

The Design of Incentives for Health Care Providers in Developing Countries

Contracts, Competition, and Cost Control

Jeffrey S. Hammer

William G. Jack

Whatever the theoretical attractiveness of certain policy options, the fact that public employees are people who make independent decisions about their careers and lifestyles can set bounds on how well government agencies can deliver promised services, such as universal health care, including in rural areas.



Summary findings

Hammer and Jack examine the design and limitations of incentives for health care providers to serve in rural areas in developing countries. Governments face two problems: it is costly to compensate well-trained urban physicians enough to relocate to rural areas, and it is difficult to ensure quality care when monitoring performance is costly or impossible.

The goal of providing universal primary health care has been hard to meet, in part because of the difficulty of staffing rural medical posts with conscientious caregivers. The problem is providing physicians with incentives at a reasonable cost. Governments are often unable to purchase medical services of adequate quality even from civil servants.

Using simple microeconomic models of contracts and competition, Hammer and Jack examine questions about:

- The design of rural service requirements and options for newly trained physicians.

- The impact of local competition on the desirable level of training for new doctors.

- The incentive power that can be reasonably expected from explicit contracts.

One problem a government faces is choosing how much training to give physicians it wants to send to rural areas. Training is costly, and a physician relocated to the countryside is outside the government's direct control. Should rural doctors face a ceiling on the prices they charge patients? Can it be enforced?

Hammer and Jack discuss factors to consider in determining how to pay rural medical workers but conclude that we might have to set realistic bounds on our expectations about delivering certain kinds of services. If we can identify reasons why the best that can be expected is not particularly good, it might lead us to explore entirely different policy systems. Maybe it is too hard to run certain decentralized systems. Maybe we should focus on less ambitious but more readily achievable goals, such as providing basic infrastructure.

This paper—a product of Public Economics, Development Research Group—is part of a larger effort in the group to analyze service delivery in the social sectors. Copies of the paper are available free from the World Bank, 1818 H Street NW, Washington, DC 20433. Please contact Hedy Sladovich, room MC2-609, telephone 202-473-7698, fax 202-522-1154, email address hsladovich@worldbank.org. Policy Research Working Papers are also posted on the Web at <http://econ.worldbank.org>. The authors may be contacted at jhammer@worldbank.org or wgj@georgetown.edu. February 2001. (15 pages)

The Policy Research Working Paper Series disseminates the findings of work in progress to encourage the exchange of ideas about development issues. An objective of the series is to get the findings out quickly, even if the presentations are less than fully polished. The papers carry the names of the authors and should be cited accordingly. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the view of the World Bank, its Executive Directors, or the countries they represent.

The Design of Incentives for Health Care Providers in Developing Countries: Contracts, Competition, and Cost-Control

Jeffrey S. Hammer

Development Economics Research Group

The World Bank

jhammer@worldbank.org

William G. Jack

Department of Economics, Georgetown University

Centre for New and Emerging Markets, London Business School

wgj@georgetown.edu

Keywords: provider incentives, contract theory, rural health care

1. Introduction

The purpose of this paper is to examine the problem of ensuring health-care coverage of rural and poor areas of developing countries. We focus primarily on the incentives facing medical service providers and analyze them in light of recent developments in contract theory. The approach is essentially theoretic but is motivated by experiences in several countries in Asia, Eastern Europe and Southern Africa.

Since 1977 the international health community has put a great deal of emphasis on ensuring universal access to basic primary care as a high priority for public action. While the rationale for this emphasis is questionable,¹ we take this goal as given. In practice, attempts to accomplish this goal have often been disappointing. Reliance on the private sector to put trained professionals in such areas is not warranted as demand for the services at the true opportunity cost of the professional's time is simply too low. State intervention is obviously needed. On the other hand, public performance has met with, at best, variable success. Disappointment stems from several sources but a common problem is the inability to staff and supply medical posts in rural areas. High rates of absenteeism (blatant or couched in terms of attending meetings, etc.), high rates of vacancies for postings, simple lack of conscientious or courteous care and frequent lack of supplies such as essential drugs are common in many public facilities.

Why is it so hard to run public clinics? We usually assume in the analysis of public expenditures that if the government wants a particular product or service, it simply pays for it and the product appears. We can debate whether these services are appropriate, but there is no ambiguity as to what is, in fact, purchased.² Here, however, we seem to face an essential inability to purchase a particular service at all. Apparently it is difficult to purchase medical services of agreed quality (including due diligence of effort, sufficient attendance, etc.) even from civil servants.

These problems with public provision do not necessarily argue in favor of abandoning the goals of basic health care coverage for the poor but they do highlight the fact that the costs of provision could be much higher than anticipated. The underlying problem is one of providing physicians with incentives at reasonable cost. The rest of the paper examines a number of theoretic models that attempt to solve aspects of this problem. For most of the paper we will use the principal-agent framework in which the government (principal) provides incentives to an agent (physician) through an explicit contractual arrangement, the terms of which will depend on, *inter alia*, the observability of actions and outcomes. We will be particularly interested in the role of "high powered"

¹ See Filmer et al. for a critique of this approach to health care policy.

² This is the essence of cost/benefit analysis or project evaluation – quantities are usually assumed to be a technological given while intellectual attention focussed entirely on their valuation.

incentives and their limitations. Some comments on the provision of incentives when contracts cannot be written or enforced conclude the paper.

As a point of departure, let us assume that the government can actually see and contract for all relevant dimensions of health care. In this case, there is no incentive problem per se as it is possible to ensure that the wage at which the universality objective is met will yield the expected services. The problem is that this level of wages may be very high – much higher than is likely to be forthcoming in any practical sense. One reason is the existence of greater earnings opportunities in richer, more urban areas of the country. On top of this, of course, is the fact that medical personnel are generally much better educated and potentially have higher incomes than the country average and tend to want the amenities that go along with urban life. If they have families they will tend to want to give their children educational opportunities usually found in cities. A recent survey of medical personnel in Indonesia discovered that the amount of money necessary to get them to relocate to the more remote areas of the country (generally away from Java) was several multiples of actual wages.³ At first glance, there does not appear to be a problem of ineffective incentives, but one of insufficient budget. There could well be good, economic, reasons for insufficient funds. If taxes to pay for government expenditures are distortionary, and most tax systems in poor countries are very distortionary indeed, then there is an additional social cost of meeting the universality objective. This gives a good reason to see if there is a way to reduce the overall cost of provision by exploiting the design of contracts to doctors (see section 2).

At the other extreme, we can assume that the government can observe nothing about physician performance – that is, it can't tell if there is a doctor in attendance in a village at all much less if she is charging (illegally if a public facility) or providing conscientious care. This is something of a straw man but does raise one point. In this case, a publicly employed doctor is indistinguishable from a completely unregulated private doctor (except that the public employee costs the government more). Except in idiosyncratic cases (such as a dedicated altruist or a doctor wanting to live in her home village regardless of income) there is no reason to believe any of the poor areas will be covered. Not only will it be costly to induce urban doctors to move (the constraint of the preceding paragraph), but without performance incentives the real cost of services will escalate dramatically. One is then left to wonder where, given the inherently high cost, the desirability of universal provision comes from.

Intermediate within these extremes are cases in which some sort of information is available or can be obtained at some cost, say by improved monitoring. In such cases we will look at the tradeoff between improving performance with incentives and controlling costs imposed by different information structures. Of course, incentives can be provided

³ Chomitz et al. (1998). Interestingly, they did find that promises of placement in graduate education would induce doctors to relocate. However, this seems to depend on a highly distorted, non-transparent system of graduate admissions, leverage that would disappear should medical education be reformed.

by people other than government bureaucrats such as competing providers (traditional healers), consumers themselves (through complaints or willingness to pay additional fees – whether legal or not) and future employers (through reputation development or career concerns of the doctor).

In the next section we abstract from performance incentives and consider the optimal policy that induces physicians to practice in rural areas, recognizing that opportunity costs differ across providers. Section 3 considers the role of competitors and consumers in providing performance incentives for rural doctors, and the implications for training policies. Section 4 reviews the lessons that can be learned from standard contract theory when explicit incentives can be provided, and section 5 concludes.

2. Satisfying the Participation Constraint: Getting Doctors to Serve in Rural Areas

This section examines how a government might pay physicians to serve rural areas when (i) there are fixed costs associated with relocation (including amenity costs), and (ii) some physicians are more willing to move than others. Physicians clearly need to be compensated for the costs of moving, so by itself, the fixed cost component does not add anything of much interest to this question. But when medical care providers differ in their willingness to relocate - for example, due to differences in underlying preferences or opportunity costs - the treatment of fixed costs can be used to minimize total financial costs incurred by the government. (We assume in this section that there is no monitoring problem once physicians have relocated to rural areas.)

We measure the quantity of services delivered by a physician by a variable q - this could be the number of years she is resident in a rural community, or equivalently the number of (potential) patients to whom she provides services. The government attaches some welfare value to the delivery of services, and for simplicity, we shall assume this is constant at the margin. That is, the social value of q is just bq , for some $b > 0$. This can be motivated by an assumption that the under-supply of rural physicians is very large relative to the availability of new graduates (so that decreasing returns are yet to set in).

Physicians incur a fixed cost R when they relocate, and an increasing and convex variable cost $c(q)$ once in a rural community. If t is the total monetary payment made to the physician, then her net utility is when providing q is simply

$$u(t,q) = t - [R + c(q)].$$

We shall assume for now that the fixed cost of relocation, R , is common to all physicians, but that they may differ in the variable cost component. For concreteness, let us assume that

$$c(q) = \theta q^2/2,$$

and that physicians vary in their values of θ . In particular, a proportion ϕ of the available physician pool has cost parameter θ_L , and a proportion $(1-\phi)$ has cost parameter θ_H .

The government's objective is to maximize the social benefit of medical care less the financial costs of delivery (i.e., the costs of paying physicians).⁴ If both types of physician are fully employed, welfare is

$$W = \phi(bq_L - t_L) + (1-\phi)(bq_H - t_H)$$

where q_i is the quantity of services provided by a physician of type i . Note that we do not include the utility of the physicians in this measure of welfare, motivated by the assumption that public income is more valuable than (private) physician income due to the existence of tax distortions. The welfare of physicians is important however, as their pursuit of it generates incentive effects.

Pricing policies

If the government paid a constant rate b per service (e.g., b could be the annual salary), then services would be provided up to the point at which the marginal cost to the physician matched the marginal social benefit. But social welfare would be very low, $W = 0$. A large surplus is generated by this *uniform pricing policy*, but since the surplus accrues to physicians, whose welfare receives no weight in social welfare, the government can potentially improve on matters.⁵

With a single type of physician, with cost parameter θ , the optimal pricing policy is to pay physicians the uniform marginal price b and to charge them an amount $S(b) = b^2/2\theta$, less the relocation costs R , to enter the rural health service. $S(b)$ is the gross surplus earned by a physician of type θ earning b per unit of service provision: i.e.,

$$S(b) = \max bq - \theta q^2/2 = b^2/2\theta.$$

In the event that the surplus is smaller than the relocation cost, the fixed payment would be a transfer from the government to the physician, but would not be large enough to cover the whole relocation cost.

When there are two physician types, then different fixed fees would be paid, denoted $S_L(b)$ and $S_H(b)$, but the same marginal compensation rate employed. When such perfect discrimination is not possible, an optimal *two part tariff* will trade off some social

⁴ Without loss of generality, the costs of drugs and other complementary inputs are ignored.

⁵ This is of course, completely isomorphic with a model of a monopolist charging marginal cost (and hence earning zero profits).

benefits of expanded delivery against the financial savings embodied in a(n algebraically) higher fixed payment (paid by the physician).⁶

Let us assume briefly that $R = 0$. Then for any marginal price p per unit paid to the physician, the gross surplus earned by the L types is $S_L(p) = p^2/2\theta_L$ which is greater than the gross surplus earned by the H types, $S_H(p) = p^2/2\theta_H$, so the largest fixed fee the government could charge is $S_H(p)$. Independent of the fixed fee, conditional on relocating, physicians of type θ_L choose $q_L = p/\theta_L$, and those of type θ_H choose $q_H = p/\theta_H$. Thus social welfare is

$$\begin{aligned} W &= \phi[bq_L - (S_H(p) + pq_L)] + (1-\phi)[bq_H - (S_H(p) + pq_H)] \\ &= (b-p)p\left(\frac{\phi}{\theta_L} + \frac{1-\phi}{\theta_H}\right) + \frac{p^2}{2\theta_H} \\ &= \frac{(b-p)p}{\tilde{\theta}} + \frac{p^2}{2\theta_H} \end{aligned}$$

where $\tilde{\theta}$ defined by $1/\tilde{\theta} = \phi/\theta_L + (1-\phi)/\theta_H$ is the harmonic mean.

Straightforward differentiation yields the optimal marginal price (e.g., annual salary)

$$p^* = b[2 - \theta/\theta_H]$$

and optimal fixed fee

$$F^* = p^{*2}/2\theta_H.$$

If $R < F^*$, physicians have their relocation costs paid and are given a lump-sum "signing bonus" equal to $F^* - R$.

The two part tariff is a first step towards a fully non-linear incentive contract. In the simple case modeled here, with only two types of physician, such a non-linear schedule reduces to a mechanism that offers providers the option of serving in rural areas for a short stint or doing so for an extended period, with compensating payments. Formally, the government offers the alternative quantity-payment pairs, or *rural service options*, (q_L, t_L) and (q_H, t_H) to maximize social welfare. This optimization is subject to two constraints: first that all physicians are willing to work (individual rationality, this will be relaxed shortly), and second that they sort themselves between the two alternatives (incentive compatibility). That is,

⁶ See also Tirole (page 145, 1988) for a model of two part tariff pricing by a monopolist.

$$t_L - c_L(q_L) \geq 0$$

$$t_H - c_H(q_H) \geq 0$$

and

$$t_L - c_L(q_L) \geq t_H - c_L(q_H)$$

$$t_H - c_H(q_H) \geq t_L - c_H(q_L)$$

Straightforward arguments can be used to show that only the second and third of these four constraints bind, allowing the payment amounts to be solved for and substituted into the welfare function, leading to

$$W = \phi[bq_L - (c_L(q_L) + [c_H(q_H) - c_L(q_H)])] + (1 - \phi)[bq_H - c_H(q_H)].$$

The government's problem is then to maximize W by choice of q_L and q_H . The first order conditions are simply

$$c_L'(q_L) = b$$

and

$$c_H'(q_H) = b + \frac{\phi}{1 - \phi}[c_L'(q_H) - c_H'(q_H)].$$

The associated transfers are $t_L = c_L(q_L) + [c_L(q_H) - c_H(q_H)]$ and $t_H = c_H(q_H)$.

Thus the solution is characterized by physicians with low opportunity costs providing the efficient level of rural service ($c_L'(q) = b$) while earning a positive rent, although this rent is lower than under the uniform pricing policy above. Physicians with higher opportunity costs serve less than the efficient time in rural areas ($c_H'(q) > b$), but earn no rent. Indeed, if the costs of H -types are too high the government will be forced to rely solely on L -types to provide rural services, it being too costly to induce even a low level of participation by the others, making universal provision all but impossible.

3. Equilibrium Behavior of Unmonitored City Doctors in Rural Communities

The previous section examined the amounts that would need to be paid to physicians to get them to move to rural areas. This section takes a first look at their

performance, and thus increases the social value of a well-trained physician. (A well-trained physician with no incentives is, to a first approximation, no more valuable than a poorly trained one.) On the other hand, higher quality local supply, while providing stronger incentives, reduces the *need* for high quality externally provided services, so the impact of local quality on training decisions is ambiguous. Similarly, when the quality of local supply is low, it might be optimal to forsake the provision of (weak) incentives and to provide external physicians of sufficiently high quality to drive local providers from the market.

To formally model the interaction of a traditional healer and a city-trained physician, we employ a spatial competition model. Individuals within a community have a range of tastes, some preferring to visit a traditional healer, others favoring a physician. Individuals choose between the two given these tastes, on the basis of the quality of service provided, and the price charged, by each. To model this we assume individuals are uniformly distributed along the unit interval $[0,1]$, with the traditional healer located at 0 and the physician at 1. An individual located at position x gains net surplus $w_0 = u_0 - p_0 - tx$ by purchasing care from the traditional healer, compared with $w_1 = u_1 - p_1 - t(1-x)$ by purchasing care from the physician. u_i is the utility of care received from provider i (which depends on its quality), p_i is the price charged by provider i , and t represents the costs of switching from one provider to the other. \underline{u} is an individual's utility from self care, that is, her outside option.

In this model, the quality of the traditional healer is fixed, and that of the city-trained doctor is chosen by the government. The unit costs of provision c_i , which are constant on the intensive margin (i.e., as quantities vary), are assumed to be higher for providers of higher quality (say due to better outside opportunities). The providers then choose the prices at which they sell these services, given that demands respond according to the behavior of individuals who try to maximize their own well-being (measured by the surplus indicators w_i above). In general, higher prices will be observed when the costs of switching provider are large, and when one provider enjoys a larger quality advantage vis a vis the other.

When there is no physician, the local healer acts as a monopolist. Given a price p_0 , all those consumers who earn a surplus greater than \underline{u} visit the healer. That is, all consumers located to the left of x_0' purchase care, where x_0' satisfies

$$u_0 - p_0 - tx = \underline{u}$$

or

$$x_0' = (u_0 - p_0 - \underline{u})/t .$$

The healer's profit is $(p_0 - c_0)/x_0'$, which is maximized by setting $p_0 = (u_0 + c_0 - \underline{u})/2$. Writing $s_0 = u_0 - c_0$ as the economic surplus generated by the healer, as a monopolist he covers a share

$$\tilde{x}_0 = (s_0 - \underline{u}) / 2t$$

of the market, if this is less than one, and all of the market otherwise.

Now suppose the city-trained physician enters the rural community, effectively locating at $x = 1$. Let \tilde{x}_1 satisfy the condition

$$1 - \tilde{x}_1 = (s_1 - \underline{u}) / 2t.$$

Then if $\tilde{x}_1 > \tilde{x}_0$, each provider acts as a local monopolist, and there is a segment of the population, located between these values, that does not seek care from either. Otherwise, if $\tilde{x}_1 < \tilde{x}_0$, that is, if the average surplus is greater than the total travel costs, $(s_0 + s_1)/2 > t$, the firms engage in price competition. We assume this condition holds for the rest of this section, so all individuals purchase care.

Given qualities and prices, the position of the indifferent consumer is \hat{x} , satisfying

$$u_0 - p_0 - tx = u_1 - p_1 - t(1 - x)$$

or

$$\hat{x} = \frac{1}{2} + \frac{(\Delta u - \Delta p)}{2t}$$

where $\Delta z = z_1 - z_0$.

Given the qualities of the two services, market shares are determined by price competition. The healer solves

$$\max_{p_0} [p_0 - c_0] \hat{x}$$

with first order condition

$$p_0 = 2t\hat{x} + c_0.$$

Similarly, the physician solves

$$\max_{p_1} [p_1 - c_1](1 - \hat{x})$$

with first order condition

$$p_1 = 2t(1 - \hat{x}) + c_1.$$

Jointly these yield equilibrium prices

$$p_0^* = t + \frac{1}{3}(\Delta u + 2c_0 + c_1)$$

and

$$p_1^* = t + \frac{1}{3}(-\Delta u + c_0 + 2c_1)$$

The traditional healer serves a share

$$\hat{x} = \frac{1}{2} + \frac{1}{6t}(\Delta u - \Delta c)$$

of the market, and the physician serves the complement.

Policy Implications

Taking the quality of traditional healers, u_0 , as given, the government's problem is to choose how much training to give physicians that it wants to send to the rural area. Training is costly, and once the physician has located to the countryside, she is outside the direct control of the government. Thus the government assumes that given u_1 , the equilibrium prices determined above will be generated by local competition.

We assume welfare is the sum of local residents' net utility from services less the cost of training the physician, $K(u_1)$. We do not include the utility of the physician or the local healer in our measure of welfare. The government then chooses u_1 so as to maximize

$$W = \int_0^x (u_0 - p_0 - tx)dx + \int_x^1 (u_1 - p_1 - t(1-x))dx - K(u_1)$$

subject to equilibrium price setting by the providers. Thus the first order condition for the optimal level of u_1 given u_0 is

$$s'(u_1)[\alpha + \beta s(u_1)] = K'(u_1)$$

where $s(u_1) = s_1$ from above, and $\alpha = 1/2 - \beta s_0$, and $\beta = 1/18t$. We assume there is a unique solution to this equation satisfying the second order condition, within the neighborhood of which the left hand side is decreasing.

Not surprisingly, if the marginal cost of training increases, physicians who are sent out to rural areas should be less well trained. Also, the higher the *net value* of the healer's services, $u_i - c_i$, the lower the optimal quality of the city-trained physician. However, at low levels of healer quality, an increase in u_0 should be matched by a *reduction* in u_1 , but for high levels of healer quality, such an increase should be matched by an *increase* in physician quality. That is, when existing local capacity is weak, externally provided quality should be used as a *strategic substitute* for local quality, and when local capacity is strong, it should be employed as a *strategic complement*.⁷ This result can be interpreted as indicating that, at low general levels of local capacity, resources should be directed to those communities especially lacking in health care services, but when average levels are higher, they should be directed to communities with relatively high existing capacities.

Finally, if training costs are low enough, then it may be socially optimal to provide sufficiently high quality that the local provider is driven from the market. The city-trained physician then acts as a monopolist with the power to increase its price above the duopoly level, without losing customers. While this pricing behavior is not distortionary from small price increases (since demand is inelastic), it is welfare reducing, due to the fact that providers' profits receive no weight in the government's objective function. A price ceiling might then be required (if it can be enforced).

⁷ These terms are those introduced by Bulow et al. (1985).

4. Contract Design with Imperfectly Observable Performance

In this section we examine the role of explicit contractual incentives provided by an imperfectly informed government. In this class of models, some components of physician performance can be observed, with an error, and payments can be made based on these signals. The question we are interested in answering is "How should the payments depend on the signals?"

The essence of the problem is that paying strictly for output, in this case for health status per se, exposes a risk-averse provider to too much risk. The government, assumed to be risk neutral, can provide some implicit insurance by accepting a greater share of uncertainty, paying somewhat less and having this be to mutual advantage. On the other hand, absorbing all the risk by paying only a salary that is not conditional on performance blunts incentives to exert effort (such as not showing up for work). The optimal payment scheme is, as might be expected, a combination of salary and performance payments.

The basic model⁸ assumes that a medical provider produces some output, x , by exerting effort, e , but that the output also depends on factors beyond the provider's control. For example, output might be determined by the stochastic relationship $x = e + \eta$ where η is normally distributed with mean zero and variance v . Effort is costly to the physician (or, leisure is valuable) and the marginal cost of working increases according to the quadratic cost function $C(e) = ce^2/2$. The physician's utility function depends on money income paid by the principal (here the government) y less effort costs, $y - C(e)$, and exhibits constant absolute risk-aversion so $U(y - C(e)) = -\exp(-r(y - C(e)))$.⁹

If a linear incentive scheme is employed,¹⁰ so that $y = k + mx$, then the marginal payment to the agent that maximizes the principal's expected return is

$$m = 1/(1+rcv),$$

with k adjusting to satisfy the agent's participation constraint (i.e., k is large enough to guarantee that the doctor will accept the contract). We say incentives are high-powered if m is large (i.e., close to one), and low-powered if it is small (close to zero).

⁸ This formulation follows the notation used by Dixit (1999), building on the seminal work of Holmstrom (1979) and Holmstrom and Milgrom (1987, 1991).

⁹ These specific assumptions can be relaxed without changing the qualitative nature of the results.

¹⁰ Holmstrom and Milgrom (1987) show that the linear incentive schemes are optimal in a specific class of moral hazard problems.

The basic conclusion, from the formula for m , is that explicit incentives are more high-powered, i.e., are tied more closely to observable outcomes, the lower the risk aversion of the physician, the lower the marginal cost of effort and the lower the variance of the error of observation. While overly simple, the model gives one insight into appropriate payment schemes for doctors in public service. In health care any indicator of the change in health status will be subject to large variations. Actual health status (however measured) is determined by many factors not in the doctor's control and any measure of health status will be plagued by measurement errors itself.¹¹ For a risk averse doctor, applying the above formula will involve substantial blunting of incentives due to this uncertainty. Hence, there must be an appreciable wage component, k , independent of effort and performance in order to induce the doctor to accept the contract at all. High-powered incentives in real life are usually tied to inputs directly, as in fee for service, rather than contingent on output. Of course, that introduces its own set of distortions.

Several extensions can be made to overcome some of the unrealistic aspects of the above model. The first relaxes the assumption of a single "output" of a doctor. Doctors in public service are called upon to do many things, some closely related to health outcomes, some less so. Besides providing curative care, public health providers frequently have responsibility for immunizations, preventive health information campaigns, policy meetings, clinic accounting, inspection of water and sanitation facilities, community relations and so on. How does the existence of competing responsibilities affect appropriate payment schemes?

Suppose the principal can make payments respond to each of a number of different outputs separately. If some of the tasks we ask of the provider are easy to observe (such as attendance at district staff meetings) while others are quite difficult (such as the degree of diligence, or even just courtesy, extended toward patients), the standard intuition would suggest providing stronger incentives for the first type, and weaker incentives for the second. However, the insight provided by formally modeling the multi-dimensional problem (Holmstrom and Milgrom, 1991) is that this standard intuition is weakened. If the two tasks compete for the time of the provider, some of the extra benefit from rewarding the easily observed task comes at the expense of effort for the more difficult to observe task. Therefore, you will not pay as much for the former.

As Dixit notes, the interactive effect of incentives could show up even if the tasks have the same costs and variances in observation. This would arise if there were a technological link between the two such that they share a common cost function: $C(e_1, e_2) = c.(e_1^2 + 2ge_1e_2 + e_2^2)$ where e_i is effort on each of two tasks and g ($-1 < g < 1$) reflects the degree of complementarity ($g < 0$) or substitutability ($g > 0$) between them. In this case, the optimal payment is

¹¹ If the report of health status change is given by the doctor herself, this changes the game considerably and places the problem in the category of "costly state verification".

$$y = k + [1/(1+(1+g)rcv)] \times (x_1 + x_2).$$

The degree to which measured output is rewarded by income depends on whether the two tasks are complements or substitutes. If substitutes, incentives need to be weakened since increases in one form of effort detract from other desirable outputs as before. If complements, however, incentives are even sharper since increases in effort in one dimension actually contribute to other outputs. This adds a further argument for the combining of tasks into groups defined by their degree of complementarity. Not only is there a technological, cost reduction justification¹² but it also allows for the use of greater output orientation on the part of the payment scheme.

A second extension recognizes that there may be other relevant “principals” besides the government employer. Local residents may be able to exert their own influence on the behavior of the doctor. For example, it has been noted in the Indian states of Kerala and West Bengal that communities will take matters into their own hands and occasionally beat up public officials who are seen to be negligent of their duties.¹³ How this behavior changes the results is, unfortunately, sensitive to the way such sanctions are modeled. For example, Mirrless (1975) showed that with penalties of unlimited size (getting beaten up might verge on this), strong incentives can be provided at approximately zero cost. On the other hand, when infinite penalties are not desirable, local residents can still improve matters by reducing the information asymmetry at the heart of the incentive problem. By threatening to punish inattentive providers they effectively reduce providers' net effort costs, c . The principal/government can then adjust the payment mechanism in favor of higher powered incentives, that is higher m . If punishment by over-zealous local residents is subject to mistakes, providers will have to be compensated, effectively with danger money, that is with a higher base salary. So, the cost function in the original problem would be modified by adding a term representing local sanctions: $s(e^2/2 + \phi)$ and the solution replaces c with $c-s$ and increases k to compensate for the variance of ϕ . The net effect would be larger monetary transfers from the state budget, and better performance.

A third extension combines the previous two. Suppose there are two outputs – attention to clinical services and the execution of public health activities such as inspections of sanitary conditions, IEC activities, immunizations and other preventive services. Local demands are usually much greater for clinical services and absence from clinics can be a cause of dissatisfaction on the part of clients whereas inattention to public health activities may not be noticed. This situation will require direct supplementary payment for the public health services or else a separation of responsibilities between providers so that substitution possibilities do not arise. If demands on the physician are not paid for by clients (i.e., there are no fees for clinical services) the pay structure

¹² See Wilson (1989) for an organizational perspective on the complementarity issue.

¹³ Caldwell (1986).

induced by local pressure may require, in addition to direct, high powered pay for public health services, a higher salary as well since the time used for public health may decrease satisfaction of the community and cost the doctor good will (or, in the case of West Bengal, safety). Finally we note that the fact that there is active demand for clinical care but little or none for public health is sometimes cited as one reason why the former should be paid for by the patient and the latter supported by government via its salaried civil servants. In this example, the inability to charge patients may require both payments for public health as well as increased salaries, increasing public costs on two counts. Fees reduce demand and the attendant claims on providers' time. This diversion of time is additional to the diversion of funds between the two activities.¹⁴

5. Conclusion

This paper has applied a selection of models of incentives, competition, and contracts to various aspects of the problem of supplying medical services in rural areas of poor countries. This information can be used in two ways. First, it is possible to identify the scope and, perhaps more importantly, the limitations of the use of direct incentives to encourage better performance. The multi-dimensional nature of the output, the opportunity costs facing professionals, the welfare cost of taxation, the wide variety of information and monitoring structures available to government all have implications for the proper management of widely dispersed services. Second, such models may set realistic bounds on our appraisal of delivering certain kinds of services at all. Regardless of the attractiveness of policy options on theoretical grounds, recognizing that public employees are people who make independent decisions about their careers and lifestyles can set bounds on precisely how well government agencies can deliver promised services. If we can identify reasons why the best that can be expected is not particularly good, it may lead us to explore entirely different policy options. Maybe it is too hard to run certain decentralized systems. Maybe we should focus on less ambitious, but more readily implementable, goals such as provision of basic infrastructure.

We have not attempted to be comprehensive – the circumstances of delivering services to the poor in poor countries vary widely. In fact, as with the rest of the literature, there are few robust results on optimal incentives with imperfect information. Each circumstance requires a careful analysis of provider behavior. General pronouncements on the best way of delivering services should be treated with extreme skepticism. We might recall the words of Lao Tzu: “He who says he knows the way, does not know the way.”

¹⁴ Gertler and Hammer (1997) come to a similar conclusion by treating public health expenditures as a residual payment in the budget.

References

- Bulow, Jeremy, John Geanakoplos, and Paul Klemperer. (1985). "Multimarket oligopoly: Strategic substitutes and complements." *Journal of Political Economy* 93: 488-511.
- Caldwell, John C. (1986). "Routes to Low Mortality in Poor Countries." *Population and Development Review* 12(2).
- Chomitz, K., Gunawan Setiadi, Azrul Azwar, Nusye Ismail, Widiyarti. (1988). "What Do Doctors Want?: Developing Incentives for Doctors to Serve in Indonesia's Rural and Remote Areas." Policy Research Working Paper 1888. Development Research Group, World Bank, March 1998. Processed.
- Dixit, Avinash. (1999). "Incentives and organizations in the public sector: An interpretive review." Paper presented at the National Academy of Sciences conference on *Devising Incentives to Promote Human Capital*, December.
- Filmer, D., J. Hammer and L. Pritchett. (2000). "Health Policy in Poor Countries I: Weak Links in the Chain." *World Bank Research Observer* (August).
- Gertler, Paul and J. Hammer. (1997). "Financing and Allocating Public Expenditures in the Health Sector." In G. Scheiber and D. Chollet, eds., *Health Reform in Developing Countries*. Washington, D.C.: World Bank.
- Holmstrom, Bengt. (1979). "Moral hazard and observability." *Bell Journal of Economics* 13(2, Autumn): 324-40.
- _____, and Paul Milgrom. (1987). "Aggregation and linearity in the provision of intertemporal incentives." *Econometrica* 55(2, March): 303-28.
- _____ and _____. (1991). "Multi-task principal-agent analysis: Incentive contracts, asset ownership, and job design." *Journal of Law, Economics, and Organization* 7: 24-52.
- Mirrlees, James. (1975). "The theory of moral hazard and unobservable behaviour, Part I." Reprinted in *Review of Economic Studies* 66(1, January) 1999: 3-21.
- Tirole, Jean. (1988). *The theory of industrial organization*. Cambridge, Mass.: MIT Press.
- Wilson, James Q. (1989). *Bureaucracy: What Government Agencies Do and Why They Do It*. Basic Books.

Policy Research Working Paper Series

	Title	Author	Date	Contact for paper
WPS2530	The Role of Nongovernmental Organizations in Primary Education: A Study of Six NGOs in India	Shanti Jagannathan	January 2001	S. Kumar 87021
WPS2531	Growth Implosions, Debt Explosions, and My Aunt Marilyn: Do Growth Slowdowns Cause Public Debt Crises?	William Easterly	January 2001	K. Labrie 31001
WPS2532	Market Presence, Contestability, and the Terms-of-Trade Effects of Regional Integration	Maurice Schiff Won Chang	January 2001	L. Tabada 36896
WPS2533	How Much War Will We See? Estimating the Incidence of Civil War in 161 Countries	Ibrahim Elbadawi Nicholas Sambanis	January 2001	H. Sladovich
WPS2534	Indigenous Ethnicity and Entrepreneurial Success in Africa: Some Evidence from Ethiopia	Taye Mengistae	January 2001	R. Bonfield 31248
WPS2535	Trade Policy Reform in the East Asian Transition Economies	Will Martin	January 2001	L. Tabada 36896
WPS2536	Malaysian Capital Controls	Ron Hood	January 2001	H. Abbey 80512
WPS2537	Regulatory Reform in Mexico's Natural Gas Industry: Liberalization in the Context of a Dominant Upstream Incumbent	Juan Rosellón Jonathan Halpern	January 2001	K. Nemes 85440
WPS2538	Designing Natural Gas Distribution Concessions in a Megacity: Tradeoffs Between Scale Economies and Information Disclosure in Mexico City	Juan Rosellón Jonathan Halpern	January 2001	K. Nemes 85440
WPS2539	Disorganization or Self-Organization? The Emergence of Business Associations in a Transition Economy	Francesca Recanatini Randi Ryterman	January 2001	M. Johnson 36320
WPS2540	Recapitalizing Banking Systems: Implications for Incentives and Fiscal and Monetary Policy	Patrick Honohan	February 2001	A. Yaptenco 31823
WPS2541	Household Schooling Decisions in Rural Pakistan	Yasayuki Sawada Michael Lokshin	February 2001	P. Sader 33902

Policy Research Working Paper Series

	Title	Author	Date	Contact for paper
WPS2542	Checks and Balances, Private Information, and the Credibility of Monetary Commitments	Philip Keefer David Stasavage	February 2001	P. Sintim-Aboagye 37644
WPS2543	When Do Special Interests Run Rampant? Disentangling the Role in Banking Crises of Elections, Incomplete Information, and Checks and Balances	Philip Keefer	February 2001	P. Sintim-Aboagye 37644
WPS2544	The Uniqueness of Short-Term Collateralization	Leora Klapper	February 2001	A. Yaptenco 31823
WPS2545	Financing the Future: Infrastructure Needs in Latin America, 2000–05	Marianne Fay	February 2001	A. François 37841
WPS2546	Gender Dimensions of Pension Reform in the Former Soviet Union	Paulette Castel Louise Fox	February 2001	J. Smith 87215