

**Institutional Solutions to the Principal-Agent
Problem in African Health Care**

By

Kenneth L. Leonard, Columbia University
David K. Leonard, University of California at Berkeley

October 1998

Discussion Paper Series No. 9899-02

dp 9899-02

pages: 53

Institutional Solutions to the Principal–Agent Problem in African Health Care

Kenneth L. Leonard*
Department of Economics
Columbia University
KL206@columbia.edu

David K. Leonard†
Department of Political Science
University of California at Berkeley
leonard@stat.berkeley.edu

October, 1998

Abstract

Free markets for health care in Africa do not function properly, in that patients exhibit willingness to pay for health care and yet practitioners are unable to sell their services. It is widely acknowledged that health markets everywhere are troubled with imperfect information. Therefore it is no surprise that free markets and spot contracts do not lead to an efficiently functioning market for health care. When issues of agency are not resolved we find practitioners specializing in the sale of pharmaceuticals but not using their skills as diagnosticians. Mechanisms that can reduce agency cost are beneficial to both patients and practitioners.

This paper draws on theory and empirical evidence to examine what institutions are necessary to solve the problems of imperfect information in this context. We dismiss government regulation because the regulatory capacity does not exist in most African countries. Theory suggests that self-regulation by professional bodies should arise as privatization continues. Empirical evidence, however, suggests that this conclusion is overly-optimistic. On the other hand, referral networks perform much the same function but do not require centralized control. The most successful institution for the delivery of quality medical care in Africa is that of independent, pre-existing value-based organizations (missions) and we suggest their choice of institutional form has contributed to their success.

JEL Classification: D2 I1 O2

Keywords: Institutions, Rural Health Care, Principal-Agent, Adverse Selection, Moral Hazard

*Research funded by NSF grant #94-22768 and a University of California Rocca Fellowship.

†Research funded by NSF grant #94-22768 and the Ministry of Foreign Affairs of the Netherlands

1 Introduction

With Africa in prolonged economic crisis, most of its states in fiscal collapse, and pressures from the World Bank for Structural Adjustment, most sectors of the continent's economies have been increasingly privatized. All too frequently, however, the "miracle of the market" has not followed. This failure is particularly acute in the health sector. It is widely acknowledged that markets for health everywhere fail to meet one of neoclassical economics' assumptions for fully competitive markets. They are troubled with *imperfect information*, in that someone who is sick and seeks medical treatment (the *principal*) cannot assess the appropriateness and quality of the effort the health practitioner (her *agent*) is making on her behalf. Without institutions or mechanisms to overcome the principal-agent problem, the market for health will not operate in an economically optimal manner. Therefore it is no surprise that free markets and spot contracts do not lead to an efficiently functioning market for health care. Individuals can simply try to cope with the problem or institutions can arise to govern and facilitate relationships between individuals. We will show that American solutions often do not work in Africa. What institutions might then overcome this principal-agent problem in medicine on the continent?

To answer this question we examine the institutions currently operating in Africa to mitigate these problems. We draw on the research completed in a seven country African collaborative project in which we participated (Leonard, ed 1999). The project examined the changing markets for both human and animal health care on the continent, with independent research projects in each country. Although we concentrate on medical care for humans in this article, we also use research done on veterinary medicine. The biological science underlying the two professions is nearly identical and they are organized in very similar ways. Veterinary care in Africa is particularly interesting, because over the last dozen years it has moved rapidly from being more dominated by the state than human health was to now being substantially more private. The health and veterinary markets in Africa therefore

nicely illuminate one another¹.

In addition to the information gained from the research project we draw from three bodies of literature. The literature on institutions of health care delivery in the west offers a contrast to the African situation and illuminates the importance of 'macro' institutions (such as functioning court systems) for the sustainability of other institutions. The literature on health care delivery in Africa informs our discussion of the behavior of individuals in their search for health care. A third source is the literature on mechanism design which provides a framework for the discussion of existing structures and patterns and allows us to make predictions about the success or failure of potential mechanisms or institutions.

The areas studied in the research project and in the literature on health care in Africa offer variety in both institutional and technological environments, allowing us to draw conclusions about the effectiveness of existing and potential institutions. The paper is framed by a theoretical exposition of agency, starting with moral hazard, observable and unobservable efforts, and discussing institutions and technologies that allow for enforceable contracts to deliver unobservable effort. Subsequently problems of adverse selection are analyzed and potential solutions are examined. Many of the candidate institutions for moral hazard and adverse selection in African health in fact are similar.

We identify three types of mechanisms or institutions that can resolve these problems. The first type are decentralized mechanisms that exist because of particular technological features either of their environment or of pre-existing institutions. These mechanisms do not rely on governments or other pre-existing organizations for their existence and in theory should form wherever the exogenously determined environment allows and where superior institutions do not exist. By assuming some degree of rationality (and drawing from studies of patient behavior to back this up) we can learn a great deal about conditions necessary for these types of institutions, because we can infer that where they do not exist there is a technological

¹For more on the mutual relevance of veterinary and human medicine, especially in Africa, see Schwabe (1984); Majok and Schwabe (1996, p. 128).

or institutional barrier to their formation. These mechanisms, although helpful, are nonetheless sub-optimal (they cannot achieve the solutions that could be obtained in the absence of agency), implying that there is room for other institutions that can achieve superior solutions.

There is also the potential for institutions that are autonomous of existing organizations but require some degree of coordination. These would be institutions managed by medical professionals and independent of governments or pre-existing organizations. We specifically examine professional self-regulation and referral networks. Such institutions are not generally present in Africa, but it is difficult to conclude why, because collective action issues coupled with the short time period since deregulation clouds evaluation of 'rationality'. In other words, it is difficult to conclude that because they have not formed they will not form. We do, however, find strong evidence to suggest that self-regulation by professionals is unlikely to occur, but hold out the possibility that effective referral networks might form and serve the role of reducing agency costs.

A third type of institution is one that relies on a pre-existing organization. In Africa the relevant organizations are government agencies and non-governmental organizations (NGO's), specifically churches and missions. Both organizations use the institution of hierarchical supervision. We find strong evidence that existing mission health facilities are effectively using this institution to deliver high quality health care that resolves, to some degree, the problems of agency. On the other hand governments, though using the same institution, fail to utilize it effectively and generally provide low quality care despite great expense in training, facilities, and supplies. Unlike decentralized mechanisms or autonomous institutions, the existence of effective organizations is an exogenous event. This study recommends that existing and effective organizations take or be allowed to take a more active role in the provision of health care, both extending their current reach and replacing many traditionally government services.

2 The Established View of Health Markets

2.1 In the Economics Literature

Economic theorizing about health care implicitly derives from the “medical model” of Parsons (1951, Chap. 10). Parsons saw the patient as an inadequately informed and hence passive recipient of care from an omniscient physician. Economics has generalized this medical model as a principal–agent one and characterized the patient/principal as suffering from imperfect information in her health care transactions (Arrow 1963, Dranove and White 1987, Pratt and Zeckhauser 1985).

The costs of principal–agent transactions are particularly manifest in what are called moral hazard and adverse selection problems. *Moral hazard* arises when one or both of the parties to a transaction is unable to effectively monitor the effort the other devotes to the production of the contracted good. If a principal is unable to observe the extent to which the agent is doing quality work on her behalf, she may have difficulty providing adequate incentives to induce that quality. If this problem is unresolved the principal will not pay for quality even if it is potentially available. *Adverse selection* occurs when some agents possess a characteristic that is of use in treatment but cannot be detected by principals. These quality practitioners will not be paid for the extra value of the product they possess, and when acquiring the characteristic has costs, they will be driven out of the market by those who do not acquire the valuable characteristic. Moral hazard and adverse selection create losses for both the principal and the agent; the former is unable to get quality for which she is willing to pay, while the latter is unable to be paid for quality that he is able and willing to provide.²

Given losses introduced by the nature of the relationship between the patient and the physician the question then is, what kind of incentive or institutional structure

²This branch of the literature is in superficial contrast to the branch concerning issues of insurance in health care. In that literature moral hazard refers to an increased demand for services by patients because insurance reduces the marginal price. Adverse selection refers to the possibility that insurance companies might attract patients with above average risks (Pauly 1974). In both literatures, however, moral hazard refers to hidden action problems and adverse selection refers to hidden information ones (Arrow 1985). In any case we do not discuss insurance in this article.

can be created that will induce the agent to consistently provide high quality services even though principals do not know whether or not he has done so? The literature on agency in health care suggests three, interacting, potential solutions to this question.

First, long-term associations increase the value to agents of their relationships with principals and give principals more information about the likelihood that effort is being provided. This makes limited monitoring by the principal more effective, for it increases the value of the threat to sever future connections if something is found to be wrong. However, in Africa most health care is currently subsidized, removing the sting of any such threat. Though privatization is occurring it is currently difficult to lure professionals to private practice outside elite, urban markets and thus patients have few non-subsidized alternatives. The literature also suggests that clients might have long-term associations with organizations such as hospitals, primary-care physicians or insurance companies to the same benefit. We will examine such relationships in this paper.

Second, a “large stock of value that could be lost through bad behavior, such as reputation or assets subject to suit, is a strong incentive for good behavior.” If the penalty is a large fine and is paid by the faulty practitioner to the aggrieved client, as in a malpractice suit, the latter has considerable incentive to collect it (Pratt and Zeckhauser 1985, pp. 12-13). This solution requires both that the failure to engage in appropriate professional behavior can be proven to a third party (verifiability) and that institutions exist through which a suit can be successfully brought. Verification of professional effort is difficult in medicine, as there is a stochastic element to recovery and as the client is often the only person who knows the outcome of treatment. As the client stands to gain from the malpractice suit, her testimony may be biased. For these reasons, the courts work only as an institution for punishing bad professional behavior, not for inducing excellence. A further issue makes this solution problematic in Africa, however — access to the legal system is expensive and the courts themselves may well be corrupt. The ability of the average African to use a court successfully against a member of the educated elite is doubtful. As North (1990, p. 33) remarks, “one cannot take enforcement for granted. It is (and always

has been) the critical obstacle to increased specialization and the division of labor.”

Third, as is more common, the loss of a “large stock of value” for poor performance by the practitioner can be the revocation of the right to practice or have access to clients. Unlike malpractice suits, the loss to the practitioner is no gain to the client. This makes the testimony of the client about the success of treatment more reliable, by removing a major incentive for her to be biased. On the other hand it also removes her incentive to seek reparation and therefore this solution depends on the existence of a body that will monitor and enforce professional standards. In France and large parts of Africa the state plays this regulatory function. The same governmental incompetence and corruption that makes privatization attractive to Africa renders the state unreliable as a regulator, however. Anglo-Saxon countries instead have relied upon self-regulation by professional associations, which not only control licenses but hospital privileges as well. In addition referral networks among physicians can accomplish many of the same results. Self-regulating professional associations and referral networks do not currently exist in Africa; we will discuss their potential later in this paper.

Western markets for health rely on three centralized institutions that we do not observe in Africa. They require either effective courts, a group of professionals who will benefit from and can lead their profession to self-regulation or effective referral networks, or a government that is willing and able to perform a regulatory function. None of these currently function effectively in Africa north of Zimbabwe.

2.2 In the African Health Literature

The empirical literature on health care in Africa demonstrates quite conclusively that, in contrast to the “medical model,” the sick are not passive in their use of the health care system but instead are actively making choices between an array of options. For example, babies in Burkina Faso may be delivered by professionals, traditional birth attendants, or the “old woman of the village” (Noug tara et al. 1989). The sick frequently have been documented as seeking assistance from among government clinics and hospitals, mission facilities, for-profit private practitioners and

traditional healers. The choices made often are explained by factors unrelated to medicine, such as price and distance, or by static attributes of the facilities, such as availability of drugs or perceived quality (Cameroun: Litvack and Bodart (1993); Ghana: Waddington and Enyimayew (1989); Nigeria: Omorodion (1993) and Stock (1983); Swaziland: Yoder (1989); and Tanzania: Abel-Smith and Rawal (1992)). None of these factors contradict the image of the sick as passive recipients of health care from providers whose methods of operation are obscured in a 'black box.' Another group of studies, however, indicates that choice is being guided by the character or severity of the illness (Benin: Bichmann et al. (1991); Burkina Faso: Sauerborn et al. (1989); Kenya: Mwabu (1986); and Cameroun: Leonard (1998a), Ndeso-Atanga (1999)). Sometimes these studies cast doubt on the appropriateness of the judgements being made (Mali: Coppo et al. (1992)). But as a group this body of research suggests that the sick may be less ignorant about the kinds of medical treatment they require than the traditional medical model has implied.

That the sick in Africa are self-referring suggests a deficiency in the classical principal-agent health model, which portrays them as wholly guided by regulated professionals. Society-wide institutions to assure professional standards, which are assumed by the medical model, are not effective in Africa. Patients instead are choosing between institutions and organizations that regulate their practitioners in different ways and to different degrees. We will present evidence in Section 3.7 that these choices reflect understanding of the ability of different institutions to solve the agency problem.

3 Moral Hazard

As noted above, moral hazard arises when one or both of the parties to a transaction is unable to effectively monitor the effort the other devotes to the production of the contracted good. If a principal is unable to observe the extent to which the agent is doing quality work on her behalf, she may have difficulty providing adequate incentives to induce that quality. If this problem is unresolved, the principal will

be willing to pay only the value of inferior quality work, for that is all she can be sure of getting. To analyze the possibilities for solutions, we first must establish a general model.

3.1 Mathematical Preliminaries

We follow Grossman (1975) and model the benefit of health care as an increase in health stock. Health stock determines time available for work and leisure. This stock depreciates over time, can be increased by investment in health care and is subject to negative shocks from illness. In our model individuals purchase health care because they have received a negative shock to their health stock that creates a demand for investment in health.

There are two possible outcomes (X) of treatment, either $X = \bar{\mu}$ (cured) or $X = \underline{\mu}$ (not cured). Without loss of generalization we set $\underline{\mu} = 0$. There are two binomial distributions that describe treatment, ϕ_{full} (meant to invoke the idea that all parties to treatment contribute all possible resources) and ϕ_{none} (no resources are contributed), where the probability of $X = \bar{\mu}$ is equal to p_h for the distribution ϕ_{full} and the probability of $X = \bar{\mu}$ is equal to p_l for the distribution ϕ_{none} .

We introduce an arbitrary technology e ($0 \leq e \leq 1$) that describes a combination of the ϕ_{full} distribution and the ϕ_{none} distribution. Thus

$$X = e \cdot \phi_{\text{full}} + (1 - e) \cdot \phi_{\text{none}} \tag{1}$$

We deliberately based this description of X on the Spanning Condition of Grossman and Hart (1983) and the Linear Distribution Function Condition (LDFC) of Hart and Hölmstrom (1987), which will allow us to characterize incentive compatibility constraints as first order conditions³.

³Mirlees (1975) (as cited in Hart and Hölmstrom (1987)) shows that the first order conditions do not describe globally optimal actions for distributions such as $X = e + \theta$ or $X = e \cdot \theta$ when θ is any of the standard candidates for random distributions. Thus in order to obtain some theoretical results the choice of functional form for X is crucial. We follow this convention so that our results match with those of the theoretical literature, however, to our knowledge no similar restriction has been shown to exist in practice. “The extreme sensitivity to informational variables that comes across from this type of model is at odds with reality.” (Hart and Hölmstrom 1987)

We can rewrite e as a function of the best possible outcome by introducing a new variable μ and setting it equal to $e \cdot \bar{\mu}$. Thus we can rewrite the above as

$$X = \frac{\mu}{\bar{\mu}} \cdot \phi_{\text{full}} + \left(1 - \frac{\mu}{\bar{\mu}}\right) \cdot \phi_{\text{none}} \quad (2)$$

We can solve for the expected value of X as follows.

$$E(X) = \sum_{l=0, \bar{\mu}} \left(\frac{\mu}{\bar{\mu}} \cdot \text{pr}(X = l \mid \phi_{\text{full}}) + \left(1 - \frac{\mu}{\bar{\mu}}\right) \cdot \text{pr}(X = l \mid \phi_{\text{none}}) \right) \quad (3)$$

This simplifies to $\mu(p_h - p_l) + p_l \cdot \bar{\mu}$. The expected outcome of treatment is the probability of full recovery with no intervention plus the premium on the spanning technology $(p_h - p_l)$ times the magnitude of the intervention, μ .

The value of treatment is the outcome multiplied by a unit constant parameter, ω . In the case of human health care ω can be interpreted as the opportunity cost of healthy time (wage), and in the case of veterinary care, the value of productivity in a healthy animal.

If the technology μ could be purchased on the open market the client would purchase $\bar{\mu}$ units of μ if $\omega \cdot (p_h - p_l)$ were greater than ν_μ and zero units if $\omega \cdot (p_h - p_l) < \nu_\mu$, where ν_μ is the per unit cost of μ . However μ (health production) is not a good that can be purchased on the open market but rather is the product of goods, some of which can be purchased and some of which cannot. Thus $\mu = f(\Psi_{\text{client}}, \Psi_{\text{pract}}, \Psi_{\text{condition}}, a, b)$ where Ψ_{client} is the set of characteristics of the client, Ψ_{pract} is the set of characteristics of the practitioner and $\Psi_{\text{condition}}$ is the set of characteristics of the condition reported, a is the effort of the practitioner and b is the effort of the client.

3.2 Capacity

Practitioners provide two things in the production of μ , capacity ($\pi \in \Psi_{\text{pract}}$) and effort (a). It is helpful to make this distinction between the capacity of a medical provider and the effort he exerts in using it. Capacity such as the presence of

equipment is relatively easy for clients to assess. Capacity such as the skills of providers can be more difficult to assess. To the degree that capacity is observable a market will develop in which practitioners who possess the capacity can charge higher prices than those who do not. The choice to invest in capacity will then depend on benefit of that capacity to clients as well as its cost. For those elements of capacity that are unobservable there is the potential for adverse selection and under-investment. This is discussed in section 4.

3.3 Observable Effort

Being in the best hospital with the best physicians is of little use if no one will pay any attention to you. The value of the effort can be generally represented in the continuous case as $\frac{\omega}{\mu} \cdot \frac{\partial \mu}{\partial a} \cdot (p_h - p_l)$ and if effort were sold in the market the patient would purchase effort up to the point where its marginal value equals its per unit cost (ν_a).

There are features of a health facility and its clinicians that do indicate effort and that would be visible even to the untrained observer – (i) the degree to which adequate quantities of pharmaceuticals are maintained in stock; (ii) the propensity to prescribe treatment and the quantity of drugs given; (iii) the frequency with which prescriptions are given by injection rather than orally; (iv) the ‘clinical manner’ (also known as ‘bedside manner’) of the attending practitioners; and (v) the adequacy of instruction given to patients as to how to use their prescriptions and how to prevent recurrence of the problem. The first four items on the preceding list are acknowledged in the literature as influential in patient choice and satisfaction in Africa (Bichmann et al. 1991, Litvack and Bodart 1993, Omorodion 1993, Stock 1983, Waddington and Enyimayew 1989).

When health care is sold in the free market these efforts can be purchased at their value to the patient. If $a = \sum_l a_l \cdot \gamma_l$ the value of any type of effort is $\gamma_l \cdot \frac{\omega}{\mu} \frac{\partial \mu}{\partial a} \cdot \frac{1}{\mu} \cdot (p_h - p_l)$ and clients will continue to purchase it until its value equals the cost. In a market for health, practitioners could compete on their provision of observable effort rather than negotiating the level with each patient. Even when

their services are subsidized or prices are set exogenously we expect that side-payments (very common in Africa) would lead to improved provision of observable efforts.

3.4 Unobservable Effort

Effort is not always observable. Assume that a is not observable, that $\mu = a$ (μ is a direct function only of practitioner effort), and that $a \in \{0, 1\}$ (the practitioner can either exert a fixed amount of effort or not exert any effort). The client engages the practitioner who then chooses whether or not to exert effort. The patient does not observe the practitioner's choice of a but does observe X . In only two trivial cases will the patient know for certain that the practitioner exerted effort.

The first is the case where $p_t = 0$ and therefore when the client observes $X = \bar{\mu}$ (or $X = \underline{\mu}$) she can know effort was (was not) exerted. In the second case if the practitioner were known to prefer $a = 1$ over $a = 0$ then this knowledge, rather than any information gathered from observing X , would inform the patient that the practitioner had exerted effort. In those circumstances even the observation of $X = \underline{\mu}$ should not change the client's belief that effort was exerted. Both are trivial examples; patients often recover from diseases that were poorly diagnosed and people rarely prefer exerting effort to shirking, *ceteris paribus*.

There is a disutility to exerting effort $c(a)$. For this example we set the disutility of no effort equal to zero and the disutility of exerting effort ($a = 1$) equal to 1. Now the value to the practitioner of exerting effort is -1 and the value of not exerting effort is 0. Any fixed payment to the practitioner will not change the dominance of the no effort strategy over the effort strategy. If the client knows these disutilities she will realize that no matter what outcome is observed the practitioner did not exert effort. When $\mu = a$ there is therefore no reason to visit a practitioner.

If the patient offers to pay the practitioner some sum of money that is contingent on the outcome observed (P_X) she can induce the practitioner to exert effort. A good example is the "pay only if cured" contract of Dranove and White (1987). For a risk neutral practitioner the difference in expected values of the contingent

payment when he exerts effort and when he does not must be greater than or equal to the difference in the disutility of exerting effort and the disutility of not exerting effort. Thus he will exert effort only if:

$$\begin{aligned} \underbrace{P_X \cdot p_h - 1}_{\text{exert}} &\geq \underbrace{P_X \cdot p_l - 0}_{\text{not exert}} & (4) \\ P_X \cdot (p_h - p_l) &\geq 1 \end{aligned}$$

Assuming a contingent fee that is exactly enough to induce the practitioner to exert effort the contract is beneficial to the client when the net benefit with the contingent fee exceeds the benefit without the contingent fee. Thus the contingent fee contract is better when the value of healthy time is higher or when the difference between high and low effort expected outcomes is higher.

$$\begin{aligned} \underbrace{\omega \cdot ((p_h - p_l) + p_l \cdot \bar{\mu}) - P_X}_{\text{with cont. fee}} &\geq \underbrace{\omega \cdot (p_l \cdot \bar{\mu})}_{\text{without cont. fee}} & (5) \\ \omega \cdot (p_h - p_l)^2 &\geq 1 \end{aligned}$$

This scheme is what we will refer to as a contingent-fee contract. There can be other fees paid but any contract that contains a contingent-fee is referred to as a contingent-fee contract. P_X can depend on a number of factors other than the outcome (such as the value of the outcome) but must vary according to the outcome. If there is more than one type of effort the patient can still purchase the observable effort in the usual way and then write a contingent-contract over the unobservable effort. The co-existence of efforts does not change the basic solution to the problem.

Numerous anthropological studies have observed that traditional healers in rural Africa charge a nominal initial fee and expect the bulk of their income to come from ‘appreciation’ payments by the patient after *successful* treatment (Zaire: Korse et al. (1989), Burundi: Baerts (1989), South Africa: Edwards (1983), Nigeria: Oyenyé and Orubuloye (1985), the Ivory Coast: Lasker (1981), Botswana: Staugård (1985), and Zimbabwe: Gelfand et al. (1985) and general sources: Conco (1972)). A random sample of 800 households in South-West Province of Cameroun (Leonard 1998a)

suggests that patients who visit traditional healers make different payments depending on the success of treatment. Two types of payments were identified in the survey, those made before and those made after treatment. For patients who were cured of their ailments the average total payment both before and after was 6,545 CFA, whereas those not cured paid 3,338 CFA ($p=0.077$). Payments made after the initial consultation varied even more: 4,990 CFA in successful cases compared to 1,384 CFA in unsuccessful cases ($p=0.047$). For visits to modern practitioners payments are higher in unsuccessful cases⁴, though the difference is not significant.

It is rare (and generally considered un-ethical) for a modern medical practitioner to make his fee contingent on the outcome of treatment. On the other hand we do find circumstances in which the fixed fees are billed *after* the service is provided, which either implicitly or explicitly involves a loan. If the practitioner knows that his chances of collecting the fees are small if the outcome of the treatment is a failure, the contract becomes a contingent payment one, with the implied incentive effects to medical effort. We do *not* find this form of contract among modern human health care providers in Africa. But we *do* find it in private veterinary practice, where deferred fees are granted to livestock owners who have no liquidity problems and clearly are using them to gain the benefits inherent in a contingent contract (Kenya: D.K. Leonard, personal observation and G. Njiru, personal communication). Having observed this phenomenon and its incentive effects in Africa, we then realized that a similar informal contract existed in US private practice before health insurance became pervasive. Physicians of that era reported substantial amounts of uncollectable fees (a third of gross was not uncommon). If repayment rates were associated with outcomes a form of implicit contract was in use. Doctors were reluctant to press the family for payment when the patient died, for example, an inhibition that did not exist when an insurance company could be billed. Implicitly, then, the development of insurance removed from the American health system one of the institutions providing incentives for medical effort. Since the US had other institutions in place that could achieve this incentive effect, the change may have

⁴Perhaps because unsuccessful cases require continued medication at additional expense.

had only minor quality consequences there, although it might have contributed to the decline in personal attention from physicians. Its absence in Africa, where the anonymous character of most human health care and the weak loyalty of families to particular medical providers makes credit impossible, is much more consequential for practitioner effort.

3.4.1 Client Effort and Optimality Results

Under full information the optimal provision of health care would equate the marginal value of practitioner effort with its marginal cost; $\frac{\partial E(\omega \cdot X)}{\partial a} = \frac{\partial c(a)}{\partial a}$. When effort is unobservable the client cannot choose the level of effort and the payment independently — she can choose only the payment scheme. When both principal and agent are risk neutral⁵ and only the agent undertakes unobservable effort a contract in which the agent captures the full marginal value of his effort achieves the full information solution. This does not mean that the agent has to be paid the full value of the outcome, but that he receives the value of the marginal outcome. Thus if $P_X = \omega X - \omega \bar{\mu} p_h + P_F$ (the value of the outcome minus the expected value of the outcome plus a fixed fee) the expected payment is P_F , but the practitioner captures the full value of the marginal income because P_X changes one to one with ωX . With a contract of this type the practitioner will exert effort such that the change in his payment with the incremental unit of effort is exactly equal to its cost, achieving the full information solution.

An important element in medical treatment, however, is the effort of the client — either the sick individual (or her family), or the herder. Since the majority of this effort takes place out of sight of the practitioner it is also unobservable. Were the client to enter a contract in which all the reward or punishment from outcomes were borne by the practitioner her incentive to exert effort would be reduced. It is easiest to visualize this potential problem from the point of view of a herder. Herder and practitioner enter a contract over the treatment of a sick animal. The herder agrees to pay the practitioner the full value of restoring the animal to health

⁵Relaxing this assumption complicates the solution without changing the results substantially

(minus fixed payments) if it is cured and nothing if it dies. This contract aligns the incentives of the practitioner correctly and he will exert effort until its marginal value equals its marginal disutility. But for a part of the treatment the animal is in the herder's care. The herder must make sure the medication is properly taken, avoid conditions that might stress the animal and generally nurse the animal back to health. These actions require effort — effort that has opportunity costs — but the practitioner cannot observe whether or not these things were done. In the north, doctors (veterinarians) can admit humans (animals) and take the patient's care into their own hands, but this usually does not happen in Africa. Even when patients are admitted, the family frequently performs most of the routine nursing functions as this reduces paid labor costs.

Thus μ is generally a function of both medical (a) and patient (b) effort as well as other possible characteristics of the client, the practitioner or the disease condition. When both a and b are unobservable it does not make sense for either the practitioner or the client to earn the full value of the marginal outcome. To achieve the best possible solution the contingent-payment contract will take the form of a sharing-rule, in which the value of the marginal outcome is divided between the client and the practitioner. Thus, for example, the practitioner might receive $P_X = r\omega X + P_F$ and the client retain $(1 - r)\omega X - P_F$. If the value of treatment must be completely consumed by payments to either the client or the practitioner (conservative or budget-balanced contracts) we cannot achieve the same efficiency as is produced by full information. This is because each of the parties receives only a share of the return to their effort and will therefore exert less effort. However sharing-rules are a distinct improvement over contracts that lack contingent-fees⁶. A proof of this concept examines a contract in which the amount of the output shared with the practitioner is very small. At this point the client has very strong incentives to exert unobservable effort and the practitioner has almost no incentive to exert unobservable effort. If the cost of exerting a unit of effort gets larger for

⁶For a theoretical discussion of contracts with joint production involving unobservable efforts see Hölmstrom (1982). For a discussion of their application in this context see Leonard (1998b, pp. 8-12).

each unit exerted (there is a relatively low cost to exerting the first unit of effort or no increasing returns to scale for effort) then there is some share small enough, but greater than zero for which it makes sense (in terms of expected outcome) for the client to increase the share paid to the practitioner. In this way the patient will decrease her effort a little and the practitioner will increase his effort a little and the expected outcome will improve. Thus we demonstrate that some sharing is better than no sharing, when budgets are balanced⁷.

Under different sets of restrictions contracts that do achieve the full information solution are possible. Since budget-balanced sharing-rule contracts fall short of the full information solution, this admits the possibility of another class of contracts that do better than sharing-rule contracts. We will show later that relaxing the condition that payments are conservative allows contracts to achieve the full information solution.

We turn now to the issue of enforceability: though contingent-payment contracts are improvements on the agency problem there are important conditions that must be met before they can be used.

3.5 Enforceability and Verifiability

Though many aspects of effort are unobservable, we have claimed that outcomes are observable. Unfortunately for most contracts in which payments are contingent on the observation of outcome, *outcomes need to be verifiable as well as observable*. Thus outcomes must be observable to a third party such as a court that can adjudicate disputes in the resolution of contracts.

When outcomes are not verifiable contracts are generally not enforceable because the practitioner or the client might refuse to honor the contract. In the case of the contingent payment contract introduced above the client is supposed to make an additional payment to the practitioner if the outcome is good. However she no longer has an incentive to make that payment because the outcome is not observed by any institution that can enforce payment or punish lying. If the practitioner

⁷A formal proof can be found in Appendix A.

anticipates that a contract is unenforceable, he will not accept it or will exert effort as if his incentives were guided only by fixed payments. We have said that, in the absence of other institutions, contingent-payment contracts reduce agency costs, benefiting both practitioners and clients. Therefore the client is hurt by her ability to violate the contract. If she were able to bind herself to always tell the truth or could utilize a technology that made outcomes verifiable she would be better off.

The success of the traditional healer contract depends on the fact that the success of the outcome is effectively verifiable and therefore the contract is enforceable. In principle the patient could deny that she had been successfully treated and refuse to make contingent payments. However, most Africans believe that the healer has the spiritual powers to know the true outcome of the treatment (or is the agent of spirits with such powers) and that he can invoke supernatural punishment on those who seek to cheat him. Patients believe outcomes are verifiable (that 'powers' observe the outcome) and this binds them to tell the truth. Thus, although contingent contracts reduce agency costs, modern medical practitioners usually lack the traditional healer's ability to enter into enforceable contracts and are unable to transact business in this way.

A 'modern' medicine variant of the contingent contract was observed for maternities by Ndeso-Atanga (1999) in his study for our project in Cameroun. With the severe decline in real government salaries from Africa's prolonged economic crisis, an informal convention has developed in Cameroun that the family pay a substantial 'appreciation' to the government clinic mid-wife after the successful delivery of a healthy child. This outcome is readily observable and verifiable to all the parties to the transaction, and the payment is made before the mother and child leave the clinic (further aiding enforcement). The existence of a contract that creates incentives for effort on the part of the mid-wife has dramatic effects. There is no doubt that government hospitals and clinics in Cameroun have staff and facilities with a medical *capacity* equal to or greater than that of the missions. For a whole range of illnesses, however, substantial numbers of people go to missions because they believe that the quality of *effort* exerted by their staff will be superior to that of their

government counterparts. Since the maternity contingent contract is able to induce that effort in government clinics, however, the numbers of people going to missions for deliveries is dramatically lower (3% vs. 30% for other conditions).

3.6 Unverifiable Outcomes

In most types of medical transactions outcomes are not verifiable and therefore contracts are not as easily enforced as in the preceding two circumstances. Are there institutional mechanisms that can assure effort in these, more typical circumstances? We examine institutions that allow enforcement of contingent payment contracts without verifiability and institutions that insure effort using mechanisms other than contingent contracts.

3.6.1 Reputation

Though they are uncommon, there exist practitioners who provide high quality effort despite the absence of incentives to do so. They might have a taste for exerting effort or derive utility from the utility of the population they serve. The socialization literature, particularly for professionals, examines processes that seek to create individuals who exert effort independent of direct incentives but the mechanisms have high probability of success only when reinforced by other indirect incentives (Leonard 1991, Chap. 12);(Freidson 1970). There is no evidence that such indirect mechanisms operate effectively for government and private medical practitioners in Africa. Hence in Africa we are concerned strictly with *self*-motivated practitioners.

With imperfect information, however, such an 'extraordinary' (as we shall refer to him) practitioners attracts more clients— because they can count on his effort without observing it. Therefore potentially he could earn a higher income. This higher income is not why he behaves as he does but it is possible that an *ordinary* practitioner could benefit by seeking to *imitate* an 'extraordinary' practitioner. In this light we model reputation as the changing beliefs of clients such that, correctly or incorrectly, they believe a given practitioner will always provide effort no matter

what the incentives are in a given interaction. Clients have a prior belief about the probability that they are dealing with an ‘extraordinary’ practitioner and this belief can be changed by observing the outcomes of many interactions with the same practitioner.

Practitioners who practice in a limited geographical area can gain a reputation when patients observe a long string of mostly positive outcomes from treatments (no practitioner can *always* achieve good outcomes). We consider a ‘good’ practitioner to be one who is imitating an ‘extraordinary’ practitioner. ‘Good’ practitioners still experience disutility of effort. When a practitioner arrives in an area that has no information about his type he faces a choice either of acting rationally in the context of each interaction or of forgoing current consumption, providing effort and hoping to gain a reputation. His practice when he has a ‘good’ reputation must offer some reward that is worth waiting for.

Again, assume $\mu = a$, that $a \in \{0, 1\}$ and that clients have a prior belief, p_0 of the type of practitioner they are dealing with. p_0 varies between 0 (‘ordinary’) and 1 (‘extraordinary’) and can be updated for each observed treatment, such that p_n is the belief as to the type after n observations of treatment. The initial prior should reflect the client’s knowledge of the distribution of kinds of practitioners. We assume clients update their beliefs using a Bayesian updating process. This follows the model outlined by Kreps and Wilson (1982) of the decision making process of a firm trying to decide whether or not to enter a market when the established firm may or may not be (or be imitating) a ‘crazy’ firm that irrationally protects market share with price-cutting. All potential clients in a given area know the history of every past visit to a particular practitioner, i.e. all potential clients have the same prior belief at any moment in time. If beliefs were not updated it would never be rational to try to gain a good reputation.⁸

Figure 1 shows how the prior is updated for a few different types of distribution

⁸This differs from the reputation model of Satterthwaite (1979) and Pauly and Satterthwaite (1981) in that the cost of gathering information is basically zero. Their model examines the cost of gathering information when there are differing numbers of practitioners. In our environment the number of practitioners is low so information gathering is a less interesting element.

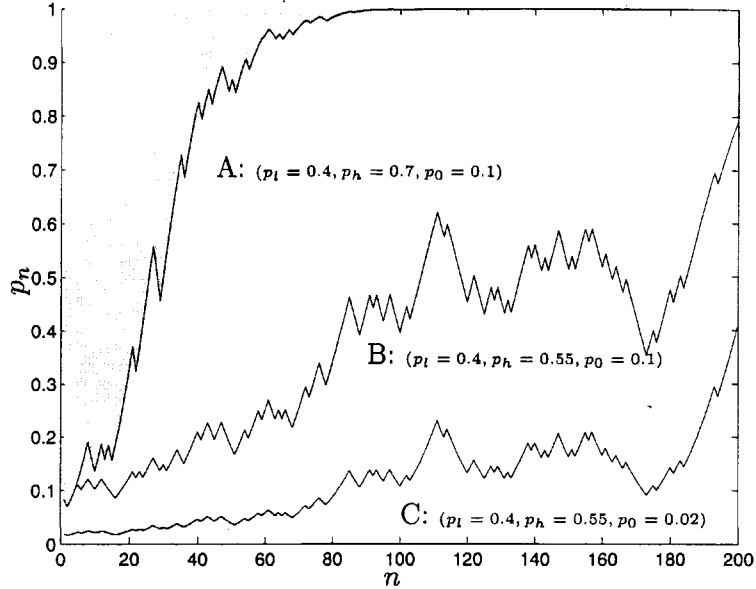


Figure 1: Belief updating with observation of outcome

functions over outcomes, (X) and two different initial priors, p_0 . In all three cases the practitioner is trying to gain a reputation by exerting effort. The zig-zag of the curve shows the random element of outcome despite consistent effort.

When clients think 10% of the practitioners they observe are likely to be ‘extraordinary’ and there is a sizeable spread between p_l and p_h (line A) clients discover that the practitioner is ‘good’ relatively quickly, although far from immediately. Of greater importance than the number of observations required for convergence to certainty is the slope of the ‘belief curve’, implying that practitioners see some rewards from their strategy relatively quickly. In contrast when the spread is tighter (the return to effort is lower) updating occurs at a slower rate (line B). When clients are pessimistic about the profession ($p_0 = 0.02$) updating occurs at a very slow rate, despite the effort exerted (line C).

What is the benefit to a practitioner of imitation? The true, underlying, value of visiting either an ‘extraordinary’ or a ‘good’ practitioner is constant, but the expected value is rising and falling with p_n . The immediate effect of an increase in the expected value is not necessarily an increase in the price that can be charged,

but an increase in the number of visits. Clients weigh the perceived value of a visit against its cost, and as the value increases we expect more conditions to meet the threshold condition to initiate a visit. This effect is clearly visible in veterinary medicine. Ly (1999b) has found that clients purchase more, and a greater variety of services from practitioners in whom they have confidence than from other practitioners. It is possible that under certain market conditions the practitioner could extract higher fees from clients as beliefs about him changed, further increasing the returns to reputation. However, this *might* serve as signal to clients that a practitioner is *not* 'extraordinary', since 'extraordinary' practitioners appear to clients to be indifferent to market forces.

The strategy of imitating 'extraordinary' practitioners cannot be sustained in a finite game. Since the 'good' practitioner is not 'extraordinary', the game unravels from its end-period⁹. Even if the game has an infinite (or finite but unknown) horizon it is not always clear that it makes sense to play 'good'. The unknown horizon insures that there is always an expectation of future value from a 'good' reputation, but at every stage this future value must exceed the cost of being 'good'.

There is an opportunity cost to the fact that clients do not know if a practitioner is 'extraordinary' or 'good'. The area between any of the curves shown in Figure 1 and the line where $p_0 = 1$, for all n (for line A this area is shown as the shaded area) is a measure of the difference in revenues that could have been earned if patients knew quality was high and the revenues earned while reputation is gained. If a practitioner could assure patients he was of high quality he would not have to forgo these revenues. Imagine that the payoffs to reputation are such that all practitioners have an incentive to gain a reputation. If clients knew this, could p_0 become 1? Clients do not really care if a practitioner is 'good' or 'extraordinary', since both behave in the same manner. So could they potentially view all practitioners as being either 'extraordinary' or having incentives to imitate the 'extraordinary'?

Just as clients learn whether or not a practitioner is 'good' or 'extraordinary' they would also learn that a practitioner is 'ordinary' (when most practitioners are

⁹For discussion of the properties of repeated games see Axelrod (1985).

'good' perhaps a better term would be 'bad'.). An updating curve analogous to those shown in Figure 1 would begin at the initial prior and fall to zero. The area beneath such a curve would represent the benefit to being perceived as 'good' when in fact a practitioner was 'bad'. As the prior rose the area underneath the updating curve would also rise. However, every time a practitioner moved to a new area he could begin with a new prior belief of his type. The temptation to move into an area and then gain from the perception of being 'good' while in fact exerting no effort becomes larger as the initial prior rises.

An equilibrium in which large numbers of practitioners are 'good' and are earning income as if they were 'good' is sustainable if there is a sufficiently large cost to moving to a new area and starting again. From the point of view of a client the arrival of a new practitioner who could somehow bind himself to not moving would be a signal that this practitioner was 'good'. The easiest way to bind oneself (in a society in which courts do not function properly) is to invest in assets that are specific to one individual's particular practice (Williamson 1985). Specific assets serve as a type of bond, foregone if a practitioner leaves one practice to begin another.

What candidates for such specific assets exist? In a well functioning market there are few assets that are specific to one practitioner in one place. The purchase of medical equipment does not meet this test since it can easily be moved. The construction of a laboratory is similarly not specific since the facility could be sold to another practitioner. However when there are very few practices in a rural area the construction of a laboratory is more of a specific investment because there are few potential buyers and the building cannot be moved. (In the US MD's frequently purchased the practices of 'good' physicians; if they then were 'bad' the sale value of the practice would drop severely.

As we introduce potential solutions to imperfect information costs we must take into account the fact that setup costs for private practice in rural Africa are already seen as being too high. Though proper investment according to the above mechanism could also increase the returns to the practice of medicine, we must recognize that mechanisms that insure quality and thus higher revenues without requiring large

investment costs will naturally be more attractive. Thus there seems to be an important opportunity for other mechanisms to insure quality other than reputation.

More common on the continent is 'brand name' reputation, whereby individual missions and church denominations become known for high quality medical care. We have observed missions being quite deliberate in the way they use their names in medical markets and have witnessed the ability of the better ones to create instant customer demand for facilities in new locations through this device (similar to franchising)(K.L. Leonard, personal observation). In this way once a denomination has gained a reputation for being 'good' they can extend their reach to new areas without having to re-earn a reputation.

3.6.2 Repeated Interaction

As we have stated, contingent contracts with verifiable outcomes improve on the solution obtained with only fixed payments. Thus if a contingent-payment contract could be sustained, it is likely to lead to superior outcomes both in the short and long run for both client and practitioner. There is a twist with sustaining contingent contracts when the client agrees to pay the practitioner; it is the reputation of the client that matters, not that of the practitioner. Under what conditions does the client have incentives to truthfully reveal unverifiable outcomes even when it increases the payment she must make without increasing her utility from treatment?

MacLeod and Malcomson (1982) observe that many contracts are written on unverifiable outcomes and address our question.

[P]iece rate [contingent] contracts are actually used, in many cases even where the arrangements for recording output are sufficiently informal that verifying it in court would be difficult. Moreover, in some jobs a substantial part of pay takes the form of a performance-related bonus (in effect an informal piece-rate [contingent-contract]) which employees expect to be paid, and firms actually pay, despite there being no legal obligation to do so. . . . The fundamental requirement for an implicit contract to be self-enforcing is that there exist sufficient economic surplus from continuing it

over what the parties can jointly get if it is terminated.

Thus when the client faces a more costly outside alternative to the relationship with a particular practitioner she has an incentive to tell the truth. Recall that with observable but unverifiable outcomes the practitioner knows whether or not the client is telling the truth and can therefore refuse to do business (or refuse to enter into certain types of contracts) with her in the future.

The enforceability of the contract with the possibility of repeated interactions will depend on at least three things: the probability of needing the services of a practitioner in the future; the additional cost incurred in seeking a new practitioner (local monopoly increases this cost); and the value of the services provided (skill of practitioner, expected disease conditions and in the case of animals, the value of the animal). These factors help to explain why we observe contingent-payment contracts in use among veterinarians but not among doctors. Veterinarians often enjoy a wider geographical monopoly and transportation costs for animals are significantly higher than for humans (Woods 1999). Furthermore we predict that areas with high value animals or where herders have large herds (more interactions with the veterinarian per client) are more likely to sustain contingent-payment contracts.

The interesting finding is that practitioners who enjoy some form of local monopoly (even if only in their skill class) are more likely to be able to form contingent-payment contracts, because they know clients have less incentive to lie about outcomes. This increases their welfare, but it also increases the welfare of clients. Thus breaking a local monopoly without introducing any other form of institutional guarantee of quality could hurt, not benefit herders.

3.6.3 Budget-Breaking (non-conservative) Incentive Systems

Both of the above mentioned mechanisms allow an individual practitioner to get around some of the problems posed by agency, while avoiding problems of enforceability. However there exists a class of contracts that can both obviate the need for verifiability and approach, if not achieve, the full information solution. These require the existence of an individual other than the parties to the contract who can

break the budget balance of payments.¹⁰ We will suggest that this third party role is an important function of institutions.

The client lies when payment is required in an unverifiable contingent-payment contract because after the outcome is observed payment can only decrease her utility. If she says the outcome was bad, she doesn't change the outcome but makes no payment. If she says the outcome was good, she does not improve the outcome but must make a payment. One could ask the practitioner about the outcome but the same problem would occur.

Consider instead that there are two transfers, not one. The payment made (or not made) to the practitioner depends on the declaration of the client, but the payment made (or not made) by the client depends on the declaration of the practitioner. This requires a third party who can add or remove money when these two payments do not match. Now the statement of the client does not change her utility. She gets the value of the outcome and a payment that depends not on what she says but on what the practitioner says. She has no reason to lie, and if we assume a mild propensity to tell the truth she will tell the truth. By similar logic the practitioner is induced to tell the truth. Now the outcome will be known and payments can be made appropriately.

Imagine an arbitrator who cannot observe effort, but offers practitioner and client a scheme as we have outlined. Both practitioner and client declare the outcome to have been good and the arbitrator then requires the client to pay the practitioner P_X as set out in the original contract. The amount paid is exactly equal to the amount received. However, this scheme is a non-conservative or budget-breaking contract because it involves unequal payments off the equilibrium path. If the client had chosen to lie she would still have to make a payment (this depends on the declaration of the practitioner) but the practitioner would not receive the payment — this is an unbalanced set of payments. Thus even with unverifiable outcomes, if

¹⁰In the general literature these mechanisms are referred to as Groves mechanisms (Clarke 1971, Groves 1973). The role of budget-breaking in multiple agent problems was advanced by Hölmstrom (1982). Budget-breaking is also cited as a mechanism for insurance companies to reduce the costs of moral hazard in insurance by Ellis and McGuire (1990, pp 377).

(but only if) the budget does not have to be balanced a contract can be sustained that makes payment contingent on outcome.

Super-Principals and the Role of Institutions In order to break a budget a third party (whom we will refer to as the super-principal) is required who can inject or remove payments when necessary. The super-principal does not perform any action that contributes to treatment and does not need to observe the outcome of treatment. We suggest that this is the primary role of an organization that employs practitioners (both government and mission, human and animal health services).

Breaking the budget gets around the problem of verifiability but it also improves on the utility of both parties when both client and practitioner effort are unobservable: the full information solution can be achieved. A simple example of this is a contract where both client and practitioner receive the full value of the marginal outcome. The practitioner receives a contingent payment $P_X = \omega X - \omega \bar{\mu} p_h + P_F$, where P_F is a fixed payment and $\omega \bar{\mu} p_h$ is the expected value of the outcome when effort is exerted. The client, on the other hand makes a fixed payment of P_F and the value to the client of treatment is $\omega X - P_F$. Both client and practitioner face the marginal value of the outcome and therefore have high powered incentives to exert effort. Thus by being a member of an organization that can break budgets practitioners can achieve solutions superior to those of practitioners who are not members of such organizations.

Signal Technologies In this form of contract the payment to the client can be made without knowing the outcome of treatment. Since she retains the value of the treatment, that is her payment. Furthermore it is not necessary that payment to the practitioner be directly tied to the outcome of treatment. Any signal will achieve the same solution if it is correlated with the outcome of treatment such that a contract can be written that leads to the same incentives for the practitioner. For example, if the practitioner is paid, not on the basis of one outcome, but on the average outcome of two or more treatments, the payments can be designed such that he faces the same incentives.

This is an important modification to the class of potential contracts because, despite the fact that with broken budgets clients can be induced to tell the truth, there are significant transaction costs to gathering information about outcomes. Outcomes are frequently not known until well after the administration of treatment, and patients (animals or people) have generally returned home by this point. The presence of transaction costs requires that organizations choose between signal technologies that differ by correlation to outcomes and cost of collection.

Budget breaking institutions use signal technologies to reduce transaction costs but any technology available to an organization is also available to an individual practitioner or an organization that does not break budgets. Their use in this latter case, however, would introduce another transaction cost as well as forcing the client to bear extra risk. One potential signal is the outcome of a randomly chosen treatment from among a large set of treatments. This reduces transaction costs by reducing data collection requirements and if the practitioner does not know, a priori, which treatment will be selected and payments are appropriately scaled he faces the same incentives to provide effort as if clients were polled on outcomes for every treatment. Now, if budgets are balanced, each client must be approached after the chosen outcome is known and asked to make the appropriate payment. If the outcome was good each client will be asked to make the contingent-payment that was part of the original contract. If it was bad, money might be returned to them. Class action lawsuits in the US operate in basically this manner. This re-introduces, however, the exact transaction costs that this technology was supposed to help organizations avoid. Furthermore, except for the client whose treatment was randomly chosen the risk to which clients are exposed has increased¹¹. Potentially the client could be faced with both a bad outcome and the imposition of an extra payment, a double penalty.

Though observed institutions use both budget-breaking and outcome-correlated signals, we suggest that the comparative advantage of organizational provision comes in their ability to break budgets. The existing and potential institutions we discuss

¹¹We had assumed risk neutral clients but relax the assumption slightly here.

below differ in many aspects of their practices but all break budgets and collect correlated signals rather than information about outcomes.

Organizational Quality The technology of information collection most commonly observed in Africa and throughout the world is direct monitoring of the practitioner. The employer of the practitioner does not seek to know the outcome of treatment but does observe other outcomes that give information about the effort of the provider. Practitioners produce both health for the patient and what we call *organizational quality*. This second output is observed by the employer. Records are kept of the various activities that go into producing health. Typically a selection of records are examined during a site visit. The patients' symptoms and complaints are part of all records and therefore procedures and records should follow protocols developed for each set of complaints. If a particular record or collection of records is determined to be in violation of standards, the practitioner is punished in accordance with the gravity of the deviation. (One of the major steps forward in assuring medical quality in the US was the creation of the Joint Commission of Medical Accreditation standards for medical records (Savage 1997, Langlois and Savage 1997)).

Organizational quality is a measure specifically designed to be highly correlated with outcomes. It does, however, include some information that is deliberately different from outcomes. For example, prescribing too many drugs (poly-pharmacy) decreases organizational quality but usually does not affect individual health outcomes. Organizations monitor this activity in an attempt to 'internalize' the externality imposed by excessive use, for example, of antibiotics. In this way the signal of organizational quality might be superior to the signal provided by outcomes. On the other hand, organizational quality ignores patient effort. Examination of records would show that correct diagnostic procedures were used and that correct medicines were prescribed but they would be unable to show whether or not the patient took her medicines correctly. Thus a practitioner evaluated on the basis of organizational quality will be less likely to encourage patient effort than a practitioner evaluated

on the basis of outcomes.

This prediction is confirmed by the results of observations of 450 consultations in both church and government health facilities in Tanzania¹². In only 16% of observations did the doctor tell the patient what the diagnosis was or what medicine was being prescribed. In only 6% of cases did the doctor inform the patient of any activity she could undertake to increase the chance of recovery or to avoid a similar illness in the future. Of cases in which a dispensing nurse gave the patient drugs requiring that the patient know how or when to take the drug in only 32% did the nurse check to see if the patient had any idea how to do so. These observations confirm that despite the emphasis these organizations put on health education, organizational quality does not create strong incentives to encourage patient effort.

Non-Conservative Institutions for the Provision of Medicine We have stated that regulation by the state does not, and is unlikely to, function properly in Africa due to constraints on the administrative capacity of most African governments. Here we explore other institutions that can play a regulatory role. These institutions all break budget balance in that the rewards or punishment of practitioners are independent of the payments made by clients.

Self-Regulation Unlike France and many African countries that rely on the state to regulate the practice of medical professionals, Anglo-Saxon countries instead have relied upon self-regulation by professional associations, which not only control licenses but hospital privileges as well. Due in part to very small private markets and in part to statist traditions, professional self-regulation has been quite weak in Africa north of Zimbabwe. The agency literature on health discussed earlier in this paper predicts that as private medical markets emerge professional associations would become strong regulators so as to enable their members to collect the higher prices the market would pay for the assurance of quality effort (Pratt and Zeckhauser

¹²K.L. Leonard participated in research sponsored and carried out by Dr Mpuya under the auspices of the District Medical Office of Iringa, Tanzania with the support of Centro Universitario Aspiranti Medici Missionari.

1985, pp. 19-29; Arrow 1963, 1985, pp. 38, 40). Ly (1999a) examined this hypothesis in Senegal for our project. There was a special reason to expect that professional self-regulation would emerge for full doctors of veterinary medicine (DVMs) in that country; the down-sizing of the state had forced large numbers of veterinary practitioners into the market and the political power of the union representing veterinary para-professionals had won them the legal right to practice in competition with DVMs. Full vets therefore had a particularly strong need to convince the livestock producer that paying a premium for their services would purchase higher quality. DVMs did create a professional body that gained state recognition, but they have not made a meaningful effort to regulate one another. Ly concludes that the agency literature underestimated the collective action problem of such regulation. There is a premium to self-regulation but it is gained by all members of the profession. The creation of the necessary mechanism would require a level of effort from leading DVMs that would exceed the benefits they personally would obtain.

Referral Networks In our opening literature review, we noted the suggestion that referral networks fulfill much the same function as self-regulation by professional associations but in an informal and less costly manner. Such networks are essentially non-existent in Africa. Patients very often self-refer to higher-level facilities, the referrals that are made are to facilities and not to individual practitioners, and interaction between the referring practitioner and the specialist does not occur (Sauerborn et al. 1989, K.L. Leonard, personal observation). This observation does not preclude such an institution's emerging at some point in the future; it does function effectively in the United States. But it is not a part of the solution to the quality of effort problem in contemporary Africa. We will return to the potential for referral networks in the discussion of adverse selection.

Hierarchical Supervision Most of Africa therefore lacks institutional structures that would solve the imperfect information problem for medical markets as a whole. The problem can be and is resolved within the context of specific organizations, however, where hierarchical supervision can take the place of contracts and

professional regulation. Government and church-run (mission) health systems are therefore players within the institution we call organizational provision. Both governments and missions collect information on organizational quality, but government health systems fail to take advantage of their institutional context. Hierarchical supervision in government does not function effectively, as it is too difficult to fire practitioners for deficiencies, the personnel systems are either too bureaucratic or too politicized to reward good performance, and supervisors have too little incentive to overcome the inertia of the system.

Hierarchical supervision does work effectively to assure quality effort in the health facilities of non-governmental organizations (NGOs), most of which are run by churches or missions. Service tends to feature prominently among the preferences of the leadership of these NGOs, so it derives satisfaction from exercising the strong supervision needed to gain staff effort. These general NGO motives for quality are even stronger for church-related facilities, for church leadership expects to derive the tangible and intangible benefits of increased membership from having a strong service reputation and interacting with clientele using their facilities. These incentives and the supervision that follows from them assure a type of quality medical effort on the part of mission medical staff. This is an area in which Africa has a readily available, highly-functional institution for dealing with moral hazard.

It is very clear that mission health facilities have a strong market position in most African countries and that many people are prepared to pay a premium over government facility charges in order to use them. This quality premium can not be due to superior mission capacity, for staff qualifications and equipment in NGO facilities typically is equal to or worse than those in their government counterparts (Mliga 1999, Leonard 1999). It could be due, however, to aspects of observable effort, such as better drug supplies and superior 'clinical manner' (Mliga 1999, Bichmann et al. 1991, Litvack and Bodart 1993, Omorodion 1993, Stock 1983, Waddington and Enyimayew 1989). To demonstrate that mission facilities also are attractive because they help to solve the moral hazard problem, we need to show that a client is most likely to chose one of them as her agent when she knows that unobservable

dimensions of quality of medical effort, and not just capacity and observable aspects of effort, are desirable for her particular medical condition. To prove that missions serve as an institution to reduce the imperfect information problem we therefore also must examine the extent to which African patients have any knowledge of their medical needs.

3.7 Self-Diagnosis

The medical model almost certainly overstates the extent of the imperfect information problem; in Africa the patient does know something about her condition and what is appropriate to it. We have noted earlier that Africans are making choices between medical providers based on their own understanding of their illnesses. The question is not whether these choices are being exercised but whether they are based on some real knowledge and therefore are good choices.

If medical effort were observable patients would seek providers who had the appropriate skill at the lowest cost and then insure that the necessary effort was exerted. Government-run facilities are clearly the least expensive in terms of fees, drug costs and travel costs for the majority of patients. Therefore, if skills, capacity and drug supplies are adequate at government facilities, we would expect to see clients visiting government clinics and being referred to government hospitals when necessary, or, if the delay was substantial, self-referring to hospitals. Visits to mission clinics or hospitals would occur only when the difference in travel costs was such that it was less expensive to visit mission clinics or hospitals. However, in a study undertaken in Cameroun, deliberately chosen to coincide with a German Technical Agency project to insure adequate drug supply and facilities at government operated clinics and hospitals we found no evidence of patterns of visits along these lines (Leonard 1999, Ndeso-Atanga 1999). Seventy-nine percent of all visits were to a provider who was not the closest one. Of the visits to clinics and hospitals only 53% were to the closest clinic or hospital. Of visits to hospitals 39% involved patients traveling past a government hospital to visit a mission hospital.

If some elements of effort are unobservable then different patterns of visits can

emerge. We have stated that the contingent-payment contract of the traditional healer fails to achieve the full information solution, though it is an improvement on treatment without contracts. The contract available at organizational providers can potentially be superior to that of the traditional healer, but it downplays the role of patient effort in treatment. Eswaran and Kotwal (1985) theorize that three different forms of land-labor contracts (sharecropping, land rental and wage labor) co-exist because each form of contract is particularly appropriate for certain types of crops. They suggest that crops differ in the extent to which they require unobservable inputs from laborers and landowners. By analogy we can consider diseases to be like production technologies that differ in the degree to which they require unobservable inputs from medical practitioners and patients and then can show that different forms of contracts are appropriate for different diseases.

Expanding the formulation of the technology that produces health, μ , we introduce two new variables that differ by disease, the responsiveness of the disease condition to medical effort, γ , and the responsiveness of the disease condition to patient effort, δ . Given the existence of two different institutions between which to choose and the existence of two different organizations within the institution of 'modern' medicine, patients make their choices based on costs, as well as their own estimate of the value of these two parameters, γ and δ .

Leonard (1998b) assumes that the production of health takes the form of a Cobb-Douglas production function such that $\mu = (a\pi)^\gamma (b\eta)^\delta$ where $\delta + \gamma < 1$ and η is the skill level of the patient. γ and δ are quantitative values given for each disease condition from a medical evaluation of the symptoms and characteristics of the patient. The parameters of the various contracts can then be recovered from observed choices. Figure 2 shows the predictions of visits when the skill of various providers in the treatment of certain conditions as well as the skill of patients are held constant.

The results of this and other empirical investigations (Leonard 1998a) suggest that patients simulate a type of imperfect market for health care by choosing between various types of institutions for the provision of effort and skill. Thus clients are

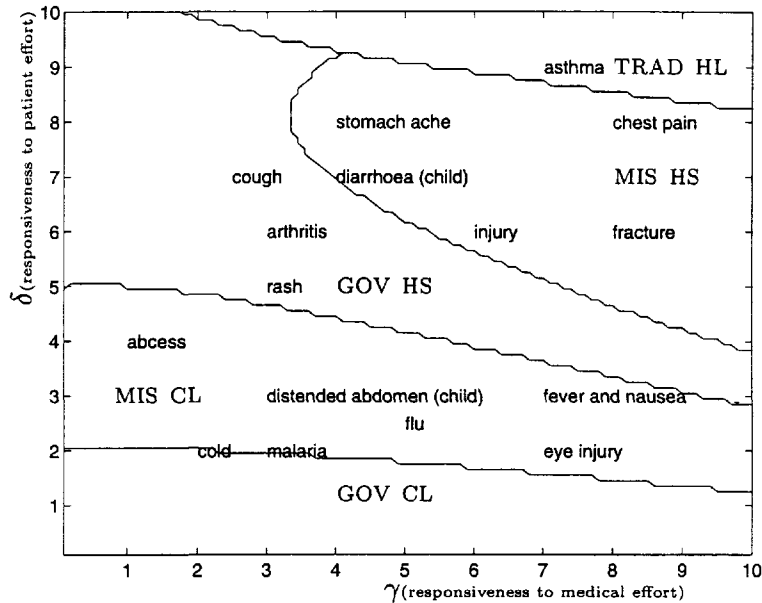


Figure 2: Two dimensional projection of disease characteristics

aware of the incentive effects of various institutional structures and act accordingly.

We do not suggest that poor, uneducated patients are *unfailingly* choosing the appropriate place to go for medical treatment. The African literature does suggest that self-referrals to hospitals, by-passing primary care facilities, are often wasteful. But the same literature also can be read as suggesting that self-referrals often are appropriate, that there is error on both sides (Mwabu 1986, Vogel 1993, p. 93). The fact that hospitals generally are more distant than clinics means that self-referrals are costly in time and transport for the sick, and therefore are unlikely to be chosen if they are believed to be unnecessary.

Moral hazard in the provision of health services creates agency costs for both the client and practitioner. When this problem is not resolved by either institution or contract we find practitioners who specialize in the sale of pharmaceuticals (observables) but do not use their skills as diagnosticians (unobservable) – a widespread phenomenon in the privatization of both veterinary and human health services in Africa. There exist some solutions to this problem that do not require institutions

(the credit/ contingent–contract system of veterinarians with local monopolies) or for which viable institutions have long existed (the contingent contract of traditional healers). There are, as well, many organizations that independently make use of the institution of hierarchical supervision to resolve these issues. However the extension of health care (or veterinary) services in Africa will rely on the existence of pervasive institutions that allow modern practitioners to be paid for the skills that they possess. We have shown that government supervision has been ineffective and that it is unlikely to become more effective. Self-regulation by professional associations has not become an effective institution and we see little reason to believe that it will become more so. We will discuss referral networks in greater detail in the section on adverse selection.

4 Adverse Selection

We turn now from the moral hazard problem to one that is related, especially in practice — adverse selection. In markets with imperfect information about skills, higher-cost quality providers have difficulty surviving. If they are unable to demonstrate the superior quality of their capacity and effort and to persuade patients of its relevance to their conditions, a ‘market for lemons’ will exist in which lower-price, low quality practitioners will compete them out of existence (Akerloff 1970).

Currently in rural Africa, where the presence of private practitioners of human medicine is small, clients tend to associate skill levels with the location in which one practices. Hospitals and their staff are differentiated from clinics and their staff. However, for private practitioners to succeed patients must be able to discern their skill without the aid of an established organizational setting. Whether this is the only block to the establishment of more extensive private medical practice is not clear, but resolving this issue has already become a top priority of private veterinary medicine.

Unlike practitioners in human health, veterinarians either visit herders in the field or work on their own away from buildings and physical facilities that might

suggest skill levels. Our project's studies of the emerging private veterinary markets in Senegal and Uganda show that full vets (DVMs) are having difficulty competing effectively against their para-professional counterparts (Ly 1999a, Koma 1999). In Uganda DVMs enjoyed no price premium over para-vets when they performed the same procedures. Their sole market advantage came when they provided services or did operations of which their less-qualified competitors were incapable, despite the fact that their training as diagnosticians should increase the value even of these simpler services.

4.1 Solutions

If clients can correctly identify the skill level of a given practitioner and *that additional skill is worth its extra cost* both the practitioner and the client will gain from its identification. The returns to skill appear to be high enough in human health care to justify its use. The value of skill in veterinary medicine, however, differs from region to region (with varying market prices for meat and dairy products and the type of animals). Still, even in the areas with the highest value for animals, when appropriate institutional guarantees of skill are absent, higher skill professionals suffer, whereas in areas with proportionately lower valued animals the existence of institutions that solve the problem of adverse selection lead to higher uses of professionals (Ly 1999a, Koma 1999).

In this section we investigate some possible solutions to the adverse selection problem and discuss their potential in the African context.

4.1.1 Labeling

The most common way in which quality providers seek to prevent adverse selection and establish a market share for themselves is through the way in which they present themselves to the public. Professional associations publicize the value of certain kinds of qualifications, certify those who possess them and prosecute those who falsely claim them, while the individual practitioner advertises that he is so qualified. Though labeling can involve public education of the benefit of qualifications, its most

important role is to allow differentiation. If skill has benefit any label attached to a high skilled practitioner that cannot be borne by a low skill practitioner will mitigate the adverse selection problem.

This version of labeling has not worked well in Africa, for subsidized medical organizations have had a motive to obscure these distinctions, as we have seen. This deficiency is reversible, but only with considerable effort by professional associations, which we have already characterized as suffering from collective action problems in Africa. Even if differences in medical qualifications were better known to the public, however, they would signal only differences in capacity, not the medical effort which we have argued is so important.

4.1.2 Signaling

A second way in which providers can counter the tendency to adverse selection in markets characterized by imperfect information is by ‘signaling’ their higher quality. This mechanism involves doing or showing things that high skill practitioners find easy (inexpensive) and inferior ones find difficult (expensive.) Spence (1974) suggested a form of game in which workers invest in education, not to improve their productivity, but to signal their productivity. If education is more expensive for the low ability type than for the high ability type (schooling has a greater disutility), there can exist an equilibrium in which only high ability types seek additional education and employers can know by looking at the level of education exactly the type of worker they are hiring and pay accordingly — even if education itself serves no productive purpose in the workplace.

The use of observable procedures that are associated only with higher levels of training and expertise, such as surgery, are one example. It would be difficult for a para-professional to learn to open and close an incision successfully in the presence of witnesses, whereas a full medical professional has already been taught how to do it. Koma (1999) did find that as private veterinary practice emerged in Uganda full vets would be known for their ability to do surgery, even among livestock owners who could not tell how their credentials differed from those of

para-professionals. Medical practitioners who want to establish a reputation for skill therefore would find it desirable to occasionally perform surgery in a highly visible setting. To do so on humans (but not animals) requires access to a hospital, and thus hospital privileges become another signaling device. Note that in this example, the role of hospital access is not to provide additional revenue, but to provide information to potential clients — who do not need surgery — about the skill of the practitioner. Another example of signaling is the use of microscopes to examine blood slides and the presence and use of other expensive pieces of medical equipment. This mechanism is visible in the de-facto privatization of the Chinese medical system, where county hospitals that are financially hard-pressed have been observed to invest in expensive new equipment in order to attract patients away from other facilities (Chen 1997, Tang 1997). Some Kenyan vets are sought out by pastoralists because of their use of sophisticated diagnostic procedures (D.K. Leonard, personal observation). Koma (1999) concluded that the failure of full vets in Uganda to purchase and use the expensive diagnostic tools in which they had been trained was one of the sources of their difficulty in establishing market position and avoiding adverse selection.

4.1.3 Separating Contracts

It is possible for the client to offer a contract that induces practitioners to truthfully declare their type, or skill level. Assume $\mu = \pi \cdot a$, $\pi \in \{1, 2\}$, $a \in \{1, 2\}$, and $c(a) = a$. There are two types of practitioners, those with low and those with high skill. An element of a is observable and therefore can be purchased in a spot contract. Thus at least one unit of a is provided, but the second unit is unobservable. The client does not know the skill level of the provider but offers the following two contracts to practitioners, designed such that the high skill practitioner will always choose the second contract and the low skill practitioner will always choose the first contract. The first contract is a fixed fee contract, $P = 1$, and the second contract is a contingent-payment contract $P = 1 + P_X$. The contingent-payment contract is designed by looking at the payoffs of practitioners, which are as follows for both

skill types

$$\underbrace{P_X \left(\left(\frac{2\pi}{\bar{\mu}} \right) p_h + \left(1 - \frac{2\pi}{\bar{\mu}} \right) p_l \right) - 2}_{\text{exert effort}} \geq \underbrace{P_X \left(\left(\frac{\pi}{\bar{\mu}} \right) p_h + \left(1 - \frac{\pi}{\bar{\mu}} \right) p_l \right) - 1}_{\text{don't exert effort}} \quad (6)$$

$$P_X \frac{\pi}{4} (p_h - p_l) \geq 1$$

Thus if P_X is chosen to fall strictly within this range

$$\frac{2}{p_h - p_l} < P_X < \frac{4}{p_h - p_l} \quad (7)$$

the high skill practitioner will benefit by the second contract and the low skill practitioner will lose by it. We do not expect one client to offer to an unknown practitioner a choice of contracts. Rather, using animal care as an example, we predict that herders with valuable animals that have unfamiliar ailments or require complicated procedures will offer a contingent contract and attract high skill practitioners and that otherwise herders will offer a fixed fee contract and attract practitioners with lower levels of skill. (The latter is clear in Uganda from Koma (1999), and the former is strongly implied by personal field observations of David Leonard and George Njiru in Kenya.)

4.1.4 Referral Networks

The fourth solution involves the development of meaningful referral networks. We noted earlier that these are not functioning effectively in Africa at present. Nonetheless they are the mechanism whereby most real professional self-regulation and enhancement of the quality of practice takes place in the United States (Freidson 1970, Savage 1997). Since they are networks of participating professionals rather than organizations, their creation can be more decentralized and accomplished at lower cost than many of the other moral hazard and adverse selection solutions we have discussed in this paper. It therefore is worth exploring further how professional referral networks work and what the prospects may be for their emergence in Africa. Freidson (1970) found in the US that physicians tend to develop networks of relations

with others similar in quality to themselves. Because medical personnel are better trained to understand the nature of one another's work and because they see many more examples of the outcomes of their colleagues' treatments, they are in a much better position to overcome the imperfect information problem of medicine than is the average patient and to evaluate accurately the technical capacity and quality of effort of their fellow practitioners. Thus if frequent interaction teaches the patient that their primary care physician provides a quality of care that they regard as satisfactory, they can use that MD's referrals to find other medical providers of similar quality. Haas-Wilson (1990) shows that (for psychotherapy services) referred practitioners earn higher fees, a result compatible with the idea that referrals identify superior practitioners.

Savage (1997) shows that these networks perform functions for professionals beyond providing mutual referrals of business. In fact professionals learn from one another in their network relationships and through their interactions collectively become more competent — testing insights and learning new procedures and developments in the field. She calls this creation of new knowledge 'professional co-production' and argues that it represents one of the distinct advantages of professional knowledge systems. These self-organizing referral networks are rarely visible in African health and veterinary systems, however. There seem to be several reasons why this is so.

(a) Historically veterinary systems were subsidized and human health systems still are. This created an excess demand for the more highly qualified practitioners. The response was to blur distinctions among personnel, so that the cheaper staff, with lower qualifications, could take up the slack. There is an attempt to get patients to enter these system at the level of the cheaper providers, where the simpler cases can be filtered out. The function being fulfilled here is an efficiency, not a quality enhancing one (Vogel 1993, p. 93). When referrals are made they are made to facilities, not to particular practitioners. Given the depersonalized character of these relationships, there is very little personal interaction or even communication among medical practitioners about the cases they have shared. Thus there is no learning,

no 'professional co-production' of knowledge, and little information generated that might teach personnel about one another's quality.

(b) To the extent that referrals take place across levels in the government and mission systems (health and veterinary) they are between subordinates and those who are their supervisors or are closely connected to them. It is a well-established principle of organizational behavior that communication takes place more readily and accurately between peers than up a hierarchy (Blau 1955).

(c) The sharing of knowledge (both about medicine and about one another) takes place most effectively through informal and frequent personal interactions. Large numbers of professional staff in the African health and veterinary systems are working in isolation. Hospitals don't exist for veterinary medicine, (human) health practitioners in clinics don't have hospital privileges, and rural hospitals tend to have very small professional staffs. Until medical professionals in Africa interact more with one another, in non-hierarchical settings, and with the possibility of personalized referrals, the professional networks that recognize, reward and encourage quality are not likely to emerge.

It does seem possible that the increased commercialization of African medical systems may cause this to occur. Since such networks can be organized in decentralized, low-cost ways, they seem more likely to emerge than self-regulation by professional associations, which is troubled with collective action problems. Particularly in veterinary medicine, where both DVMs and para-vets have been thrown into the private market, it seems reasonable to hope that such networks will be created. DVMs are finding it difficult to compete with para-vets, for it is irrational for the livestock owner to pay them extra to do something a para-vet can do. They need para-vets to refer to them the cases that need higher order skills, and the para-professionals will want to serve their clients with better quality referrals. DVMs will eventually realize that they can't make a living by doing procedures that are appropriate to para-vets and will want to have a set of these para-professionals to whom they can recommend their clients as well. At the same time, as time progresses the para-vets will begin to suffer from being cut off from further developments in their

fields and would benefit from the opportunity to advance themselves professionally by learning from DVMs. Thus mutually symbiotic relationships that enhance quality should and could emerge at low cost. They have not yet done so in Uganda (Koma 1999), but the evolution of the veterinary market there is still young. Ly (1999b) did observe such a network between a DVM and para-vets created by a mission in Senegal's Ferlo. The result of the assurance of quality that comes from a well-functioning referral network was at least a doubling in herder purchase of preventive treatments from para-vets.

It will be harder for the same spontaneous development of referral networks to take place in human health, where large, subsidized organizations dominate medical provision. More organizational effort will be required in these settings, but we believe that as the subsidy element of these systems declines, particularly in the mission ones, the competitive pressures to provide quality service to one's clientele will make them increasingly likely.

Of the solutions mentioned above to the problem of adverse selection, two (signaling and separating contracts) can be sustained by an individual practitioner. We are not aware of conclusive evidence to show that such schemes are being used in Africa (especially in the context of veterinary medicine), but we have observed behavior compatible with signaling. Labeling and referral networks require some level of coordinated effort on the part of practitioners, though referral networks should require less organization. Because the hurdle for organization is lower for referral networks we are more likely to see successful attempts at this in the future.

5 Conclusion

Imperfect information in the provision of health services creates agency costs for both the client and the practitioner. When this problem is not resolved by either institution or contract we find (as we do currently in African privatization) that many practitioners are specializing in the sale of pharmaceuticals (observables) but not using their skills as diagnosticians (unobservable). Autonomous and id-

iosyncratic solutions to the agency problem in the provision of professional medical services abound in rural Africa. Some do not require institutions (the credit/contingent-contract system of veterinarians with local monopolies) and others are built on viable institutions that have long existed (the contingent-contract of traditional healers). However, they depend in large part on the existence of a particular exogenous technological or institutional environment and therefore are difficult to replicate. Furthermore, since client effort is important in the treatment of diseases, these solutions are generally sub-optimal, suggesting a role for better institutional mechanisms.

We identify, through theory and comparison to the west, three forms of institutions that might do a better job of solving the problem. We dismiss government regulation because the regulatory capacity does not exist in most African countries. Self-regulation, by professional bodies, is a very attractive option and theory suggests that such regulation should arise as privatization continues. However, empirical evidence suggests that this conclusion is over-optimistic. There are significant collective action problems to the formation of self-regulatory bodies and since these organizations require (by their very definition) centralized control, we are pessimistic that these problems can be overcome in the present African context. On the other hand, referral networks perform much the same function but do not require centralized control. For this reason, we hold out the hope that they will yet spontaneously form, or that they can be coaxed into existence.

The role of Missions The most successful institution for the delivery of quality medical care in Africa is that of independent, pre-existing value-based organizations that use the institution of hierarchical supervision to insure quality. This result parallels a large body of empirical and theoretical work on health care in the US, where non-profit hospitals frequently arise in discussions of quality. Many authors have suggested that non-profit administrators maximize over different variables than for-profit administrators (Newhouse 1970, Pauly and Resisch 1973). We certainly believe that the bishop of a church does not operate health services with the simple

goal of earning profits. However, in Africa, we do not have enough for-profit hospitals to draw any conclusions about the role of missions specifically as non-profits.

The existence of missions hospitals and health services is an invaluable resource that should be tapped to its fullest potential and encouraged to grow. Many African governments implicitly or explicitly block the expansion of these services, policies that should be reconsidered. Some countries, such as Tanzania, effectively subcontract government subsidized care to missions, which are permitted to charge fees but sell their services at below market prices. This arrangement makes effective use of the missions ability to supervise their staff more effectively than government does its. A more aggressive policy would be to allow and encourage mission services to franchise by selling their name to individual practitioners. Such a policy should require that only missions that operate health services can sell their name (otherwise they would have no capacity to regulate) and that the person to whom the name is sold *must* publicize his relationship to the mission (giving strong incentives to regulate).

References

- Abel-Smith, B. and P. Rawal, "Can the Poor Afford 'Free' Health Services? A Case Study of Tanzania,," *Health Policy and Planning*, 1992, 7, 329–41.
- Akerloff, George A., "The market for "Lemons": Quality uncertainty and the market mechanism," *Quarterly Journal of Economics*, 1970, 84, 488–500.
- Arrow, Kenneth J., "Uncertainty and the Welfare Economics of Medical Care," *American Economic Review*, 1963, 53, 941–973.
- , "The Economics of Agency," in J. Pratt and R. Zeckhauser, eds., *Principals and Agents*, Cambridge, MA: Harvard Business School, 1985.
- Axelrod, R., *The Evolution of Cooperation*, New York: Basic Books, 1985.
- Baerts, Martine, *Guerisseurs et plantes medicinales de la region des cretes Zaire-Nil au Burundi* number 18. In 'Annales Science Economiques.', Tervuren, Belgie: Koninklijk Museum voor Midden-Afrika, 1989.
- Bichmann, W. et al., "District Health Systems: Users' Preferences for Services in Benin," *Health Policy and Planning*, 1991, 6, 361–70.
- Blau, P.M., *The Dynamics of Bureaucracy*, Chicago: University of Chicago Press, 1955.
- Bloom, G. and A. Wilkes, eds, *Health in Transition: Transforming China's Rural Health Services*, Vol. 28 of *IDS Bulletin*, IDS, 1997.
- Chen, J., "The Impact of Health Sector Reform on County Hospitals." In Bloom and Wilkes, eds (1997) pp. 48–52.
- Clarke, E., "Multipart pricing of public goods," *Public Choice*, 1971, 8, 19–33.
- Conco, W. Z., "The African Bantu practice of medicine: Some preliminary observations," *Social Science and Medicine*, 1972, 6 (3), 283–322.
- Coppo, P. et al., "Perceived Morbidity and Health Behavior in a Dogon Community," *Social Science and Medicine*, 1992, 34, 1227–35.
- Dranove, David and William D. White, "Agency and the Organization of Health Care Delivery," *Inquiry*, 1987, 24, 405–415.
- Edwards, S.D, *Indigenous African healers and western trained psychologists: a comparative study of interview methods* number 38. In 'Publication series of University of Zululand, series B.', Kwa Dlangezwa, South Africa: University of Zululand, 1983.
- Ellis, Randall P. and Thomas G. McGuire, "Optimal Payment Systems for Health Services," *Journal of Health Economics*, 1990, 9, 375–396.
- Eswaran, Mukeshi and Ashok Kotwal, "A Theory of Contractual Structure in Agriculture," *American Economic Review*, 1985, 75, 352–367.
- Freidson, E., *Profession of Medicine: A Study of the Sociology of Applied Knowledge*, New York: Harper and Row, 1970.
- Gelfand, M. et al., *The Traditional Medical Practitioner in Zimbabwe: His principles of practice and pharmacopoeia*, Harare, Zimbabwe: Mambo Press, 1985.
- Grossman, Michael, "On the Concept of Health Capital and the Demand for Health," *Journal of Political Economy*, 1975, 80, 223–255.
- Grossman, Sanford and Oliver Hart, "An Analysis of the Principal Agent

- Problem," *Econometrica*, 1983, 51 (1), 7-45.
- Groves, T.**, "Incentives in Teams," *Econometrica*, 1973, 41, 617-631.
- Haas-Wilson, Deborah**, "Consumer Information and Providers' Reputations: An empirical test in the Market for Psychotherapy," *Journal of Health Economics*, 1990, 9, 321-333.
- Hart, Oliver and Bengt Hölmstrom**, "The Theory of Contracts," in Truman F. Bewley, ed., *Advances in Economic Theory: Fifth World Congress*, number 12. In 'Econometric Society monographs.', Cambridge: Cambridge University Press, 1987.
- Hölmstrom, Bengt**, "Moral Hazard in Teams," *Bell Journal of Economics*, 1982, 13, 324-40.
- Koma, L.M.P.K.**, "Can Private Veterinary Services Survive in Uganda?" In Leonard, ed (1999) chapter 6.
- Korse, Piet et al.**, *Sorcellerie, maladie et chasseur de sorciers chez les Mongo (Rep. du Zaïre)* number 103. In 'CEEBA, series 11.', Bandundu, Zaïre: Publications (CEEBA), 1989.
- Kreps, David M. and Robert Wilson**, "Reputation and Imperfect Information," *Journal of Economic Theory*, 1982, 27, 253-279.
- Langlois, R. N. and D. Savage**, "Standards, Modularity, and Innovation: The Case of Medical Practice," 1997. Paper presented at the Conference on Path Dependence and Path Creation, Copenhagen.
- Lasker, Judith N.**, "Choosing among therapies: illness behavior in the Ivory Coast," *Social Science and Medicine*, 1981, 15A (2), 157-68.
- Leonard, David K.**, "The Supply of Veterinary Services: Kenyan Lessons," *Agricultural Administration*, 1987, 26, 219-36.
- , *African Successes: Four Public Managers of Kenyan Rural Development*, Berkeley: University of California Press, 1991.
- , ed., *Africa's Changing Markets for Human and Animal Health Services*, London: Macmillan, 1999.
- Leonard, Kenneth L.**, "African Traditional Healers: Incentives and Skills in Health Care Delivery," Discussion Paper Series 9798-13, Columbia University 1998.
- , "Institutional Structure of Health Care in Rural Cameroun: Structural Estimation of Production in Teams With Unobservable Effort," Discussion Paper Series 9798-16, Columbia University 1998.
- , "The Role of Incentives in the Delivery of Rural Health Care in Cameroun." In Leonard, ed (1999) chapter 4.
- Litvack, J. and C. Bodart**, "User Fees Plus Quality Equals Improved Access to Health Care: Results of a Field Experiment in Cameroon," *Social Science and Medicine*, 1993, 37, 369-83.
- Ly, C.**, "Veterinary Professionals in Senegal: Allocation of Priorities and Working Behavior." In Leonard, ed (1999) chapter 7.
- , "Veterinary Services Delivery to Livestock Producers: Management and the Impact of Auxiliaries on Pastoral Production in Senegal." In Leonard, ed (1999) chapter 9.

- MacLeod, W. Bentley and James M. Malcomson**, "Implicit Contracts, Incentive Compatibility, and Involuntary Unemployment," *Econometrica*, 1982, 57, 447-480.
- Majok, Aggrey A. and Calvin W. Schwabe**, *Development among Africa's Migratory Pastoralists*, Westport, Conn.: Bergin and Garvey, 1996.
- Mirlees, J.**, "The Theory of Moral Hazard and Unobservable Behavior - Part I," Mimeo, Nuffield College, Oxford 1975.
- Mliga, G.M.**, "Decentralization and the Quality of Health Care." In Leonard, ed (1999) chapter 8.
- Mwabu, G.M.**, "Health Care Decisions at the Household Level: Results of a Rural Health Survey in Kenya," *Social Science and Medicine*, 1986, 22, 315-19.
- Ndeso-Atanga, Sylvester**, "Health Care Quality and the Choice of Care Providers in Rural Cameroon." In Leonard, ed (1999) chapter 5.
- Newhouse, J.S.**, "Toward a Theory of Nonprofit Institutions," *American Economic Review*, 1970, 60, 64-74.
- Normann, Hans**, *A selected bibliography of references on indigenous theory and practice and related literature* Indigenous Theory and Practice Project, Cape Town, South Africa: University of Cape Town, 1990.
- North, Douglass C.**, *Institutions, Institutional Change and Economic Performance*, Cambridge: Cambridge University Press, 1990.
- Nougbara, A. et al.**, "Assessment of MCH Services Offered by Professional and Community Health Workers in the District of Solenzo, Burkina Faso. 1 Utilization of MCH Services," *Journal of Tropical Pediatrics*, 1989, 35, 2-9.
- Omorodion, F.I.**, "The Socio-Cultural Context of Health Behavior Among Esan Communities, Edo State, Nigeria," *Health Transition Review*, 1993, 3, 125-36.
- Oyenyé, O.Y. and I.O. Orubuloye**, *Some aspects of Traditional Medicine in Bendel State of Nigeria: An exploratory study* number 14. In 'Niser monograph series.', Ibadan, Nigeria: Nigerian Institute of Social and Economics Research, 1985.
- Parsons, Talcott**, *The Social System*, Glencoe, Ill: The Free Press of Glencoe, 1951.
- Pauly, Mark V.**, "Overinsurance and Public Provision of Insurance: The Roles of Moral Hazard and Adverse Selection," *Quarterly Journal of Economics*, 1974, 88, 44-62.
- and **Mark A. Satterthwaite**, "The Pricing of Primary Care Physicians' Services: A Test of the Role of Consumer Information," *Bell Journal of Economics*, 1981, 12, 488-506.
- and **Michael Resisch**, "The Not-For-Profit Hospital as a Physicians' Cooperative," *American Economic Review*, 1973, 63, 87-99.
- Pratt, J. and R. Zeckhauser**, *Principals and Agents*, Cambridge, MA: Harvard Business School, 1985.
- Satterthwaite, Mark A.**, "Consumer Information, Equilibrium Industry Price, and the Number of Sellers," *Bell Journal of Economics*, 1979, 10, 483-502.
- Sauerborn, R. et al.**, "Low Utilization of Community Health Workers: Results from a Household Interview Survey in Burkina Faso," *Social Science and*

- Medicine*, 1989, 29, 1163-74.
- Savage, D.**, "Self-Regulation and the Professions," Technical Report, Institution for Social and Policy Studies, Yale University, New Haven 1997.
- Schwabe, C.W.**, *Veterinary Medicine and Human Health*, third ed., Baltimore: Williams and Wilkens, 1984.
- Spence, A. Michael**, *Market signaling: informational transfer in hiring and related screening processes*, Cambridge: Harvard University Press, 1974.
- Staugård, Frants**, *Traditional Medicine in Botswana*, Vol. 1, Gaborone, Botswana: Ipelegeng Publishers, 1985.
- Stock, R.**, "Distance and the Utilization of Facilities in Rural Nigeria," *Social Science and Medicine*, 1983, 17, 563-70.
- Tang, S.**, "The Changing Role of Township Health Centres." In Bloom and Wilkes, eds (1997) pp. 39-47.
- Vogel, R.J.**, *Financing Health Care in Sub-Saharan Africa*, Westport, Conn: Greenwood Press, 1993.
- Waddington, C.J. and K.A. Enyimayew**, "A price to pay: The impact of user charges in Ashanti-Akim district, Ghana," *International Journal of Health Planning and Management*, 1989, 31, 17-47.
- Williamson, Oliver**, *The Economic Institutions of Capitalism: Firms, Markets, and Relational Contracting*, New York: The Free Press, 1985.
- Woods, P.S.A.**, "The Importance of Proximity, Transport and Gender as Transaction Costs in the Use of Veterinary Services in Zimbabwe." In Leonard, ed (1999) chapter 3.
- Yoder, R.A.**, "Are People Willing and Able to Pay for Health Services?," *Social Science and Medicine*, 1989, 29, 35-42.

A A Proof

Proposition: With unobservable efforts some output sharing between patients and practitioners leads to a strictly superior expected outcome.

We simplify $\omega\mu(p_h - p_l)$ to μ without loss of generality for simplicity of exposition. The problem then is to maximize the expected net outcome (net of disutilities of effort) subject to the incentive compatibility (I.C.) and individual rationality constraints (I.R.) of the patient and the practitioner.

$$\begin{aligned}
 & \max_{a,b} \mu - c(a) - d(b) - T \\
 & \text{subject to} \\
 & a^* \in \operatorname{argmax}_a r\mu - c(a) + T \quad (\text{I.C. medical}) \\
 & b^* \in \operatorname{argmax}_b (1-r)\mu - d(b) - T \quad (\text{I.C. patient}) \\
 & r\mu c(a) + T = V_m \quad (\text{I.R. medical}) \\
 & (1-r)\mu - d(b) - T > V_p \quad (\text{I.R. patient})
 \end{aligned}$$

where T is the transfer from the patient to the practitioner,

The use of the LDFC condition allows us to ‘relax’ the incentive compatibility constraints and characterize them as first order conditions. Assuming the individual rationality constraint of the practitioner is binding the lagrangian is formed as follows;

$$L = \mu - d - c + \lambda_1 \left(r \frac{\partial}{\partial a} \mu - \frac{d}{da} c \right) + \lambda_2 \left((1-r) \frac{\partial}{\partial b} \mu - \frac{d}{db} d \right)$$

Using the envelope condition we solve for the change in utility with respect to a change in the sharing rule.

$$\frac{\partial L}{\partial r} = \lambda_1 \frac{\partial}{\partial a} \mu - \lambda_2 \frac{\partial}{\partial b} \mu$$

differentiating the lagrangian with respect to a and b we solve for the two multipliers.

$$\lambda_1 = - \frac{\frac{\partial}{\partial a} \mu - \frac{d}{da} c + \lambda_2 \frac{\partial^2}{\partial a \partial b} \mu - \lambda_2 \left(\frac{\partial^2}{\partial a \partial b} \mu \right) r}{r \frac{\partial^2}{\partial a^2} \mu - \frac{d^2}{da^2} c}$$

which then leads to an expression of λ_2 .

$$\lambda_2 = - \frac{\left(\frac{\partial}{\partial b} \mu - \frac{d}{db} d - \frac{r \left(\frac{\partial^2}{\partial a \partial b} \mu \right) \frac{\partial}{\partial a} \mu}{r \frac{\partial^2}{\partial a^2} \mu - \frac{d^2}{da^2} c} + \frac{r \left(\frac{\partial^2}{\partial a \partial b} \mu \right) \frac{d}{da} c}{r \frac{\partial^2}{\partial a^2} \mu - \frac{d^2}{da^2} c} \right)}{\left(- \frac{r \left(\frac{\partial^2}{\partial a \partial b} \mu \right)^2}{r \frac{\partial^2}{\partial a^2} \mu - \frac{d^2}{da^2} c} + \frac{r^2 \left(\frac{\partial^2}{\partial a \partial b} \mu \right)^2}{r \frac{\partial^2}{\partial a^2} \mu - \frac{d^2}{da^2} c} + \frac{\partial^2}{\partial b^2} \mu - \left(\frac{\partial^2}{\partial b^2} \mu \right) r - \frac{d^2}{db^2} d \right)}$$

Note that

$$\frac{\partial}{\partial b} \mu - \frac{d}{db} d = e \frac{\partial}{\partial b} \mu$$

and it should be clear that $\lim_{r \rightarrow 0} \lambda_2 = 0$. This makes sense since all of the incentives to provide effort lie with the patient when the sharing rule is equal to zero the cost of the incentive compatibility constraint is smaller and smaller. However λ_1 does not go to zero as r goes to zero.

$$\lim_{r \rightarrow 0} \lambda_1 = (1 - r) \frac{\frac{\partial}{\partial a} \mu}{\frac{d^2}{da^2} c}$$

Substituting both multipliers into the envelope condition we find

$$\lim_{r \rightarrow 0} \frac{\partial}{\partial r} L = \frac{\left(\frac{\partial}{\partial a} \mu \right)^2}{\frac{d^2}{da^2} c}$$

This condition is positive given that there are increasing marginal costs to effort. Note that the system is not solvable if this is not the case. Thus as r approaches zero there is a benefit to increasing r . Note that this does not rely on any assumptions about the sign of the cross partial of output with respect to a and b .

1998-99 Discussion Papers
Economics Department, Columbia University

Series no.	Title	Authors
9899-01	Propensity Score Matching Methods for Non-experimental Causal Studies	Dehejia, R.
9899-02	Institutional Solutions to the Principal-Agent Problem in African Health Care	Leonard, K. Leonard, D.