Asymmetric Information and the Role of
NGOs in African Health Care

by

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

July 1999

Asymmetric Information and the Role of NGOs in African Health Care*

Kenneth L. Leonard†   David K. Leonard‡

July, 1999

Abstract

In African health care the "miracle of the market" has not occurred. Patients exhibit willingness to pay for health care and yet practitioners are unable to sell their services. Simultaneously non-governmental organizations (NGOs) are running successful health facilities for which patients are willing to pay.

We develop a model of the demand for health care in the presence of asymmetric information that allows us to view African health care in the framework of the New Institutional Economics literature. We use previously published empirical results to support the validity of this view and show that NGOs have the institutional capacity to deliver high quality health care, whereas private practitioners, even with good intentions, will not easily succeed.

Having arrived at the well documented conclusion that NGOs provide high quality services through theory allows us to draw policy conclusions on ways to extend the provision of health services.

JEL Classification: D2 I1 O2
Keywords: Institutional Economics, Africa, Rural Health Care, NGOs, Non-Governmental Organizations, Not–for–Profit.

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 are troubled with imperfect information. The patient cannot assess the appropriateness and quality of the activities of the health practitioner. Therefore it is no surprise that free markets and spot contracts do not lead to an efficiently functioning market for health care.

There are many potential solutions to problems of asymmetric information. With a few important exceptions, however, the only consistently functioning solution in Africa is non-governmental organization (NGO) health systems. Science-based private practice is not well established and government services are almost universally of poor quality, despite the fact that NGO personnel are hired from the same medical schools and often paid less than their government counterparts. The success of NGOs is not a new result; however we explain their success by appeal to theory. This allows us to make policy conclusions about asymmetric information in the market for health care in general, the possibility for private markets and the future role of NGOs in this market.

Throughout most of Sub-Saharan Africa there is very little private (for profit) provision of health care, despite great cut-backs in government provision (Bennett et al. 1997, Hanson and Berman 1998). Most countries have no private sector of which to speak. Some countries have done better; in Kenya 23% of health facilities are private and in Ghana this figure reaches 40%. However, even these facilities do not appear to be serving the expected function in health care. "In Kenya 94 per cent of private for-profit facilities are below the health center level, compared with 79 per cent in the mission sector and 60 per cent in the public sector" (Berman et al. (1995a) as reported in McPake (1997, p. 26)). It appears that private providers most often are selling drugs or clinical tools, but not diagnostic services; they are acting as pharmacists not doctors. This is confirmed by the disparity between hospital beds

and percentage of total expenditure. In Zambia, for example, 0.2% of hospital beds but 22% of total health expenditure is in the private sector (Berman et al. (1995b) as reported in McPake (1997, p. 25)). In this respect, at least, privatization reforms have led to the proverbial "race to the bottom," where private providers exist only in small profitable market niches and the majority of health care needs are not being satisfied by this sector.

At the same time however, NGOs provide significant portions of health care services in Africa and are generally seen as being of high quality. Ghanaian NGOs provide 40% of clinical care needs, 27% of hospital beds and 35% of outpatient services. In Tanzania NGOs provide half of all hospitals and beds and receive half of all curative visits. In Zimbabwe they supply 35% of all hospital beds and 96% of NGO facilities are located in disadvantaged rural areas. (Gilson et al. 1997). These providers are not serving only small market niches.

Arrow (1963, pg 965) is the first among many authors to suggest that non-profit hospitals serve a special role in the health care market and that this is due to asymmetric information. He suggested that patients prefer organizations that are not maximizing profits when they are unable to judge quality directly, supposing that the profit motive has adverse consequences when the patient has imperfect information about quality. As we have stated, there are few private practitioners in Africa so the relevant comparison usually is not between profit and not–for–profit but between governmental and non–governmental organizations. We seek to explain why employees who are otherwise identical (in terms of skill, training and for the most part their instinct to seek profit) behave and are perceived to behave in such different manners depending on their employer's identity. It is not enough to say that non–profits maximize something other than profit; we must explain how they manage to extend the proper incentives to their personnel.

We introduce three categories of mechanisms or institutions that could poten-tially resolve these problems and discuss why we believe only NGOs are currently solving the problem on an effective scale. The first type of mechanisms are *au-tonomous outcome–contingent contracts*. Such contracts can be formed between any patient–provider pair if the correct environment exists. The fabled "pay only if cured" contract of Dranove and White (1987, pg 409) is one such mechanism.

Another is contracts with the implicit sanction of malpractice suits against practitioners. Repeated interactions or the expectation of repeated interaction can also fall under this category. Repeated patronage is dependent on the observation of a series of outcomes and an implicit contract can be formed in which future patronage is contingent on past outcomes. The pay–only–if–cured contract, however, depends on the verifiability of outcomes and malpractice suits depend, at the very least, on the existence of functioning courts. If these exogenous conditions do not exist then autonomous contracts cannot be formed. The absence of effective courts rules out malpractice as a mechanism for insuring quality in Africa but we will show that the pay–only–if–cured contract does exist under some circumscribed conditions. Implicit contracts or repeated patronage are common in African veterinary care but we will show why their success is unlikely to extend to human health care.

There is also the potential for *practitioner–coordinated contracts*, requiring some degree of coordination between practitioners. Examples are professional self-regulation and referral networks, both of which take advantage of the fact that other practitioners do not face the same disadvantage as the patient in evaluating effort. Since practitioners suffer when they are unable to sell their services they could benefit from publicly credible commitments to provide high levels of effort. These networks or professional bodies could, in theory, be formed by providers themselves and not rely on outside organizations (such as the government) for their existence. Furthermore they present practitioners with incentives to provide effort that are independent of any contract with patients. These institutions do not generally exist in Africa and we will present empirical evidence that self–regulation, at least, is not likely to become an important institution.

A third type of institution is the *hierarchically–enforced contract*, which relies on pre–existing organizations. In Africa the relevant organizations are the government and NGOs, generally churches and missions. We claim that both types of organization use the mechanism of hierarchical supervision. We find strong evidence that existing mission health facilities are effectively using this mechanism 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 employ it effectively and generally provide low quality care despite great expense in training,

3

facilities, and supplies.

In the following Section of this paper we will introduce a model of health care seeking behavior with asymmetric information. In Section 3 we examine the potential solutions to asymmetric information suggested by the literature. We weigh these solutions against empirical evidence. The majority of this evidence is drawn from 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, 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 another.[1] In Section 4 we examine empirical evidence specifically concerning NGOs and the ramifications of policy towards NGOs. We conclude in Section 5.

2 Asymmetric Information in Health Care

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. In contrast, the empirical literature on health care in Africa demonstrates quite conclusively that the sick are not passive in their use of the health care system but instead are actively making choices between an array of options. 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 (Burkina Faso: Nougtara et al. (1989); Cameroun: Litvack and Bodart (1993); Ghana: Waddington and Enyimayew (1989); Nigeria: Omorodion (1993) and Stock (1983); Swaziland:

---

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

4

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.

Thus we develop a model of health care in which, although the patient is unable to evaluate what is being done for her health, she does have some information about the characteristics of the condition under which she suffers. This information will allow her to make some educated decisions about which type of practitioner is most likely to provide the inputs she needs in health care.

2.1 The Health Production Technology

We begin with an individual who has fallen sick.[2] She knows her illness condition (the collection of symptoms) but not her diagnosis. We assume that there are only two outcomes to health interventions; a good outcome and a bad outcome (full recovery and no recovery, for example). The characteristics of these outcomes depend only on the disease condition and not on any characteristics of the patient or the practitioner she visits. The probability of achieving either outcome is determined by two binomial distributions; the 'true diagnosis' distribution and the 'false diagnosis' distribution. We motivate these distributions as follows: if the patient's condition is correctly diagnosed, and the proper treatment regime is prescribed, understood and followed, the patient will have a set probability of full recovery. If the diagnosis is incorrect the patient faces a different (lower) probability of recovery. These two distributions contain the key source of 'error' in health care that allow us to model health care as a principal–agent problem. In health, often everything is done as it

---

[2]This exposition is presented with greater rigor in the Appendix.

should and the patient does not recover — the probability of recovery for the 'true diagnosis' distribution is not generally one. On the other hand patients frequently recover when nothing has been done for their health (or when incorrect actions have been taken) — the probability of recovery for the 'false diagnosis distribution' is not zero.

Health care is a set of technologies that probabilistically span these two distributions. A 'better' technology is one that has a higher probability of choosing the 'true diagnosis' distribution than another technology. The 'best' technology has a 100% chance of correct diagnosis and leads to the highest possible chance of recovery (not generally one). The 'worst' technology has a 100% chance of choosing among the incorrect diagnoses and leads to the lowest possible chance of recovery (not generally zero).[3]

The properties of the two binomial distributions are given by the illness condition. The patient cannot choose the distribution under which to seek health care, but she does have some control over the magnitude of health technology. The technology is generally a function of practitioner effort, practitioner skill, patient effort and patient skill (the patient's efficiency at transforming health inputs into health). An increase in any of these factors, *ceteris paribus*, increases the probability of choosing the 'true diagnosis' distribution. The role of each of these factors will vary according to the illness condition.

If the probability of cure under 'true diagnosis' is not much greater than the probability of cure under 'false diagnosis' then the patient might decide not to invest in any technology and just take her chances. On the other hand, if the difference is great enough and she suffers from a condition that is easy to diagnose but requires a therapeutic technique only available at one type of provider, she will know that the greatest chance of recovery comes from a visit to that provider. Another condition might indicate a range of diseases each of which is easily treated, but knowing which therapy is correct would require the exertion of diagnostic effort. In this case the patient will seek a provider who has incentives to exert effort.

---

[3]This description meets the Spanning Condition of Grossman and Hart (1983) and the Linear Distribution Function Condition of Hart and Hölmstrom (1987), which allows us to characterize incentive compatibility constraints as first order conditions.

Unobservable efforts imply that the patient does not ever observe the technology used in treatment; she only observes the outcome. Since *both* outcomes are possible with *all* technologies the patient can never know physician effort even if she knows her own level of effort, her own skill and the practitioner skill. In the example of a condition that indicates a range of diseases all practitioners will make a diagnosis from among these diseases but the patient cannot tell which is right and which is wrong. The practitioner who exerted no diagnostic effort has a positive chance of being correct, but the practitioner who did exert effort has a greater chance. Since the exertion of effort causes disutility, patients can expect only effort compatible with incentives. These incentives, in turn, are set by the contract for health care provision from that provider. Patients can affect the technology of health care through their choice of practitioner and their own level of effort.

Patient effort (either the sick individual or her family) is an important element in medical treatment. Since the majority of this effort takes place out of sight of the practitioner it is also unobservable to the other party in the contract. When patients alone gain all the benefits and suffer all the cost of their subsequent health condition they bear all of the risk and this unobservable effort is theoretically uninteresting. But when they share risks with practitioners (as they do in contracts such as the "pay only if cured" one) this effort and incentives for its provision can become important.

Utility depends on the outcome, costs incurred and the disutility of the patient's own effort. Costs can include fixed fees, contingent payments, drug expenses and (importantly) travel costs. All of these factors will therefore be important in the patient's choice of practitioner.

3   Solutions to the Principal–Agent Problem

The classic solution to the principal–agent problem is for the patient to pass on to the agent the residual value of the production process. Thus, after fixed payments, the agent/practitioner should be paid the full marginal value of medical effort to the patient. He would then naturally equate his marginal effort with his marginal costs and the full information solution (the solution that would be achieved in the

absence of imperfect information) would be achieved. In our case this would involve the patient agreeing to pay the practitioner an amount that varied one for one with her marginal utility. So a patient suffering from a headache would offer to pay the practitioner a fee that would be larger if she was cured of her headache and smaller if she was not. If the difference in payments with and without the cure equals the value of the difference in utility the full information solution is achieved.

However, due to unobservable patient effort, we do not have the classic model but rather a model of production in teams following Hölmstrom (1982). This model is more complicated and in general the full–information solution cannot be obtained. If the patient pays the practitioner the full value of the change in her utility from health then she is at exactly the same final level of utility at both outcomes. If she were not, the contract would not have been the best contract. But if she is at the same level of utility then what reason does she have to exert costly effort to change the outcome? This is the core of the problem with moral hazard in teams — both parties cannot face the full incentives at the same time.

Though we cannot achieve the full information solution we can do better with some form of contract than with no contract.[4] One such contract would be for the patient and provider to share the value of the outcome. Rather than paying the practitioner the full value of the change in utility the patient would pay some share of that value. Then both principal and patient face an incentive to provide effort — though not at full information levels.

3.1 Autonomous Outcome–Contingent Contracts

Leonard (1998a) has shown 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. A random sample of households in South-West Province of Cameroun found that patients cured of their ailments made payments to traditional healers twice the size of payments made by patients who were not cured. Payments made after the initial consultation were 3.6 times higher when a patient was cured than when she was not. Payments to 'modern'

---

[4]For a proof of this proposition, see the appendix.

8

practitioners in contrast are higher in unsuccessful cases.

It is rare for a modern medical practitioner to make his fee contingent on the outcome of treatment. On the other hand if fees are billed *after* the service is provided, and practitioners know that their 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).

3.2 Enforceability and Verifiable Outcomes

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 externally verifiable as well as observable. When outcomes are not verifiable contracts are generally not enforceable because the practitioner or the patient might refuse to honor the contract and no outsider can verify which party is violating the contract. In the case of the contingent payment contract introduced above, the patient 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 if the outcome is not observed by any institution that can enforce payment or punish lying. If the practitioner 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. Since contingent–payment contracts reduce agency costs, benefiting both practitioners and patients, the patient 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 an exogenous feature of the institution of traditional medicine. Most Africans believe that healers are the agents of spiritual powers and it is these powers that have cured them. They believe these powers cannot be deceived about the outcome and could revoke a cure, cause

9

a new illness or cause the death of the patient.[5] Patients thus believe outcomes are verifiable and this binds them to tell the truth. Modern medical practitioners could never use such a mechanism to enforce contracts.

A modern medicine variant of the contingent contract was observed for maternities in Cameroun (Ndeso-Atanga 1999). An informal convention has developed in Cameroun that the family pay a substantial 'appreciation' to the 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. 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 NGOs. 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% of deliveries vs. 30% for other conditions).

This has important implications for privatization. In Africa the wide-spread use of corruption often means that many services are *de facto* if not *de jure* privatized. However, in health care, corruption is only a hurdle to access and does not lead to increased quality. If health care did not suffer from asymmetric information side–payments should lead to increased quality. Without verifiability this will not happen.

3.3 Repeated Interaction

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 service.[6] With imperfect information, however, a practitioner

---

[5] For details of this institution see Leonard (1998a).

[6] The socialization literature, particularly for professionals, examines processes that seek to create individuals who exert effort independent of direct incentives but the mechanisms have a high probability of success only when reinforced by other indirect incentives (Freidson 1970, Leonard 1991, Chap. 12). There is no evidence that such indirect mechanisms operate effectively for government and private medical practitioners in Africa.

that is known to be extraordinary (as we shall refer to him) 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 (one with distaste for effort) could benefit by seeking to imitate an extraordinary practitioner. Thus reputation is the changing beliefs of clients such that, correctly or incorrectly, patients believe a given practitioner will always provide effort independent of direct incentives. 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.

If there was very little randomness in outcomes, 'good' and 'ordinary' practitioners would be almost immediately distinguishable. However, when, (more realistically) the difference in the probability of cure with the 'true diagnosis' and 'false diagnosis' is not great, it can take a large number of outcomes to distinguish practitioners. Practitioners who practice in a limited geographical area can gain a reputation when patients observe a long string of mostly positive outcomes from treatments.[7] 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.

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 the patient's belief. 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

---

[7]This follows the model of reputation outlined by Kreps and Wilson (1982) and is in contrast to the model of Satterthwaite (1979) and Pauly and Satterthwaite (1981) in that, for us, the cost of gathering information is zero. Their models examine the cost of gathering information when there are differing numbers of practitioners. In our environment the number of practitioners is always low.

veterinary medicine. Ly (1999b) found that clients purchase more, and a greater variety of services from practitioners in whom they have confidence than from other practitioners.

When most practitioners are not extraordinary or good there is an investment cost to trying to gain a reputation. Here we must take into account the fact that setup costs for private practice in rural Africa are already seen as being too high, especially in the absence of long-term credit markets. Mechanisms that insure quality and thus higher revenues without requiring large investment costs will naturally be more attractive.

3.3.1 Repeated Interaction and Outcome–Contingent Contracts

As we have stated, contingent contracts with verifiable outcomes improve on the solution obtained with only fixed payments. If a contingent–payment contract could be sustained over repeated interactions even without verifiability, it is likely to lead to superior outcomes both in the short and long run for both patient and practitioner. There is a twist with sustaining contingent contracts when the patient agrees to pay the practitioner upon the realization of a successful outcome; it is the reputation of the patient that matters, not that of the practitioner. Under what conditions does the patient 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. ... 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 patient 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 patient 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 ability of the practitioner to identify clients with their history of interactions; the client's probability of needing services in the future; and the additional cost incurred in seeking an alternative practitioner. These factors help to explain why we observe contingent–payment contracts (deferred fees that cannot be collected when outcomes are bad) in use among veterinarians but not among doctors. Veterinarians often enjoy a wider geographical monopoly because transportation costs for animals are significantly higher than for humans (Woods 1999, Leonard 1987). The reputation is attached to a herder but each herder represents a large number of animals which could fall sick, increasing the likelihood of future visits as well as making identification of clients easier for vets. Human medicine practitioners see a larger number of clients per case, and their client can more easily find alternative providers. This helps explain why we do not observe such contracts in health care.

3.4 Budget–Breaking Incentive Systems

Autonomous contracts can be formed without the direct intervention of the state or other outside organizations, but they rely on exogenous circumstances or substantial investment. However, in most types of medical transactions outcomes are not verifiable and repeated interaction cannot offer enough long-term benefit. Are there institutional mechanisms that can assure effort in these, more typical circumstances? We examine in this section institutions that allow enforcement of contracts without verifiability.

The patient lies when payment is required in an unverifiable contingent–payment contract because after the outcome is observed payment can only decrease her utility. Consider instead that there are two transfers, not one; a payment from patients and a receipt by practitioners. The payment received by the practitioner depends on the declaration of the patient, but the payment made by the patient depends on

13

the declaration of the practitioner.[8] This requires a third party who can add or remove money when these two payments do not match. Now the statement of the patient does not change her utility. She gets the value of the outcome and makes a payment that depends only on the practitioner's declaration. She has no reason to lie, and if we assume a mild propensity to tell the truth she will tell the truth.[9] By similar logic the practitioner is induced to tell the truth. Now the outcome will be known and payments can be made appropriately. Thus even with unverifiable outcomes, if (but only if) the budget does not have to be balanced a contract can be implemented that makes payment contingent on outcome. Furthermore since both patient and practitioner can face the full marginal value of health when budgets are broken, the full information solution can be achieved.

Once a budget is broken implementation of outcome-contingent contracts without verifiability is just the beginning of the types of contracts that can be introduced. Breaking budgets introduces the possibility that payments can be made not just on different statements about the outcome of one process, but on statements about the outcome of different signals.

3.4.1 Correlated Signals

With the contract outlined above payments to patients are easy to make. If a patient makes a fixed payment at the beginning of treatment and then returns home, she will automatically gain the value of outcome, since this is just her own health. The third party does not need to find the patient in order to pay the patient, but the patient must be located and polled in order to make the practitioner's payment. However, in some circumstances, it is not necessary that the payment to the practitioner be tied to the outcome of treatment. Any signal will achieve the same solution if it is correlated with the outcome of treatment and is verifiable, such that a contract can

---

[8]In 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).

[9]This mild propensity to tell the truth has bothered other writers (it is a weakly dominant strategy) and other solutions have been proposed, namely Nash implementation. Thus if the declaration of both parties is in agreement then the original contract must be fulfilled according to the declaration but if they disagree then both are penalized. Agreement on the truth is therefore a Nash equilibrium: however, so is agreement on a lie (Moore 1992). The reality of the contracts actually observed makes these distinctions less important.

14

be written that leads to the same incentives for the practitioner. For example, if the practitioner is paid, not on the basis of each outcome, but on the outcome of a randomly selected treatment from among a series of treatments, payments to the practitioner can be designed such that he faces the same incentives. The outcome of the selected treatment is still not verifiable, but since payments to that patient do not depend on her declaration the contract is implementable.

This is an important modification to the class of potential contracts. Despite the fact that with broken budgets patients 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 have generally returned home by this point. The presence of transaction costs means that organizations face a trade-off between correlation with 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. Polling a random patient about the outcome would greatly reduce the cost of gathering information about outcomes, but would not be implementable because this outcome is still un-verifiable. The patients declaration will influence her payment because budgets are balanced and she will therefore have an incentive to misrepresent the outcome.

Though observed institutions use both budget–breaking and 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 below differ in many aspects of their practices but all break budgets and collect correlated signals rather than direct information about outcomes.

3.5 Practitioner–Coordinated Contracts

There are many ways to break a budget. For example, the practitioner could form a side-contract on outcomes with other practitioners that would break the symmetry between payments and receipts. The main contract is a balanced one between patient and practitioner, but if the practitioner violates the side-contract he will loose some large stock of value such as his relationship with other practitioners

or his right to practice. The value of this asset is not collected by the patient or by other practitioners and therefore the side-contract and the main contract are not balanced. Self-regulation and referral networks are examples of these types of side-contracts.

3.5.1 Self–Regulation

Many countries rely upon self–regulation by professional associations to provide quality-assured medical care. These not only control licenses but hospital privileges as well. Since professionals gain from credibly committing to provide quality care, they gain by the creation of such an organization. They can do so because practitioners can asses the quality of work done by their peers better than patients can.

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 predicts that as private medical markets emerge professional associations should 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 1985, pp. 19-29; Arrow 1963, 1985, pp. 38, 40). Ly (1999a) examined this hypothesis among veterinarians 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.

16

France and large parts of Africa use state regulation as a proxy for self–regulation. In theory the extended powers of the government can provide the same regulation while avoiding the pitfalls caused by voluntary agreements. However, the same governmental incompetence and corruption that makes privatization attractive to Africa renders the state unreliable as a regulator.

3.5.2 Referral Networks

Referral networks fulfill much the same function as self–regulation by professional associations but in an informal and less costly manner. They are the mechanism whereby most real professional self-regulation and quality enhancement takes place in the United States (Freidson 1970, Savage 1997). 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. Thus if frequent interaction teaches the patient that their primary care physician provides a quality of care that they regard as satis-factory, they can use that MD's referrals to find other medical providers of similar quality. Haas-Wilson (1990) shows that (for psychotherapy services) referred practi-tioners earn higher fees, a result compatible with the idea that referrals benefit both practitioner and patient. 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.

Referral networks are not generally present in African health systems. There seem to be several reasons why this is so. First patients very often self-refer to higher–level facilities and thus do not allow referral networks to function (Sauerborn et al. 1989). This may be the result, however, not the cause of absent networks. Second, when referrals are made they are made to facilities, not to particular practi-tioners. Given the depersonalized character of these relationships, there is very little communication among medical practitioners about the cases they have shared. Thus little information is generated that might teach personnel about one another's qual-ity. Third, the sharing of knowledge (both about medicine and about one another)

17

takes place most effectively through informal and frequent personal interactions. Large numbers of professional staff are working in isolation, 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.

3.6 Hierarchically–Enforced Contracts

We suggest that breaking a budget is an important role of any organization that employs practitioners. These organizations observe signals correlated with outcomes and then make payments to practitioners that help to align incentives.

3.6.1 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 signals 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 out-

18

comes. Organizations monitor this activity in an attempt to 'internalize' the externality imposed by excessive use, for example, of antibiotics. 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 Tanzania (Leonard 1998a). In only 16% of observations (17% at government and 11% at NGO facilities) did the doctor tell the patient what the diagnosis was or what medicine was being prescribed. In only 6% of cases (7% at government and 5% at NGO) 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% (38% at government and 30% at NGO) did the nurse check to see if the patient had any idea how to do so. Despite the emphasis these organizations put on health education, organizational quality does not create strong incentives to encourage patient effort. Furthermore, although there is strong evidence that missions provide more unobservable effort they do not address patient effort significantly more than government facilities.

We stated earlier that any signal available to one type of provider should, in theory be available to all providers. The signal used by organizational providers is verifiable and therefore seems to circumvent the previous problems that we claimed faced private providers. Is breaking a budget necessary for the implementation of this contract or can private providers use this signal to implement organizational quality–contingent contracts? In this case the patient would visit a provider and her payment to that provider would depend on the degree to which his actions conformed to protocols. She would sometimes pay more for a protocol adhering treatment and not be cured and sometime she might pay less for a protocol violating treatment and be cured, but on average her payment should correspond to the expected outcome.

The problem with this contract is transaction costs. It takes a doctor to evaluate protocols and organizations do not have monitors present at each consultation, they evaluate randomly selected treatments after the fact. A patient cannot take her own private monitor with her when she visits the doctor: if such a monitor existed she would be unlikely to need a doctor at all. A private practitioner could contract with an outside monitor to credibly punish him if he violates protocols. This is no longer a balanced budget (since the punishment would not be gained by patients), but is implementable. This contract with an outside monitor will be discussed in Section 4 when we look at franchise relationships between NGOs and private practitioners.

4   Evidence and Policy Prescriptions for NGOs

A great number of the potential solutions that we have examined rely on institutions like courts or strong professional associations or exogenous characteristics for enforceability or reputation. 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 with hierarchical supervision. Government and NGO 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 NGO facilities, however. This is an area in which Africa has a readily available, highly-functional institution for dealing with moral hazard.

4.1   Empirical Evidence

It is very clear that most NGO 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 cannot 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 patient 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.

4.1.1 Self–Diagnosis and the Choice between Providers

We have noted earlier that Africans are making choices between medical providers based on their own understanding of their illnesses. Here we draw on evidence that their choices follow patterns that would be predicted by the theory of asymmetric information advanced in this paper and therefore offer empirical support for our conclusions

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 patients visiting government clinics and being referred to government hospitals when necessary, or, if the delay was substantial, self–referring to government 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 donor project to insure adequate drug supply 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 since it breaks budgets, but the fact that organizational quality does not take into account the role of the patient means that it will not, in general, achieve the first best solution. We have developed a model of illness conditions as production technologies that differ in the degree to which they require unobservable inputs from medical practitioners and patients. With it we can show that different forms of contracts are appropriate for different diseases.

Mwabu (1986) examines the patterns of diseases reported at different providers using cluster analysis and finds significant patterns across illness conditions. Leonard (1998b) quantifies this analysis using medically assessed values for the benefit of unobservable medical and patient effort and skill than can be used in a production function. By creating explicit representations of the contracts discussed above and estimating some of their parameters he matched predicted patterns of patient behavior to observed patterns. Leonard (1998b) estimates that practitioners in mission facilities face penalties over 50% greater than government practitioners for the same violation of protocol and that this fact helps explain the patient choice to by-pass government facilities when they suffer from conditions that require high levels of unobservable effort.

The results of this and other empirical investigations (Leonard 1998a) suggest that the patterns observed are in part explained by the contracts available at different providers. Patients are aware of the incentive effects of various institutional structures and act accordingly.

4.2   Policies for the Extension of Institutional Capacity

If NGOs play a unique role in the provision of health care, as it would appear that they do, is there any way to systematically take advantage of this fact? At the very

least NGOs should not be unnecessarily deterred from expanding there role, but are there ways of encouraging this expansion? Below we discuss empirical evidence from two possible expansion schemes.

4.2.1 Franchising an NGO Brand Name

The Tanzanian government banned all private practice in 1977 and gave private practitioners the choice of joining the government service or joining the practices of voluntary organizations. In effect they could keep their practice if any religious organization agreed to serve as their sponsor and "practitioners retained ownership of the facility and paid voluntary organizations a commission for using their names." (Kanji, Kilima and Munishi 1992, pg. 9) Even as strictures against private practice were loosened the tax exemption offered to NGOs meant that private practitioners continued to seek these relationships.

The government therefore recognized and implicitly encouraged a franchise relationship between NGOs and private practitioners. This franchising should have lead to the extension of the institutional capacity of NGOs. Bearing the name of an established NGO a private practitioner could forgo the up-front costs of establishing a reputation, while the NGO could provide quality-assured services at a very low cost, using their capacity for regulation to expand the reach of their organization.

However a 1992 study on quality of clinics in Dar-es-Salaam (Kanji et al. 1992) found that these hypothetical hopes were not realized. Out of 70 clinics registered under the name of religious organizations only 15 were found to have any real relationship with an NGO. The others were private practitioners associated with organizations that did not operate any facilities on their own. Religious organizations with any self-run clinics did not sell their name to private practitioners. Thus it would appear that a religious organization would either sell its name or provide its own services, but never both at the same time.

These findings imply that the franchising we observe in Tanzania is only an attempt to get a license or avoid paying taxes but not a method to extend reputations. Furthermore established religious organizations have not shown any desire to extend their services using this method. It is common in Africa for an established NGO hospital to open 'annexes' in geographically remote locations. In Cameroun,

for example, whenever this is done the annex will bear the name of the hospital. Mbingo Baptist hospital and Mbingo annex are at least 150 miles apart and in different provinces; they bear the same name specifically because the parent institution enjoys a wide–spread reputation from which the 'annex' can immediately benefit. However, annexes remain under the hierarchical control of the parent whereas regulation of a franchised private provider would require management expertise outside of the hierarchy. It would appear that NGOs are reluctant to assume this different regulatory role.

4.2.2 Contracting Government Services out to NGO's

A number of African governments contract out health services to NGOs and some private health systems. For example, in Zimbabwe a mining company was paid to open up its company hospital to the entire local population (Mills 1997). Gilson et al. (1997) discuss the experience of Zimbabwe, Tanzania and Ghana in which mission health services were contracted to run designated district hospitals. These contracts are not put up for tender in the classic sense; there is no bidding or competition and there is limited monitoring of the service providers. The limited evidence about these experiments (the study of mission hospitals compared those hospitals to matched government hospitals) suggests that the contractors provide the same or higher level of service as the government at lower cost, though the results are mixed. These differences might be due to greater flexibility in the type of personnel hired, the ability to hire and fire and the discretion to shift money between budget categories. Additionally, "[t]here are clear signs that, as is commonly assumed of NGOs, the church hospitals studied were motivated to provide good quality care, and did not abuse the resources made available to them" (Gilson et al. 1997, pp 295). However, there is little in-depth comparison of quality in terms of outcomes or patient perceptions rather than ours.

Mills (1997) suggests that competition for such contracts might improve outcomes, as lower cost providers would be more likely to win. However, given the limited success of privatization, is the least cost provider likely to deliver quality services? Though we have stated that NGOs generally provide high quality services it would be a mistake to claim that all NGOs always provide high quality

24

services. Whether or not competitive bidding will reduce costs without reducing quality is likely to depend on some of the following factors. If the winner is to provide services in which the bulk of their revenue depends on patient fees they prefer more, not less patients. Of the above mentioned church providers one hospital in Ghana received 40% of its income in fees. However, the other 5 received almost all of their income directly from the government. We would expect the Ghanaian hospital to work harder to attract patients. This suggests an interesting implication; NGO organizations with wide-spread reputations for high quality services will attract more patients and collect more fees than those with no reputation. They therefore, might be able to out-bid competitors without any other cost advantage. In this case competitive bidding could select only higher quality providers.

On the other hand, it is not the case that low quality providers attract no patients. Government clinics and hospitals continue to attract patients and we have seen that private providers in Kenya, for example, thrive on the sale of observable goods and services. A low quality organization would run a hospital differently than a high quality organization and would attract a different mix of custom, but it is not clear that it would earn less revenue. Market discipline does not have the same implications in a market with imperfect information.

In each of the cases of mission subcontracting, the churches supplemented hospital income with their own funds. The winners of the contracts clearly had some personal interest in the provision of health services and were not seeking profit. A for-profit provider would not have agreed to these contracts. It would seem that contracting in this context is a government subsidy for NGO services. The NGO had a desire to provide the services; the government has simply made it less expensive for them to do so. Rather than giving money directly the government has given a subsidy that is contingent on the expansion of service provision. These findings suggest that the relationship between the government and NGO providers of health services is better thought of as a long-term partnership than a low cost way to contract out services.

5 Conclusion

There are many potential solutions to the problems of asymmetric information in African health care. In fact we observe many of these solutions at work in Africa. However, most of the standard solutions do not function, mostly because of the absence of 'macro' institutions like effective government regulation or a functioning court system. Even those solutions that we do observe at work in Africa (such as traditional healers and child-birth appreciation contracts) are often severely constrained by the scope of the institution.

This lack of institutions is a direct factor in the limited success of private health practices on the continent. However, NGOs to provide extensive, popular and high quality services in Africa. It should come as no surprise that these service organizations want to provide these services. But their success should be contrasted to the failure of governments that also want to provide high quality services. This work shows how NGOs transform their organizational objectives into practitioner level incentives in an environment of asymmetric information.

They differ from private practitioners in their ability to break the budget of value created when a patient is treated. Since practitioners can be rewarded or punished in ways other than payments from patients their utility can be made dependent on measures correlated with outcomes. Private practitioners could not implement such contracts because outcomes of health care are not generally verifiable. Thus we find that private practitioners specialize in observable services and goods and only missions are left providing the vast array of services that require unobservable efforts.

The goal of all African countries is to increase the availability of quality–assured health care. It had been widely hoped that allowing private health care practice would help cash-strapped governments achieve this goal. This has not happened. Policy makers should recognize and embrace NGO providers as a unique and pre-existing institution for the delivery of high quality health care.

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.

Arrow, Kenneth J., "Uncertainty and the Welfare Economics of Medical Care," *American Economic Review*, 1963, *53* (5), 941–973.

Bennett, Sara, Barbara McPake, and Anne Mills, eds, *Private Health Providers in Developing Countries: Serving the Public Interest?*, London and New Jersey: Zed Books, 1997.

_____ **et al.**, "The public/private mix debate in health care." In Bennett, McPake and Mills, eds (1997) chapter 1.

Berman, P. et al., "Kenya: Non-governmental Health Care Provision," Technical Report, Department of Population and International Health, Harvard School of Public Health April 1995.

_____ **and** _____ , "Zambia: Non-governmental Health Care Provision," Technical Report, Department of Population and International Health, Harvard School of Public Health January 1995.

Bichmann, W. et al., "District Health Systems: Users' Preferences for Services in Benin," *Health Policy and Planning*, 1991, *6*, 361–70.

Clarke, E., "Multipart pricing of public goods," *Public Choice*, 1971, *8*, 19–33.

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.

Ellis, Randall P. and Thomas G. McGuire, "Optimal Payment Systems for Health Services," *Journal of Health Economics*, 1990, *9*, 375–396.

Freidson, E., *Profession of Medicine: A Study of the Sociology of Applied Knowledge*, New York: Harper and Row, 1970.

Gilson, Lucy et al., "Should African Governments contract out clinical health services to church providers?" In Bennett et al., eds (1997) chapter 17.

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.

Hanson, Kara and Peter Berman, "Private health care provision in developing countries: a preliminary analysis of levels and compostion," *Health Policy and Planning*, 1998, *13*, 195–211.

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.', Cambidge: Cambridge University Press, 1987.

Hölmstrom, Bengt, "Moral Hazard in Teams," *Bell Journal of Economics*, 1982, *13*, 324–40.

Kanji, N., P.M. Kilima, and P.M. Munishi, "Quality of Primary Curative Care in Dar-Es-Salaam," 1992. Unpublished paper.

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.

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.

_____ , "Incentives and Rural Health Care Delivery." 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 Pastorilists*, Wesport, Conn.: Bergin and Garvey, 1996.

McPake, Barbara, "The role of the private sector in health service provision." In Bennett et al., eds (1997) chapter 2.

Mills, Anne, "Contractual Relationships Between Government and the Commerical Private Sector in Developing Coutries." In Bennett et al., eds (1997) chapter 13.

Mliga, G.M., "Decentralization and the Quality of Health Care." In Leonard, ed (1999) chapter 8.

Moore, John, "Implementation under Compelete Information," in Jean-Jacques Laffont, ed., *Advances in Economic Theory*, Oxford: Oxford University Press, 1992.

Mwabu, Germano 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.

Nougtara, 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.

Parsons, Talcott, *The Social System*, Glencoe, Ill: The Free Press of Glencoe, 1951.

Pauly, Mark V. 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.

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.

Stock, R., "Distance and the Utilization of Facilities in Rural Nigeria," *Social Science and Medicine*, 1983, *17*, 563–70.

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.

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 Mathematical Appendix

A.1 The Value of Health Care

The level of health is H. Health intervention leads to a change in the level of health, ΔH. There are only two possible outcomes; the worst outcome $\Delta H = \underline{h}$ and the best outcome $\Delta H = \bar{h}$. These outcomes depend only on the disease condition and not on any characteristics of the patient or the practitioner he visits. The probability of achieving either outcome is determined by two binomial distributions. $\phi^\star$ is the 'true diagnosis' distribution and $\phi^\emptyset$ is the 'false diagnosis' distribution. With the 'true diagnosis' distribution the probability of recovery is $p^\star$. If the diagnosis is incorrect the probability of recovery is $p^\emptyset$. The probability of failing to recover is $(1 - p^\star)$ with the 'true diagnosis' and $(1 - p^\emptyset)$ with the 'false diagnosis.'

Health care is a set of technologies that probabilistically span $\phi^\star$ and $\phi^\emptyset$. A 'better' technology is one that has a higher probability of choosing the 'correct diagnosis' distribution than another technology. We represent the technology by e $(0 \leq e \leq 1)$ where

$$\Delta H \sim e \cdot \phi^\star + (1 - e) \cdot \phi^\emptyset \tag{1}$$

We model utility following Grossman (1975) and consider health as increasing the hours of time available to consume work and leisure as well as augmenting utility directly. Thus $U = (H, I(H), d(p))$, where $d(p)$ is the disutility of patient effort and $I(H)$ is the income potential at health level H.

The expected value of health is

$$EU = ep^\star \cdot U[\bar{h}, (I(\bar{h}) - C), d(p)] + e(1 - p^\star) \cdot U[\underline{h}, (I(\underline{h}) - C), d(p)] +$$
$$(1 - e)p^\emptyset \cdot U[\bar{h}, (I(\bar{h}) - C), d(p)] + (1 - e)(1 - p^\emptyset) \cdot U[\underline{h}, (I(\underline{h}) - C), d(p)] \tag{2}$$

where C is the total cost of a visit. Of interest to the patient is the change in expected utility. We choose as a natural comparison the utility when no health care is sought. We assume a separable utility form such that $U = U'[H, I(H)] - C - d(p)$. Although income and total costs are measured in the same units and need not be separated we choose this formulation for the following reasons. The income (or earning potential of the patient) and health level for good outcomes is the same whether the patient sought health care or not; it depends on the outcome, not the process. Thus the part of utility inside the utility operator $(U'[H, I(H)])$ does not depend on the practitioner chosen or effort exerted. Costs and disutility have a linear relation to utility. This is reasonable because, although utility may exhibit decreasing marginal utility of income, the change in utility caused by exerting effort or paying medical costs is small enough so as to be approximated in a linear fashion. Patients are not risk neutral over health outcomes, but for tractability we claim that the marginal utility of income is constant over effort and costs. For ease of exposition we write $U'[\bar{h}, I(\bar{h})]$ as $\bar{U}'$ and $U'[\underline{h}, I(\underline{h})]$ as $\underline{U}'$.

The expected utility when no medical care is sought $(e = 0)$ can be expressed as

$$EU_0 = p^\emptyset \cdot \bar{U}' + (1 - p^\emptyset) \cdot \underline{U}' \tag{3}$$

Using the separable utility function the expected utility of seeking care at provider

j is

$$EU_j = \left(e(p^\star - p^\emptyset) + p^\emptyset\right)\bar{U}' + \left(1 - p^\emptyset + e(p^\emptyset - p^\star)\right)\underline{U}' - C - d(p) \qquad (2')$$

The change in the expected utility is

$$\Delta EU_j = e(p^\star - p^\emptyset) \cdot \left(\bar{U}' - \underline{U}'\right) - C - d(p) \qquad (4)$$

At this point we make a number of further simplifying assumptions. First, we assume that $\underline{U}'$ is equal to zero, a simple scaling assumption. Furthermore we assume that utility from health comes from a fixed health affect, $\bar{h} \cdot \bar{\omega}$ (where $\bar{\omega}$ is the per unit value of health) and an increased amount of time for leisure or work, $\bar{h} \cdot w$ (where w is the opportunity cost of healthy time.) We cannot separate these two effects and therefore use the combination of effects, $\bar{h} \cdot \omega$ (where $\omega = \bar{\omega} + w$.) Thus

$$\Delta EU_j = e(p^\star - p^\emptyset)\omega\bar{h} - C - d(p)$$

Without loss of generality we define the technology for health production as being a standard production function divided by a 'maximum' level of production for that function, $e = h/\bar{h}$. Thus where e varies between 0 and 1, h varies between 0 and $\bar{h}$.

$$\Delta EU_j = (p^\star - p^\emptyset)\omega h - C - d(p)$$

For simplicity of exposition we replace $(p^\star - p^\emptyset)\omega h$ with h'.

A.2 The Health Production Technology

We view the health production technology (e) as a search for the proper treatment regime. This search however is a complex function of a number of different inputs; a production function of health. We assume the following factors are important in the production of health; medical effort, patient effort, medical skill and patient efficiency at transforming health inputs into health. An increase in any of these factors, *ceteris paribus* increases the probability of choosing the 'true diagnosis' distribution. The role of each of these factors will vary according to the illness condition. Generally $h = f(m, p, \psi_k, \psi_j, \psi_i)$ where m is the effort of the practitioner, p the effort of the patient, ψ_k characteristics of the illness condition ψ_k characteristics of the practitioner, and ψ_i characteristics of the patient. k is the illness index, j the practitioner index and i the individual index;

Payments to health care practitioners differ across providers but can involve fixed fees (paid before a consultation), an outcome–contingent fee (paid after the outcome is observed) and drug costs. Traditional healers charge a fixed fee and negotiate with the patient over a final payment to be made if the patient is cured. We assume this payment depends on the value of the outcome and can therefore be expressed as a share of the value created when the patient is cured, $r\omega\bar{h}$. Traditional healers do not charge for drugs. Organizational providers charge a fixed fee for consultation and charge for all drugs administered.

Expected utility for patient i, treated at provider j for illness condition k is

$$\Delta EU_{ijk} = (1 - r)h' - C_f - C_d - C_t - d(p) \qquad (5)$$

31

Costs are composed of fixed fees (C_f), drug costs (C_d) and travel costs (C_t). The travel cost is a function of the individual (origin) and the practitioner visited (destination), the drug cost is a function of the practitioner and the disease, and the fixed cost is a function only of the practitioner. $r = 0$ for all providers except the traditional healer. The income of the practitioner is the sum of all payments by the patient, not including travel costs.

In the specification we have introduced, the patient and the traditional healer both have incentives to exert effort, so although effort cannot be purchased in a perfect market, a contract exists for its provision. The provision of effort on the part of government and mission health centers is not quite as obvious.

$$g(m, \psi_k, j) = p \cdot k \left(Q_k^* - Q(m, \psi_k) \right) \tag{6}$$

where ψ_k are characteristics of the disease, p is the probability, for any given organization, of a record being examined and k is the baseline penalty for organization. p and k cannot be identified separately thus we refer to the product as K.

We can construct the utility of the practitioner (who is risk neutral) and see that both traditional healers and providers at government and mission clinics have incentives to exert effort.

$$E(Y^m) = \underbrace{rh'(m, p, \psi)}_{\text{contingent fee}} + \underbrace{C_d}_{\text{drugs}} + \underbrace{C_f}_{\text{fixed fee}} - \underbrace{g(m, \psi_k, j)}_{\text{penalty}} - \underbrace{c(m)}_{\text{disutility}} \tag{7}$$

Only the healer has a non-zero share and drug costs are received only by organizational providers. Note that no providers receive the travel costs.

A.3 Production in Teams with Unobservable Effort

The interaction between patients and practitioners is based on the principle-agent model of production in teams advanced by Hölmstrom (1982). Increases in health stock are produced by the joint effort of two agents; the patient and the practitioner. These two players form the team. We introduce a third player, the super-principal, who declares incentive–compatible efforts for both agents and pays them according to a contract. The super-principal does not exert effort but has the task of interpreting the information that she receives and executing the contract according to this information. The super-principal maximizes the expected utility of the patient subject to total cost and the participation constraint of the practitioner. Note that all actions must be incentive compatible, even the actions of the patient. Thus the problem can be represented as

$$\max_{r,m,p} EU = (1-r)h' - C_f - C_d - C_t - d(p) - T \quad \text{(S. P. utility)} \tag{8}$$

subject to

$$m^* \in \operatorname*{argmax}_{m} rh' + C_f + C_d - g(m) - c(m) + T \quad \text{(I.C. medical)} \tag{9}$$

$$p^* \in \operatorname*{argmax}_{p} (1-r)h' - C_f - C_d - C_t - d(p) \quad \text{(I.C. patient)} \tag{10}$$

$$rh' + C_f + C_d - c(m) - g(m) + T = V_m \quad \text{(I.R. medical)} \tag{11}$$

$$(1-r)h' - C_f - C_d - C_t - d(p) > V_p \quad \text{(I.R. patient)} \tag{12}$$

where T is the transfer from the super-principal to the practitioner, V_p is the reservation utility of the patient and V_m is the reservation utility of the practitioner. The practitioner's individual rationality constraint (I.R.) is binding and we can therefore substitute it into the super-principal's objective function. We have justified the use of the relaxed incentive compatibility (I.C.) constraints by the choice of distribution for h and we get:

$$\max_{r,m,p} EU = h' - C_t - d(p) - c(m) - V_m \tag{8'}$$

$$\text{subject to} \qquad rw(\frac{\partial h}{\partial m}) - \frac{\partial c(m)}{\partial m} - \frac{\partial g(m)}{\partial (m)} = 0 \tag{9'}$$

$$(1-r)\omega(\frac{\partial h}{\partial p}) - \frac{\partial d(p)}{\partial p} = 0 \tag{10'}$$

$$(1-r)h' - C_f - C_d - C_t - d(p) > V_p \tag{12'}$$

A.4 Balanced budgets

We begin by considering the case of the interaction between the patient and a traditional healer; $K = 0$. The full information solution (where 9' and 10' do not bind) is the point where the marginal benefit of effort is equal to its marginal cost;

$$\omega(\frac{\partial h}{\partial m}) - \frac{\partial c(m)}{\partial m} = 0 \tag{13}$$

$$\omega(\frac{\partial h}{\partial p}) - \frac{\partial d(p)}{\partial p} = 0 \tag{14}$$

But the super-principal can only require effort compatible with the incentives of the practitioner and the patient as defined by the incentive compatibility constraints equations 9' and 10'. 9' and 10' are equal to 13 and 14 if $r = (1 - r) = 1$, clearly never true. When the payments to the patient and the practitioner must always sum to the value of the output (a 'balanced budget') then the share payment cannot achieve the first best solution.

Proposition 1 *A contingent contract with balanced budgets cannot achieve the full information solution in the case of both unobservable patient and practitioner effort. Proof is above.*

Given that the full information solution is not obtainable, is the introduction of a sharing rule an improvement over the solution without a sharing rule?

Proposition 2 *With unobservable efforts as above and balanced budgets any sharing rule $0 < r < 1$ gives a solution strictly superior to either $r = 0$ or $r = 1$.*

For proof see bellow.

We cannot achieve the first-best solution with a sharing rule but we can improve on the solution that would obtain without a sharing rule.

If the super-principal is an actual entity she can inject or remove value from the game; the payments do not need to sum to the total value. By breaking the budget balance the super-principal can achieve the first best solution to the game. Instead of a sharing rule $\{r_m, r_p\}$, where $r_p = 1 - r_m$, choose $r_p = 1$ and $r_m = 1$. Examining again equations 9' and 10' it should be clear that such a rule will achieve the first best solution. Note that the super-principal can then extract the expected value of these payments from the two parties without changing either incentive compatibility constraint. This is not a balanced budget because, though the sum of payments might be equal to the total value in expectation and on equilibrium paths, they do not equal the total value off the equilibrium path.

A.5 Proof of Proposition 2

The Lagrangian is formed as follows;

$$
L = h' - d(p) - c(m) + \lambda_1 \left(r \frac{\partial h'}{\partial m} - \frac{\partial c}{\partial m} \right) + \lambda_2 \left((1-r) \frac{\partial h'}{\partial p} - \frac{\partial d}{\partial p} \right)
$$

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

$$
\frac{\partial L}{\partial r} = \lambda_1 \frac{\partial h'}{\partial m} - \lambda_2 \frac{\partial h'}{\partial p}
$$

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

$$
\lambda_1 = -\frac{\frac{\partial h'}{\partial m} - \frac{\partial c}{\partial m} + \lambda_2 \frac{\partial^2 h'}{\partial m \partial p} - \lambda_2 \left(\frac{\partial^2 h'}{\partial m \partial p} \right) r}{r \frac{\partial^2 h'}{\partial m^2} - \frac{\partial^2 c}{\partial m^2}}
$$

which then leads to an expression of λ_2.

$$
\lambda_2 = -\frac{\frac{\partial h'}{\partial p} - \frac{\partial d}{\partial p} - \left(\frac{\partial h'}{\partial m} - \frac{\partial c}{\partial m} \right) \frac{r \left(\frac{\partial^2 h'}{\partial m \partial p} \right)}{r \frac{\partial^2 h'}{\partial m^2} - \frac{\partial^2 c}{\partial m^2}}}{\left(\frac{(1-r) \left(\frac{\partial^2 h'}{\partial m \partial p} \right)^2}{r \frac{\partial^2 h'}{\partial m^2} - \frac{\partial^2 c}{\partial m^2}} + \left(\frac{\partial^2 h'}{\partial p^2} \right) (1-r) - \frac{\partial^2 d}{\partial p^2} \right)}
$$

Note that

$$
\frac{\partial h'}{\partial p} - \frac{\partial d}{\partial p} = r \frac{\partial h'}{\partial p}
$$

and it should be clear that $lim_{r \to 0} \lambda_2 = 0$. 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.

$$lim_{r \to 0} \lambda_1 = (1 - r) \frac{\frac{\partial h'}{\partial m}}{\frac{\partial^2 c}{\partial m^2}}$$

Substituting both multipliers into the envelope condition we find

$$lim_{r \to 0} \frac{\partial L}{\partial r} = \frac{\left(\frac{\partial h'}{\partial m}\right)^2}{\frac{\partial^2 c}{\partial m^2}}$$

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

1998-1999 Discussion Paper Series

Department of Economics
Columbia University
1022 International Affairs Bldg.
420 West 118th Street
New York, N.Y., 10027

The following papers are published in the 1998-99 Columbia University Discussion Paper series which runs from early November to October 31 of the following year (Academic Year).

The following is the Columbia University, Economics Department's website where viewers may access more information about the Discussion Paper series.

Http://www.columbia.edu/cu/economics/

All **1997-98, 1998-99** and **future papers** will be acessible on-line at the following website:

Http://www.ssrn.com

Discussions papers that are not available on-line maybe obtained by written request to the Economics Department. Please check the Ordering Papers section for details.

Copy Requests

To order 1996-97 and prior discussion papers, please write to the Discussion Paper Coordinator at the address below, along with a check for the appropriate amount, made payable to: Department of Economics, Columbia University. Please be sure to indicate the discussion paper number or the particular series in your written request. *Orders cannot be processed without payment*, and they cannot be taken over the phone, or by fax or email.

Past discussion papers (1996-97 and prior papers) are available for purchase in
   U.S. dollars only, at the cost of:

  US: per paper $4.00 / per series: $140.00
  Canada: per paper $5.00 / per series: $150.00
  Overseas: per paper $7.00 / per series $185.00

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.
9899-03	The Economics of Repeated Extortion	Choi, J.P. Thum, M.
9899-04	To Pay or Not to Pay: Managerial Decision making and Wage Withholding in Russia	Desai, P. Idson, T.
9899-05	Wage Arrears, Poverty, and Family Survival Strategies in Russia	Desai, P. Idson, T.
9899-06	Averaged Periodogram Spectral Estimation with Long Memory Conditional Heteroscedasticity	Henry, M.
9899-07	Household Production, the Bundling of Services and Degradation, and Non-monotonic Environmental Engel Curves	Chaudhuri, S. Pfaff, A.
9899-08	A Carbon Sequestration Supply Function And Development of Feasible Clean Development Mechanism Rules for Tropical Forest Carbon Sinks	Kerr, S. Pfaff, A.
9899-09	Asymmetric Information and the Rofle of NGOs in African Health Care	Leonard, K. Leonard, D.