbalanced by many socially desirable effects. The latter include less tendency to skimp on unprofitable patients or "creamskim" profitable ones, lower likelihood of exploiting information asymmetries through "supplier-induced demand", and greater incentive to provide high quality care (Eggleston 2000).

providers compete for patients. An optimal system seeks to encourage desired efforts, while mitigating incentive for distortions such as risk selection.

III. Competition

Competition among health care providers has often been advocated as a method for promoting efficiency, allowing consumer sovereignty in choosing providers, while simultaneously exerting pressure on providers to deliver quality care at least cost. Unfortunately, provider competition in the health sector is complicated by the limited ability of consumers and payers to monitor and contract upon various dimensions of provider effort or quality of care. As is well known from the theory of multitasking, rewarding one kind of effort will often lead to distortions in other kinds of effort (Holmstrom and Milgrom 1991). In health care the externalities associated with inefficient forms of provider competition, such as risk selection, can call into question the desirability of introducing strong market forces in this sector.

The following simple model illustrates this policy dilemma. The point is straightforward: whereas competition can create incentives to exert costly effort, not all such efforts are socially desirable. In particular, the fact that patients differ from one another can make it profitable for providers to shun the sickest, an incentive that is exacerbated by competition without risk adjustment. Indeed, an optimal risk adjustment mechanism itself must take into account the fact that providers exert multiple efforts, and any systematic profits or losses associated with imperfect risk adjusters will affect provider efforts.

To understand the interaction of payment incentives with competition, we will first abstract from the multiplicity of health services, patient types, and provider efforts. Consider a simple extension to the above model of provider choice of effort, focusing on a single dimension of effort e , such as quality improvement. Providers compete to attract patients, and higher quality attracts more patients, perhaps through enhanced reputation for quality care. Demand $D(e)$ is therefore an increasing function of effort:

$D'(e)$ >0. Abstract from any other demand response.[21] Exerting effort causes disutility $G(e)$, increasing in effort.

The provider's objective is to maximize total net revenue from a panel of patients, less effort costs: $D(e)[r + dt - G(e)]$. The first order condition for choice of effort balances the effort cost of quality improvement against the marginal benefit of attracting more patients, who are, at least in expectation or on average, profitable[22]:

$$D'(e)[r + dt - G(e)] = DG'(e) . \qquad (7)$$

This condition describing provider choice of effort can be re-written as

$$\frac{r + dt - G(e)}{G} = \frac{e_{G,e}}{e_{D,e}} , \qquad (8)$$

$$\text{where } e_{G,e} \equiv G_e \frac{e}{G} > 0 \quad \text{and} \quad e_{D,e} \equiv D_e \frac{e}{D} > 0 .$$

This shows that an increase in the elasticity of demand with respect to effort, $e_{D,e}$ (i.e., competition), leads to a decrease in the ratio of net payment to effort cost--the left hand side of (8), indicating that quality effort has increased. Intuitively, more competition elicits more patient-attracting effort.[23]

The responsiveness of effort with respect to the treatment-based payment margin is given by

$$\frac{de}{dd} = \frac{t}{\frac{e_{G,e}}{e_{D,e}} G_e + G_e} >0. \qquad (9)$$

Increasing the payment margin per service leads to increased effort, since that will attract more patients and therefore more revenue. The response to payment is larger for more

---

[21] In particular, assume that treatment for a given condition is contractible (e.g., guideline treatment), not chosen by the patient, and abstract from the fact that patients' inability to re-trade health services endows providers with the ability to set treatment away from guideline level (McGuire 2000).

[22] If discrimination among patients is possible, the provider will have incentive to offer zero discretionary quality to unprofitable patients; see Ma (1994) and Chung and Meltzer (2000).

[23] Compare McGuire 2000, equation (3.11'), p.489, and Chung and Meltzer 2000, equation (4), p.10.

complex and more potentially profitable services. An increase in the "work"--marginal effort cost--of quality, G_e , leads to less responsiveness of effort to payment incentives.

What we wish to focus on here is the interaction of competition with payment incentives. It can be shown that the responsiveness of quality effort to payment is increasing in the degree of competitiveness:

$$\frac{d}{de_{D,e}}\frac{de}{dd} = \frac{t e_{G,e} G_e \Big/ e_{D,e}{}^{2}}{\left[\frac{e_{G,e}}{e_{D,e}} G_e + G_e\right]^{2}} > 0. \qquad (10)$$

An increase in the competitive pressure on a provider increases that provider's patient-pleasing efforts and increases the provider's responsiveness to profitability of the patient.

This simple model focuses on a single quality effort. The result that competition reinforces financial incentives in effort choice, however, is more general. If the provider can target effort to attract specific kinds of patients, increased competitive pressure will increase the incentive to over-provide discretionary quality to profitable patients. Provider competition therefore exacerbates incentives of high-powered payment to discriminate against high cost (or low profitability) patients.[24]

These results also point to the importance of multitasking when analyzing the effects of payment incentives and competition on provider behavior. We can see from (9) that the responsiveness of effort choice to payment incentives is negatively impacted by an increase in the marginal disutility of effort, G_e . Suppose two kinds of efforts compete for provider time and concentration. Then increasing one will increase the marginal cost of the other. If patients are more responsive to one kind of effort than another (e.g., “patient-pleasing quality effort” vs. “upgrading technical skill”), then a provider has added financial incentive to substitute away from the latter and into the former. If such a re-allocation of provider effort is socially undesirable, lowering the intensity of financial incentives can mitigate this adverse effect of provider competition.

[24] Chung and Meltzer (2000) show a similar result, and use evidence from California hospital admissions in the 1983 to 1993 period to show that competition is associated with lower costs after the introduction of prospective payment, especially among patients with the highest costs.

Multitasking in Competing Managed Care Plans

To illustrate how multitasking influences the patterns of efforts that a provider will undertake when providing different services to heterogeneous patients, and what this implies about optimal risk adjustment, this section develops a multitasking extension of the managed care model of Frank, Glazer and McGuire (2000) [hereafter FGM]. The model describes consumer choice of health plan, and how plan management has incentive to respond by attempting to restrict access to services that attract unprofitable patients. The actual effectiveness in restricting access will depend upon clinician efforts. We generalize the FGM model to include professional ethics, supply-side cost sharing, effort costs, and a travel-cost parameter to capture market competitiveness.

Our extension of FGM can illustrate that (1) the higher the degree of supply-side cost sharing (**s**), the greater the incentive to distort service-specific qualities to select good risks (Newhouse 1996); (2) increased competitive pressures exacerbate selection incentives; (3) the greater the degree of “agency” on behalf of patients (a) or the higher the effort costs, G(q), the smaller the selection distortion; and (4) risk adjustment can be a powerful policy instrument for achieving targets of efficient provision without selection distortions, but should also take account of efficiency targets for provider efforts and multitasking. The FGM model emerges as a special case of that presented here, the case of a pure profit-maximizing provider paid flat capitation who can ration services efficiently at no resource or effort cost.

Following FGM, assume consumers value a health plan according to spending on care that they will receive if they join that plan and an individual-specific non-service component of utility, m_i , which we will interpret as the "distance" to the nearest competitor. Let m_{ij} represent the spending on health service j given to individual i if she joins the plan, and $v_{ij}(m_{ij})$ represent the increasing and concave utility individual i derives from that spending. Total service-related utility from joining the plan is then captured by $v_i(m_i) = \sum_j v_{ij}(m_{ij})$. Define **c**>0 as travel cost per unit distance, so that cm_i is the travel cost of individual i to the nearest competitor.[25]

Let $\bar{u}_i$ represent consumer i's utility from joining the next-preferred plan. Consumer i will choose the health plan if $v_i(m_i) > \bar{u}_i + cm_i$, or if $m_i > \frac{\bar{u}_i - v_i(m_i)}{c}$. Since the plan does not know each individual's m_i but does know the cumulative distribution from which it is drawn, $\Phi_i(m_i)$, the plan considers the probability that individual i will join it to be

$$n_i(m_i) = \text{prob}\left(m_i > \frac{\bar{u}_i - v_i(m_i)}{c} \right) = 1 - \Phi_i\left(\frac{\bar{u}_i - v_i(m_i)}{c} \right). \qquad (11)$$

Note that demand is increasing in the spending generosity of the plan: $\frac{dn_i}{dm_i} = \frac{dn_i}{dv_i}\frac{dv_i}{dm_i} = \frac{\Phi_i'}{c}\frac{dv_i}{dm_i} > 0$. A decrease in **c** corresponds to an increase in competition and a stronger demand response to changes in spending.

The plan provides various health care services, indexed by j. The model of managed care adopts the shadow price approach first used by Keeler et al (1998). The health plan sets a "shadow price" q_j for access to health service j such that "the patient must 'need' or benefit from services above a certain threshold in order to qualify for receipt of services" (FGM, p.836)[26]:

$$v_{ij}'(m_{ij}) = q_j . \qquad (12)$$

FGM assume that the provider chooses the shadow price directly, but then suggest that this is reflective of a more general framework in which there is a "division of

---

[25] The "travel cost" between two plans could also represent differences in consumer tastes and/or premium differentials across plans. A decrease in **c** would then capture increased competitiveness by lowering plan switching costs, which could take the form of lower premium differentials for similarly attractive benefit packages and less costly search, as well as lower travel costs.

[26] This framework is similar to a menu-setting approach in which a plan sets a menu of services, and a price of each service, to maximize a given objective function; see Olmstead and Zeckhauser (1999).

responsibility between the 'management'…and 'clinicians'": "cost-conscious management allocates a budget or a physical capacity for a service. Clinicians working in the service area do the best they can for patients given the budget by rationing care so that care goes to the patients that benefit most" (FGM, p.836). FGM abstract from any resource costs associated with implementing (enforcing) rationing, or any inefficiency from less-than-fully efficient rationing of services. Yet a provider must expend diagnostic, evaluation and management effort to allocate patients efficiently across rationed services. This effort is unlikely to be uniform across services.

Departing from FGM, therefore, we assume that implementing a shadow price q_j requires provider effort, which causes disutility $G(q_j; q_{j-1})$, increasing in the shadow price. G(q) represents any effort associated with "management" to reduce spending, e.g., from the excessive amount of spending preferred by fully insured patients. In other words, more stringent rationing cannot be implemented efficiently without increasing provider effort. The effort cost of rationing service j will also depend on the effort necessary to ration other services (q_{j-1}), just as in the multitasking model presented earlier.

There are several natural interpretations of G(q), the "effort cost of efficient rationing." For example, to implement a given set of shadow prices, the plan management must rely upon providers to exert effort to obtain information on patient-specific "needs" and use clinical judgment and experience in allocating patients in such a way so as to meet the plan's spending targets for different services. The relationship between budget stringency and effort needed to implement efficient rationing is unlikely to be linear. Acting as "gatekeeper" to services is relatively easy when the threshold for use of the service is low, but the incremental effort required to ration heterogeneous patients efficiently across services almost surely increases for tighter and tighter budgets (G">0).

An alternative interpretation of the effort cost associated with rationing would be the inefficiency resulting from imperfect allocation of patients across services. In the extreme, a "lazy" provider operating under a fixed budget could offer the same average level of spending on each service to all patients. This lack of effort would cause inefficiency from mismatching of services to patients and, for a given service, over-spending for less severe cases and under-spending for more severe cases.[27] Pollack and Zeckhauser (1996) examine how rational gatekeepers respond to the incentives of fixed budgets in a dynamic framework, showing that a doctor will provide access to a service as long as the discounted expected benefit to the patient exceeds the option value of a later referral to that service. Even homogenous, individually rational gatekeepers

> will fail to produce the social optimum for at least two reasons:
>
> - Gatekeepers will tend to hoard their budgets early on, for fear of running out later. So in early periods, care is not dispensed even when the benefit to patients exceeds the social cost.
> - Individually-optimal strategies produce socially wasteful "spend-downs" toward the end of each budget cycle (Pollack and Zeckhauser 1996: 650).

The opposite may occur if providers consider their budget constraint "soft.". A provider could allow all patients access at a relatively low threshold early in the budget period (e.g., at the beginning of the fiscal year). The provider would then either have to deny access to patients towards the end of the period, or secure an increase in the budget to cover end-of-period cases. The shadow price framework of FGM implicitly assumes that the plan can credibly commit to a hard budget constraint (Kornai 1980), and will not give in *ex post* to requests for more expenditure, even if severely ill patients are denied vital services.

Avoiding these dynamic inefficiencies of fixed budgets is likely to require even more effort if providers are innately heterogeneous. Differences among clinicians in the estimation of who would benefit from a service could stem from practice style differences, which are known to be significant. Some degree of provider profiling and/or peer feedback and monitoring is likely to be needed to enforce consistent standards for threshold "need" before patients are granted access to specific services. Such profiling and monitoring systems involve resource costs.

---

[27] Ma (2000) analyzes the effects of public-sector rationing on incentives for private-sector cost reduction. In Ma's model, efficient rationing improves private sector cost incentives, but random rationing does not. Although his model abstracts from effort costs of efficient rationing, the distinction between random rationing and efficient rationing is similar to the argument that efficient rationing requires provider effort. If a provider simply allocates patients access to a service randomly, by giving each patient in a given category the same probability of receiving the service, then the provider conserves diagnostic and patient allocation effort but undermines efficient allocation of the budget for that service.

Moreover, when individual clinicians are called upon to ration care by allocating patients to treatments, they have power over patients. Note that the required allocation is according to a patient's capacity to benefit, which may correspond only imperfectly, if at all, to a patient's willingness and ability to pay for additional services. Indeed, the idea of efficient rationing to constrain *ex post* moral hazard requires some conflict between the provider and patient (Ellis and McGuire 1990). This creates an incentive for patients to voice their different willingness to pay through under-the-table payments directly to the relevant provider(s). Although such payments do not seem to be much of a factor in US managed care, they are pervasive in many systems that restrict patient choice, particularly in developing and transitional economies. Frequently patients who have personal connections to clinicians receive preferred access. Some monitoring of provider behavior may be necessary to combat these tendencies, and increasingly so the more stringent the rationing to be implemented.

These interpretations of the effort cost of efficient rationing are not mutually exclusive. The effort costs G(q), assumed to be increasing and convex, could be some combination of all the above factors.

In addition to taking account of these direct effort costs of rationing, health plans will also respond to financial incentives and consumer preferences when choosing a pattern of shadow prices for various health services. Payment for each patient takes the form of a fixed pre-payment (capitation) $\mathbf{r_i}$ plus reimbursement of $(1-s_j)m_{ij}$ for each service j . The health plan therefore is at risk for the proportion of spending $s_j m_{ij}$, and **s** denotes the degree of supply-side cost sharing, potentially different for each kind of service. A fully capitated plan would receive a positive $\mathbf{r_i}$ per enrollee and be fully liable for costs of care: $s_j = 1 \forall j$. If capitation payments are risk adjusted, $\mathbf{r_i}$ will differ according to the *risk adjusters* (such as age, sex, and diagnoses) of individual i included in the risk adjustment mechanism.

Expected net revenues will depend upon demand for services, the degree of supply-side cost sharing for each service and the effort cost associated with efficient rationing:

$$p(q) = \sum_i n_i(q)\left[r_i - \sum_j s_j m_{ij}(q_j) - G(q_j; q_{j-1}) \right]. \qquad (13)$$

Define $p_i(q)$ as the plan's gain or loss for individual i, $p_i(q) = r_i - \sum_j s_j m_{ij}(q_j) - G(q)$. Unprofitable enrollees are those for which $p_i < 0$. Which consumers are unprofitable will depend on the level of the (possibly risk adjusted) payment, the degree of supply-side cost sharing, and the effort required to ration services efficiently for those patients. The health plan can discourage unprofitable consumers from enrolling in the plan by rationing services valued by those consumers more stringently. If the health plan is at risk for some of the costs of care for service j ($s_j > 0$), the plan will want to balance the marginal financial benefit of increasing the shadow price ($-s_j m_{ij}' > 0$), which reduces spending, against the associated effort cost and the possibility of discouraging enrollment of profitable consumers. A plan whose providers care about patient benefits as well as profits will need to take account of this "agency" relationship as well when endeavoring to ration services.

Let the degree of provider "agency" on behalf of patients, or fidelity to patient interests, be denoted by a. Health plan management will need to take clinicians' agency on behalf of patients into account when setting shadow prices (e.g., in order to recruit and retain quality clinicians). The health plan therefore maximizes an objective function that includes not only expected profits but also agency weight a on patient valuation of treatment benefits:

$$EV = \sum_i n_i(q)\left[r_i - \sum_j s_j m_{ij}(q_j) - G(q) + \sum_j a v_{ij}(m_{ij}) \right] = \sum_i n_i \left[p_i + a v_i \right]. \qquad (14)$$

FGM analyze the case of a pure profit maximizer ($a = 0$) paid on a flat capitation basis ($s = 1$) who can ration efficiently at no effort cost ($G(q) \equiv 0$). In that case, (14) reduces to $p(q(s = 1))$.

The health plan chooses the shadow price for each service j to maximize (14):

$$\frac{dEV}{dq_j} = \sum_i \left[\left(\frac{dn_i}{dq_j} \right) \{ p_i + a v_i \} + n_i \left\{ \frac{dp_i}{dq_j} + a \frac{dv_i}{dq_j} \right\} \right] = 0 \text{, or}$$

$$\sum_i \left[\left(\frac{dn_i}{dq_j} \right) p_i - n_i s_j m_{ij}' \right] = \sum_i \left[n_i G_{q_j} - a \left\{ \frac{dn_i}{dq_j} v_i + n_i \frac{dv_i}{dq_j} \right\} \right]. \qquad (15)$$

This first order condition describes the trade-off a plan makes in setting the shadow price for each service. The marginal benefit to the plan of raising the shadow price (the left hand side of (15)) includes discouraging unprofitable patients from enrolling ($\left(\frac{dn_i}{dq_j} \right) p_i > 0$ if $p_i < 0$) and less spending per enrollee. The marginal cost of more stringent rationing includes provider effort cost, foregone agency benefits, and possibly discouraging profitable patients from joining the plan ($\left(\frac{dn_i}{dq_j} \right) p_i < 0$ if $p_i > 0$). As FGM note, "the idea behind competition among managed care plans is that … the plan by rationing too tightly will lose profitable customers -- to balance the plan's incentive to reduce services to the existing enrollees" (p.838). Agency on behalf of patients discourages selection by increasing the marginal cost of stringent rationing.

Shadow prices can exceed or fall short of the socially optimal value, which would equate marginal benefit to social marginal cost. Note that if the payment system does not include any supply-side cost sharing ($s_j \leq 0$), then plans will not want to restrict access to services. This is consistent with a low threshold for use, and possibly wasteful over-use, under FFS.

For simplicity, assume the elasticity of spending with respect to the shadow price threshold for service j is constant across individuals: $e_j = m_{ij}' \frac{q_j}{m_{ij}}$. FGM show that, for the special case of a pure profit maximizer ($a = 0$) paid on a flat capitation basis ($s = 1$) who can ration efficiently at no effort cost ($G(q) \equiv 0$) and **c**=1, the shadow price can be expressed as

$$q_j = \frac{\sum_i n_i m_{ij}}{\sum_i \left[\Phi_i' m_{ij} p_i \right]}. \qquad (16)$$

Generalizing to include professional ethics, supply-side cost sharing, effort costs, and a travel-cost parameter to capture market competitiveness, we find that

$$q_j = \frac{s_j \sum_i n_i m_{ij}}{\sum_i \left[\frac{\Phi_i'}{c} m_{ij} \{p_i + a v_i\} + n_i \left\{ a m_{ij} - \frac{G_{q_j}}{e_j} \right\} \right]} \qquad \text{for } 0 \le s_j \le 1. \qquad (16')$$

The numerator of (16') represents the incentive the plan has to save money on expected enrollees. The higher the degree of supply-side cost sharing, the greater this incentive and the higher the shadow price. This corresponds to the finding in the literature (e.g., Newhouse 1996) that increasing supply-side cost sharing increases the incentives to invest in risk selection. In this case risk selection is implemented through stringent (lax) rationing--high (low) shadow prices--for services most (least) costly for the plan to provide, i.e., services valued most by unprofitable (profitable) consumers.

The denominator of (16') represents the costs associated with rationing. These include effort costs of efficiently allocating patients among the restricted services, and the expected financial and professional benefits that a plan sacrifices by losing enrollees. If the effort cost of efficient rationing (G_{q_j}) increases, rationing is less stringent (q_j declines). Higher agency on behalf of patients--greater provider fidelity to patients--increases the forgone gains associated with quality distortions and therefore decreases service-specific shadow price distortions. In other words, professional ethics mitigate risk selection. Note that when there is positive agency, there is a positive marginal cost of increasing shadow prices even when there is no explicit demand response from consumers ($\frac{\Phi_i'}{c}$=0) and no effort cost of rationing ($G(q) \equiv 0$). For example, (16') shows that even if--in the absence of any demand-response discipline--plans were paid flat capitation (with the highest incentive to risk select, **s**=1), there would still be little or no distortion of shadow prices if provider fidelity to patient interests was sufficiently strong (e.g., $a = 1$).

Health plan rationing will also respond to competitive pressures. As the cost of traveling to the nearest competitor (**c**) declines, the demand response to rationing ($\frac{\Phi_i'}{c}$)

increases. This increased competitive pressure will increase the sensitivity of shadow prices to the profitability of patients, so that service-specific shadow prices increase (decrease) for services valued by unprofitable (profitable) consumers. This captures the impact of increased competition on providers' incentives to skimp on care for unprofitable patients and "cream" profitable patients through over-provision of services they value (Ellis 1998). The selection-exacerbating effect of competition is reinforced the greater the degree of supply-side cost sharing, since the marginal gain to skimping and creaming is higher when the provider is at risk for a larger percentage of spending. This corresponds to the finding of (10) above in the single-effort, single-service model. Accordingly, an effective instrument for decreasing plan incentives to ration is to decrease the degree of supply-side cost sharing for competing health plans. Reducing supply-side cost sharing, however, will also discourage cost-control effort (Newhouse 1996). Indeed, as noted previously, a plan that is not at risk for spending, or receives a positive profit margin under FFS, will not wish to discourage use of services and may encourage wasteful over-use.

Risk adjustment of capitation payments can be a potent additional policy instrument for mitigating selection incentives. We can see from the above description of shadow prices that service-specific distortions are large when the dispersion of profitability of patients is great. Consider the effect of increasing the capitation payment r_i for an originally unprofitable patient. This better match of prepayment to expected cost will increase p_i (to a smaller negative number) and decrease the marginal benefit to the plan of stringent rationing for that consumer. Similarly, a lower capitation payment for originally profitable consumers will decrease the incentive for "creaming" through very low shadow prices for services those consumers value.

Glazer and McGuire (2000) use (16) to illustrate minimum variance optimal risk adjustment. The above analysis suggests that an additional consideration in deriving optimal risk adjustment mechanisms should be the efficiency target(s) of appropriate provider effort(s). In general, risk adjusters will be imperfect, implying that risk adjusted payments will leave providers with positive and negative profit margins for different patients. These over-and under-payments will affect provider efforts for appropriate quality improvement and cost control as well as "pure" incentives to select, so that expected cost conditional on a risk-adjuster signal (i.e., conventional risk adjustment) is unlikely to represent the optimal risk adjustment mechanism.[28] This result is similar in spirit to that derived by Glazer and McGuire (1997, 1999) and Encinosa (2000).

In sum, the incentives for quality distortions to select good risks will be most acute for competing providers paid according to capitation or prospective payment, and risk adjustment can ameliorate these incentives. Therefore, risk adjustment can be a powerful policy instrument for achieving targets of efficient provision without selection distortions. But the adjustment mechanism should also take account of efficiency targets for provider efforts and multitasking. Moreover, given the multiple goals of an appropriate payment system (e.g., low cost, high quality, zero selection distortions, appropriate preventive effort, etc.), the single policy instrument of risk adjustment is unlikely to be a panacea. An additional beneficial instrument is lower-powered supply-side cost sharing (i.e., partial capitation), possibly differing by service and/or patient characteristics, such as separate FFS payments for specific procedures and/or outlier payments and high-risk pools for designated high-cost patients (van Barnevald et al. 1996, Encinosa 2000). Quality assurance programs can also be helpful, as long as policymakers bear in mind that for every measurable dimension of quality there almost inevitably is another dimension that is fundamentally noncontractible. Appropriate performance compensation should take provider multitasking into account.

IV. Conclusion

Any health care professional can tell you that the daily practice of modern medicine requires attention to diverse "tasks". Yet the theory of provider payment has largely overlooked the question of how incentives impact providers' decisions to allocate effort across tasks that may substitute for or complement formal treatment. This paper extends the theory of provider payment to take account of health care provider "multitasking," focusing on how competition among providers interacts with payment

---

[28] I am indebted to a referee for suggesting this risk adjustment application of the multitasking framework.

system incentives when the allocation of provider effort among multiple tasks is noncontractible.

A simple two-effort model of health care provider multitasking illustrates that reliance on a single policy instrument--treatment-based payment--cannot in general give incentive for appropriate levels of multiple provider efforts. The extension to a more nuanced model of managed care reinforces this result, and shows that competitive pressure increases incentive for providers to exert patient-pleasing efforts and to be responsive to the profitability of services and patients when making treatment decisions. These competitive pressures may help to control costs, but also reward service-specific quality distortions to attract the healthy and eschew the sick.

The policy implication is a simple plea for modesty: the search for the most sophisticated, "fine-tuned" incentive system, with high-powered incentives for all dimensions of wanted performance, is akin to the search for the Holy Grail. Treatment-based payment can be supplemented by additional instruments--imposing minimum quality standards, risk adjusting prospective payments, fostering reputational effects, etc.--but almost inevitably each instrument introduces distortions of its own. Moreover, given the complexity of modern medicine and the fundamental noncontractibility of many aspects of quality care, health care professionals engage in a greater diversity of tasks than there are viable instruments for individually rewarding each task.

Ultimately we must rely to a greater or lesser extent on the professional discretion of providers to allocate their time and efforts according to patient needs. Provider professional ethics can be a powerful policy instrument for achieving multiple efficiency targets. Low-powered treatment-based incentives (such as mixed payment or partial capitation) also seem to be relatively robust instruments for avoiding extreme financial rewards for distorted effort allocations. Optimal provider payment systems, including optimal risk adjustment, should take account of provider multitasking.

References

Arrow, K.J., 1963. Uncertainty and the welfare economics of medical care. American Economic Review 53, 941-973.

Chung, J., and D. Meltzer, 2000. Effects of Competition under Prospective Payment on Hospital Costs among High and Low Cost Admissions: Evidence from California, 1983-1993. Forthcoming, Rand Journal of Economics.

Cohen, A., 1999. A note on optimal health insurance and provider payment. Mimeo, Harvard University Department of Economics.

Conviser, R., S. Gamliel, and L. Honberg, 1998. Health-based payment for HIV/AIDS in Medicaid managed care programs. Health Care Financing Review 19(3), 63-82.

Cutler, D.M., and R.J. Zeckhauser, 2000. The anatomy of health insurance. Chapter 11 in A.J. Culyer and J.P. Newhouse, eds., Handbook of Health Economics, Vol. 1A (Amsterdam: Elsevier North-Holland), 563-643.

Eggleston, K., 2000. Risk Selection and Optimal Payment Systems for Medical Services. Journal of Risk and Insurance 67(2): 173-196.

Ellis, R., 1998. Creaming, skimping and dumping: Provider competition on the intensive and extensive margins. Journal of Health Economics 17, 537-555.

Ellis, R., and T.G. McGuire, 1990. Optimal payment systems for health services. Journal of Health Economics 9: 375-396.

Encinosa, W., 2000. The Theory of Risk Adjusting Health Care Payments: Firm Organization and Specialization. Mimeo.

Finkelstein, A., 2000. Minimum standards and health insurance regulation: The case of Medigap. Mimeo.

Frank, R. G., J. Glazer, and T. G. McGuire, 2000. Measuring Adverse Selection in Managed Health Care. Journal of Health Economics 19 (2000), 829-854.

Glazer, J., and T.G. McGuire, 1997. Optimal risk adjustment in markets with adverse selection: An application to managed care. American Economic Review 90(4), 1055-1071.

Glazer, J., and T. G. McGuire, 1999, "Minimum Variance Optimal Risk Adjustment," mimeo.

Hart, O., A. Shleifer, and R. Vishny, 1997. The Proper Scope of Government: Theory and an Application to Prisons. Quarterly Journal of Economics 112(4): 1127-1161.

Holmstrom, B. and P. Milgrom, 1991. Multitask principal-agent analyses: Incentive contracts, asset ownership, and job design. Journal of Law, Economics and Organization 7 (1991 Special Issue), 24-52.

Hsiao, W. C., P. Braun, D. Dunn, E. Becker, M. DeNicola, and T. Ketcham, 1988a. Results and Policy Implications of the Resource-Based Relative Value Study. The New England Journal of Medicine 319 (September 29,1988), 881-888.

Hsiao, W. C., P. Braun, D. Dunn, and E. Becker, 1988b. Resource-Based Relative Values: An Overview. Journal of the American Medical Association 260(16) (October 28, 1988), 2347-2353.

Hsiao, W. C., D. B. Yntema, P. Braun, D. Dunn, and C. Spencer, 1988c. Measurement and Analysis of Intraservice Work. Journal of the American Medical Association 260(16) (October 28, 1988), 2361-2370.

Kornai, J., 1980. *Economics of Shortage*. Amsterdam: North Holland.

Lewis, T.R., and D.E.M. Sappington, 1999. Using subjective risk adjusting to prevent patient dumping in the health care industry. Journal of Economics & Management Strategy 8(3), 351-382.

Ma, C.A., 1994. Health care payment systems: Cost and quality incentives. Journal of Economics & Management Strategy 3(1), 93-112.

Ma, C.A., 2000. Public rationing and private cost incentives. Mimeo.

Ma, C.A., and T.G. McGuire, 1997. Optimal health insurance and provider payment. American Economic Review 87: 685-704.

McGuire, T. G., 2000. Physician Agency. Chapter 9 in A.J. Culyer and J.P. Newhouse, eds., Handbook of Health Economics, Vol. 1A (Amsterdam: Elsevier North-Holland), 461-536.

Newhouse, J.P., 1996. Reimbursing health plans and health providers: Selection versus efficiency in production. Journal of Economic Literature 34, 1236-1263.

_____, 1998. Risk adjustment: Where are we now? Inquiry 35, 122-131.

Olmstead, T., and R. J. Zeckhauser, 1999. The Menu-setting Problem and Subsidized Prices: Drug Formulary Illustration. Journal of Health Economics 18 (1999), 523-550.

Pollack, H., and R. J. Zeckhauser, 1996. Budgets as Dynamic Gatekeepers. Management Science 42(5), 642-658.

Prendergast, C., 1999. The provision of incentives in firms. Journal of Economic Literature 37, 7-63.

Saltman, R.B., and J. Figueras, eds., 1997. European Health Care Reform: Analysis of Current Strategies. World Health Organization Regional Office for Europe, Copenhagen 1997.

Selden, T. M., 1990. A model of capitation. Journal of Health Economics 9, 397-409.

Tinbergen, Jan, 1991. The Functioning of Economic Research. Journal of Economic Issues 25(1) (March 1991), 33-38.

Van Barnevald, E., R. van Vliet, and W. van de Ven, 1996. Mandatory high-risk pooling: An approach to reducing incentives for cream skimming. Inquiry 33, 133-143.

van de Ven, W.P.M.M., and R.P. Ellis, 1999. Risk adjustment in competitive health plan markets. Chapter 14 in A.J. Culyer and J.P. Newhouse, eds., Handbook of Health Economics, Vol. 1A (Amsterdam: Elsevier North-Holland), 755-845.

Zeckhauser, R.J., 1970. Medical insurance: A case study of the tradeoff between risk spreading and appropriate incentives. Journal of Economic Theory 2(1), 10-26.

Table 1. Comparative Statics for the Two-Effort Model

Case	Assumptions						Results	
	t_1	t_2	G_{12}	t_{12}	d	$(dt_{12} - G_{12})$	$\frac{\partial e_1}{\partial d}$	$\frac{\partial e_2}{\partial d}$
1.1	+	+	-	0	n.a.	n.a.	+	+
1.2				+	+	n.a.	+	+
1.3					-	+/-	+/?	+/?
1.4				-	+	+/-	+/?	+/?
1.5					-	n.a.	+	+
2.1	+	-	+	0	n.a.	n.a.	+	-
2.2				+	+	+/-	?/+	?/-
2.3					-	n.a.	+	-
2.4				-	+	n.a.	+	-
2.5					-	+/-	?/+	?/-
3.1	-	-	-	0	n.a.	n.a.	-	-
3.2				+	+	n.a.	-	-
3.3					-	+/-	-/?	-/?
3.4				-	+	+/-	-/?	-/?
3.5					-	n.a.	-	-
4.1	+	+	+	0	n.a.	n.a.	?	?
4.2				+	+	+/-	+/?	+/?
4.3					-	n.a.	?	?
4.4				-	+	n.a.	?	?
4.5					-	+/-	+/?	+/?
5.1	+	-	-	0	n.a.	n.a.	?	?
5.2				+	+	n.a.	?	?
5.3					-	+/-	?/+	?/-
5.4				-	+	+/-	?/+	?/-
5.5					-	n.a.	?	?
6.1	-	-	+	0	n.a.	n.a.	?	?
6.2				+	+	+/-	-/?	-/?
6.3					-	n.a.	?	?
6.4				-	+	n.a.	?	?
6.5					-	+/-	-/?	-/?

n.a. = not applicable: no assumption on this term is needed to sign the comparative statics.

Table 2. The "Direct" and "Multitasking" Effects of an Increase in the Payment Margin

Case / Example	Direct Effect			Multitasking Effect		Reinforces or offsets the direct effect?	Applies to sub-cases
Case 1	$d\uparrow\Rightarrow$	$dt_1\uparrow\rightarrow e_1\uparrow$	$\Rightarrow$	$G_{12}<0: G_2\downarrow\rightarrow e_2\uparrow$		Reinforces	1.1 - 1.5
$e_1=$ Technical quality improvement;				$t_{12}>0: t_2\uparrow\Rightarrow$	$e_2\uparrow if\,(d>0)$	Reinforces	1.2
$e_2=$ Improvement of amenities					$e_2\downarrow if\,(d<0)$	Offsets	1.3
				Or $t_{12}<0: t_2\downarrow\Rightarrow$	$e_2\downarrow if\,(d>0)$	Offsets	1.4
					$e_2\uparrow if\,(d<0)$	Reinforces	1.5
		$dt_2\uparrow\rightarrow e_2\uparrow$		Perfectly symmetric with above, with reverse of subscripts			
Case 2	$d\uparrow\Rightarrow$	$dt_1\uparrow\rightarrow e_1\uparrow$	$\Rightarrow$	$G_{12}>0: G_2\uparrow\rightarrow e_2\downarrow$		Reinforces	2.1 - 2.5
$e_1=$ Quality improvement;				$t_{12}>0: t_2\uparrow$ (less negative) $\Rightarrow$	$e_2\uparrow if\,(d>0)$	Offsets	2.2
$e_2=$ Cost control effort					$e_2\downarrow if\,(d<0)$	Reinforces	2.3
				Or $t_{12}<0: t_2\downarrow$ (more negative) $\Rightarrow$	$e_2\downarrow if\,(d>0)$	Reinforces	2.4
					$e_2\uparrow if\,(d<0)$	Offsets	2.5
		$dt_2\downarrow\rightarrow e_2\downarrow$	$\Rightarrow$	$G_{12}>0: G_1\downarrow\rightarrow e_1\uparrow$		Reinforces	2.1 - 2.5
				$t_{12}>0: t_1\downarrow\Rightarrow$	$e_1\downarrow if\,(d>0)$	Offsets	2.2
					$e_1\uparrow if\,(d<0)$	Reinforces	2.3
				Or $t_{12}<0: t_1\uparrow\Rightarrow$	$e_1\uparrow if\,(d>0)$	Reinforces	2.4
					$e_1\downarrow if\,(d<0)$	Offsets	2.5

TUFTS UNIVERSITY
ECONOMICS DEPARTMENT

DISCUSSION PAPERS SERIES 2001

2001-01 Karen EGGLESTON; Multitasking, Competition and Provider Payment

Discussion Papers Are Available Online at http://ase.tufts.edu/econ/papers