When both states and markets fail: asymmetric information and the role of NGOs in African health care

Kenneth L. Leonard

Department of Economics, Columbia University, Mail Code 3308, 420 W 116th Street, New York, NY 10027, USA

Accepted 19 February 2002

Abstract

In African health care the “miracle of the market” has not occurred. Patients exhibit willingness to pay for quality health care and yet good private practitioners are unable to sell their services. Simultaneously nongovernmental organizations (NGOs) are running successful health facilities for which patients are willing to pay. We use 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 validate this view empirically and show that NGOs have the institutional capacity to deliver high quality health care. We show that at the same time, neither private practitioners nor regular government organizations charging fees will easily succeed. © 2002 Elsevier Science Inc. All rights reserved.

Keywords: NGO; African health care; Market failure

Since their independence, African states have provided the majority of health care services, often at low cost. Now, 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, including health care, are being increasingly privatized. The “miracle of the market,” however, has not followed for health. Despite the fact that government health care services were generally seen to be of low quality, private health care has not emerged to offer significant competition. At the same time nongovernmental organization (NGO) health services have always provided, and continue to provide, health services in Africa. Since NGOs charge significant fees and their success is widespread, the failure of private markets cannot be attributed to a lack of demand.
It is widely acknowledged that markets for health care are troubled with asymmetric information. Patients cannot assess the appropriateness and quality of the activities of their practitioners. Therefore, it is no surprise that free markets and spot contracts do not lead to an efficiently functioning market for health care. We propose that asymmetric information combined with the institutional environment in Africa explain both the success of NGOs and the failure of state-sponsored and private health care providers. We use the principal–agent framework and examine at the possible solutions to asymmetric information in this framework. Importantly, we not only look at how government, NGOs and private practitioners solve these problems, but also at a broad range of possible solutions. We can learn, not only from the successful solutions, but from the potential solutions that are not observed, since these “dog[s] that did not bark” suggest important institutional realities in Africa.

This principal–agent view, and the subsequent explanation of the success of NGOs is supported by empirical evidence. The core of this evidence is drawn from a seven-country collaborative project (Leonard, 2000) that examined the changing markets for both human and animal health care in Africa, with independent research projects in each country. Although we concentrate on medical care for humans, we also use research done on the provision of services in veterinary medicine, where the principal–agent setup and institutional environment are nearly identical.

1. Introduction

One of the key assumptions underlying structural adjustment efforts in Sub-Saharan Africa is that the implementation capacity of the continent’s states is limited and has been overextended. The World Bank recently has qualified its negative views on the state, but it is still pressing privatization for most sectors of African economies (World Bank, 1997).

The larger number of African states are confronted with most of the following: corruption; patronage (Joseph, 1987); inflated public payrolls; severe fiscal scarcity and a consequent ‘budgeting by cash-flow’ (Caiden & Wildavsky, 1974); constant shortages of critical inputs (Moris, 1977, pp. 78–83); weak performance rewards for public servants, due both to erosion of their value by inflation and the failure to use them to support organizational objectives (Montgomery, 1987); and political authorities who are insecure, indecisive, arbitrary, and interventionist (Moris, 1977; Jackson & Rosberg, 1982). Though their roots are debated, there is little disagreement that these problems are pervasive and that they imply a reduced role for the state in Africa. These problems create a set of incentives for civil servants that are variously perverse, negative, or weak, thereby creating lower levels of public sector performance in Africa than in many other parts of the world. The health care sector is not immune to these influences and the quality of government services reflects this.

Despite this underprovision of health care, throughout most of Sub-Saharan Africa there is little private (for profit) provision of health care (Bennett et al., 1997; Hanson & Berman, 1998). Most countries have no private sector of which to speak. Some countries have done better; in Kenya 23 percent of health facilities are private and in Ghana this figure reaches 40 percent. However, even these facilities are not serving their expected function in health care. It appears that private providers most often are selling drugs or clinical tools, but not
diagnostic services; they are acting as pharmacists not doctors (McPake, 1997, p. 26). This is confirmed by the disparity between hospital beds and percentage of total expenditure. In Zambia, for example, 0.2 percent of hospital beds but 22 percent of total health expenditure is in the private sector (Berman et al., 1995 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, most of which provide poor quality services, 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 percent of clinical care needs, 27 percent of hospital beds and 35 percent 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 percent of all hospital beds and 96 percent of NGO facilities are located in disadvantaged rural areas (Gilson et al., 1997). These providers are not serving niches.

Arrow (1963, p. 965) was the first among many authors to argue 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, fearing that the profit motive might have adverse consequences. 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 nongovernmental 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 in an environment where other organizations fail.

We introduce three categories of mechanisms or institutions that could potentially resolve problems of asymmetric information and discuss why, in Africa, only NGOs are currently solving the problems on an effective scale. We comprehensively examine the various mechanisms suggested by the literature because only by showing that institutions other than NGOs rarely solve these problems can we demonstrate the dominant optimality of the NGO. The first type of mechanisms are what we label autonomous 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, p. 409) is one such mechanism. Another solution is a contract with the implicit sanction of a malpractice suit against a practitioner. When repeated patronage is dependent on the observation of a series of outcomes, a contract can be formed in which future patronage is contingent on past outcomes. Therefore, repeated interactions also can be a form of outcome-contingent contract. The pay-only-if-cured contract, as we will see, 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 scarcity of effective courts and their inaccessibility to most citizens rule 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 over 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 cooperation 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 the higher quality services their patients wish to purchase, they could benefit from publicly credible commitments to provide high levels of effort. Practitioner 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 present empirical evidence that self-regulation, at least, is not likely to become an important institution outside of southern Africa.

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 mechanism, fail to employ it effectively and generally provide low quality care despite great expense in training, facilities, and supplies.

We examine these potential solutions more carefully in the following section. In Section 3 we examine empirical evidence specifically concerning NGOs. In Section 4 we discuss the policy implications of our theory. We conclude in Section 5.

2. Solutions to the principal–agent problem

The relationship between the patient and the practitioner is characterized by asymmetric information, and is similar to the classic principal–agent framework. The core of the asymmetry is the patient’s inability to evaluate the effort provided by the practitioner. It is possible for the patient to remain sick despite the best efforts of the practitioner and it is possible for patients to be cured despite entirely ineffective care. Although health care is useful to the patient, outcomes do not tell patients what was done, and she cannot evaluate effort directly. However, the problem is more complicated because patients also provide inputs into their own health and these cannot always be evaluated by the practitioner (mainly because patients provide these inputs outside the view of the doctor). Thus, patients and practitioners form a team in which neither participant can know for sure what the other is doing, yet the activities of both are important.

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’s 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.

Due to unobservable patient effort, however, 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 in general, we can do better with some form of contract than with no contract. 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.

2.1. Autonomous outcome-contingent contracts

Leonard (2000a) 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 of traditional healers who were cured of their ailments made much larger payments than those 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’ 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 find this form of contract among modern private veterinary practice in Africa, 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.¹

2.1.1. 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 could verify which party was 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, if the outcome is not observed by any institution that can enforce payment or punish lying, she no longer has an incentive to make that payment. If the practitioner anticipates that a contract is unenforceable, he will not accept it. 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 who have played a part in curing them. They believe these powers cannot be deceived about the outcome and could revoke a cure, cause a new illness or cause the death of the patient if the contract were dishonored (Leonard, 2000a). 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, 2000). 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 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 percent of deliveries vs. 30 percent for other conditions).

The importance of verifiability has strong implications for privatization. In Africa, widespread 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.

2.1.2. Repeated interaction and outcome-contingent contracts

Repeated interaction should help to resolve principal–agent problems as it allows observation of a larger number of outcomes from which information can be gathered about the effort that is provided. Practitioners should be able to earn a reputation that they provide high levels of effort. However, by relying only on reputation, practitioners will be required to forgo the benefits of their increased effort until they gain a good reputation. This is an investment cost. If reputations are won slowly the cost can be substantial.

On the other hand, if a contingent-payment contract can 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 to 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?

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 therefore can refuse to do business (or refuse to enter into certain types of contracts) with her in the future (MacLeod & Malcomson, 1982). 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 and link them to 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 physicians. Veterinarians often enjoy a wider geographical monopoly because transportation costs for animals are significantly higher than for humans (Leonard, 1987; Woods, 2000). 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 smaller number of cases per client, and their clients can more easily find alternative providers. This helps explain why we do not observe such contracts in health care.

2.2. Budget-breaking incentive systems

The remaining classes of mechanisms for solving the principal–agent problem involve ‘budget-breaking’ institutions. 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. In most types of medical transactions, however, 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 misrepresents outcomes 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 the declaration of the practitioner. This requires a third party who can add or remove money when these two payments do not match (breaking the budget). 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 do so. 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 and only if the budget does not have to be balanced (and thereby is ‘broken’) a contract can be implemented that makes payment contingent on outcome. Furthermore, since both patient
and practitioner face full incentives (each is paid the marginal value of health) when budgets are broken the full information solution can be achieved (Hölmstrom, 1982).

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 different signals.

2.2.1. 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 the outcome, since this is just her own health. The third party does not need to find the patient to pay her, but she 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 observable, 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 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. Transaction costs mean that organizations face a trade-off between the strength of correlation with outcomes and cost of collection.

Budget-breaking institutions use signal technologies to reduce transaction costs. Technically this technology is also available to an individual practitioner or an organization that does not break budgets, but it will not achieve the same effect. For a private practitioner, polling a random patient about the outcome would greatly reduce the cost of gathering information about outcomes, but would not be implementable because this randomly chosen outcome is still unverifiable. The patient’s declaration will influence her payment because budgets are balanced and she will therefore have an incentive to misrepresent the outcome. 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. We suggest, however, that the comparative advantage of organizational provision comes in their ability to break budgets.

2.3. 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. In such a case 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 and therefore the contract is not balanced. Self-regulation and referral networks are examples of these types of side-contracts.

2.3.1. 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 other practitioners can assess the quality of work done by their peers better than patients can.

Due in part to historically 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 (Arrow, 1963, 1985, pp. 38, 40; Pratt & Zeckhauser, 1985, pp. 19–29). Ly (2000a) examined this hypothesis among veterinarians in Senegal. 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.

2.3.2. Referral networks

Referral networks fulfill much the same function as self-regulation by professional associations but in an informal and less costly manner.4 Because medical personnel are better trained to understand the nature of one another’s work and because they see many more 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. Ly (2000b) observes such a network between a DVM and para-vets created by an NGO 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 practitioners. 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 quality. Third, 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 Africa 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.

2.4. Hierarchically-enforced contracts

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.

2.4.1. 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 for 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.

Organizational quality is a measure specifically designed to be highly correlated with outcomes. It does, however, include some additional information. 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. 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 NGO and government health facilities in Tanzania (Leonard, 2000a). In only 16 percent of observations (17 percent at government and 11 percent at NGO facilities), did the doctor tell the patient what the diagnosis was or what medicine was being prescribed. In only 6 percent of cases (7 percent at government and 5 percent 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 percent (38 percent at government
and 30 percent 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 than government facilities, they do not address patient effort more effectively.

Any signal available to one type of provider should, in theory, be available to all providers. The signal used by organizational providers is observable 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? For example, a patient would visit a provider and her payment to that provider would depend on the degree to which his actions conformed to protocol. She would sometimes pay more for a protocol-adhering treatment and not be cured and sometimes 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 such a contract is transaction costs. It takes a doctor to evaluate protocols and as 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 is discussed in Section 4.2.1 when we look at franchise relationships between NGOs and private practitioners.

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 NGOs 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.

3. Empirical evidence

It is 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 NGO capacity, for staff qualifications and equipment in NGO facilities typically are equal to or worse than those in their government counterparts (Leonard, 2000b; Mliga, 2000). We propose that their success is due to their ability to solve problems of asymmetric information. Previous sections have
presented our theoretical support for this proposition; the following sources offer empirical support.

3.1. Organization of NGO health services

Mliga (2000) observed consultations at government and a variety of NGO clinics and hospitals in Tanzania. NGOs were significantly different from their government counterparts in the completeness of examination, appropriate use of laboratory tests, attentiveness to the client, health education, and availability of medicine. Government health services were superior to NGO services only in the appropriateness of the combination of medicines prescribed; government practitioners generally do not over-prescribe medicines.

Mliga (2000) examines the organizational structure of three different NGO health systems as well as the government system. In Tanzania, government services are highly centralized. A chief of post cannot hire or fire, set salaries, retain any fees collected, set the composition of his staff, set fees or make any decisions about types or quantities of drugs to be purchased. Different NGOs display varying degrees of decentralization and at the opposite extreme, a chief of post at a Roman Catholic dispensary can choose the composition of, hire and fire, and set the salaries of his staff. He must collect fees to pay a significant portion of his costs, he can set fee levels, and he makes all drug purchasing decisions. At this extreme, the only difference between this latter dispensary and a private clinic is that the chief of post is accountable to (and his records are monitored by) the Diocesan Medical Officer, who reports to the Bishop.

The variety of organizational form within NGOs permits an investigation into which features are most correlated with quality. Mliga (2000) finds that decentralized control over staffing decisions increases the diagnostic effort, appropriate use of labs, effort in health education, and attentiveness while it decreases the appropriateness of prescriptions. However, decentralized control of resources (drug purchasing and level of fees) does not lead to an increase of any of these quality measures. This is strong evidence in support of the argument that the ability to effectively reward and punish practitioners increases quality; that incentives are essential.

3.2. Self-diagnosis and the choice between providers

Many authors have suggested that NGOs are superior in observable aspects of health care, such as better drug supplies and superior ‘clinical manner’ (Bichmann et al., 1991; Litvack & Bodart, 1993; Omorodion, 1993; Stock, 1983; Waddington & Enyimayew, 1989). According to the theory, we have advanced these characteristics, by themselves, cannot explain the popularity of NGOs because they could be adopted by private practice. To further demonstrate the validity of our approach, we present evidence that patients visit NGOs when they suffer from conditions that they know require unobservable medical effort.

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 or hospitals. Visits to NGO clinics or hospitals would occur only when the difference in travel costs was such
that it was less expensive to visit NGO 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, such patterns were not observed (Leonard, 2000b; Ndeso-Atanga, 2000). Though drugs were available at all clinics and hospitals only 53 percent of visits to clinics and hospitals were to the closest clinic or hospital. Of visits to hospitals 39 percent involved patients traveling past a government hospital to visit an NGO hospital.

If some elements of effort are unobservable then different patterns of visits can emerge. We stated above that the contingent-payment contract of the traditional healer fails to achieve the full information solution. 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 either.

Illnesses respond differently to different mixes of inputs (where inputs can be patient effort, medical effort, and medical skill among other things). When patients choose between different providers they are choosing between different types of contracts for the provision of unobservable effort. Some contracts are superior when the outcome is responsive to medical effort, others are superior when the outcome is responsive to patient effort, and still others are superior when outcomes are responsive to both medical and patient effort. Different forms of contracts, and therefore different providers, are appropriate for different illness conditions. Our theory can therefore be tested by observing the patterns of choice by illness condition.

Mwabu (1986) examines the patterns reported at different providers using cluster analysis and finds significant patterns across illness conditions. These patterns can be explained by our theory of incentives. In order to test the theory, Leonard (2000b) uses medical assessments of the elasticity of output with respect to unobservable medical and patient effort as well as practitioner skill. Using multinomial logit regression analysis, he shows that patterns of patient choice conform to the theory suggested by our model. NGO providers are visited over government providers when outcomes are elastic to medical effort. The characteristics of illness conditions are important in predicting the choice of provider overall; without illness condition characteristics (using only travel costs and individual characteristics) only 34 percent of visits to five categories of provider are correctly predicted. With these characteristics, 50 percent of visits are correctly predicted.

Leonard (1998) develops explicit representations of the contracts discussed above and estimates their parameters by matching predicted patterns of patient behavior to observed patterns. This leads to the estimate that at a government clinic, the benefit to cost ratio of the marginal unit of unobservable medical effort is 16 to 1. This is a significant transaction cost: if a functioning market for medical effort existed, the marginal benefit should be equal to the marginal cost. Patients are willing to pay significantly above cost for medical effort and yet the market cannot clear. At NGO facilities the gap between benefit and cost is half that of its government counterpart. Furthermore, he estimates that practitioners at NGO facilities face penalties over 50 percent greater than government practitioners for the same violation of protocol. This is strong evidence that NGOs reduce transaction costs from asymmetric information and that patients behave in manners consistent with the belief that NGOs are an effective institutional solution to problems of asymmetric information.
4. Implications for policy

4.1. Reforming government services

The finding that decentralization in church medical facilities leads to high powered incentives that in turn provide high quality service might seem to support efforts to improve government services through their decentralization. There is a general enthusiasm about decentralization at the moment. But the failure to conceive of it as a political process can lead to serious naiveté about its likely impacts (Bossert, 1998). The crisis in African health care is due to a breakdown in the functioning of the state on the continent. This comes in part from the corrosive effects of patronage on governmental institutions. If control over health institutions is being decentralized from the national government to elected local governments (devolution) or field branches of the central government (deconcentration), there is no good reason to expect that patronage will be less present at these levels. It may well be true that elected municipal and district government officers are more directly responsive to their constituents than appointed central government officials are. But which pressures do local politicians feel most acutely: those for better service and lower costs, or those for more jobs and generous contracts for their supporters? Unless the institutional setting of local politics in Africa were to be quite different from the national one, it would be naïve to expect anything other than the latter.

Deconcentration almost certainly will not work any better. If local employees of the national government are given autonomy to make operational decisions and manage tax and fee resources but are constrained by civil service regulations and guaranteed by the national budget, facility managers will be unable to exercise the discipline over personnel that improved quality requires. In these circumstances, deconcentration is likely to mean only that the facility is run for the benefit of its employees, with little in the way of improved service for the public.

Delegation of authority to autonomous government corporations is more promising. Autonomy means that these organizations will be jeopardized if their accounts do not balance (even if they receive certain subsidies for the public services they provide). It also means that they are free of civil service regulations and in principle can run a more disciplined personnel system. If each government facility were a professionally controlled corporation subject to national regulation, it would look more like the decentralized mission facilities that Mliga found worked so well in rural Tanzania.

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 their role, but are there ways of encouraging 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 MD’s could keep their private 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, & Munishi, 1992, p. 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 led 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 its capacity for regulation to expand the reach of its organization.

However, a 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. Even these 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. Established religious organizations have not shown any desire to extend their services using this method.

On the other hand, 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. Mbongo Baptist hospital and Mbongo 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 is external. The fact that NGOs are reluctant to assume the role of a franchiser is probably due again to weak, or non-existent, contract enforcement mechanisms, i.e., ineffective courts. Hierarchical organizations can provide their own enforcement mechanisms without recourse to the outside.

4.2.2. Contracting government services out to NGOs

Rather than having NGOs rescue private practice, many African countries have tried to have NGOs rescue government health services. Leonard (1998) shows, using simulations, that if government health services held their personnel to standards more like those of NGOs, they could increase fees charged and still improve patients’ utility. Reducing the transaction cost imposed by asymmetric information can make everyone better off. We doubt, however, that more effective government regulation is a practical policy proposal. If, on the other hand, governments contracted with NGOs to provide services, they could increase quality and reduce expenditure simultaneously.

A number of African governments contract out health services to NGOs. Gilson et al. (1997) discuss the experience of Zimbabwe, Tanzania and Ghana in which mission health services were contracted to run 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 evidence on 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, consistent with our theory that NGO hierarchical supervision reduces transaction costs from asymmetric information. Gilson et al. (1997) suggests that improvements are 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—further evidence of the importance of decentralization and incentives. It is important to note that these improvements came without any effective regulation on the part of governments. Other than the work done for our project, however, there is little in-depth comparison of quality in terms of outcomes or patient perceptions. Thus, the evidence of Ly (2000b), Mliga (2000) and Ndeso-Atanga (2000) that NGO facilities in three different countries are providing superior services is particularly important.

Mills (1997) suggests that competition for such contracts might improve outcomes, as the provider who can meet contract terms at the lowest cost will win a correctly implemented bidding. If NGOs differ in the degree to which they mitigate losses from asymmetric information, the government should prefer to contract with the NGO implementing the most effective solution. Under what circumstances is the low cost provider the most effective provider?

As a weak regulator, the government cannot force the winner of a contract to provide high quality services and therefore bidding will be successful only if low quality providers cannot outbid high quality providers. All contracts observed involve a subsidy on the part of the government, thus outbidding another provider involves running a hospital for a smaller subsidy than any other potential provider. As the government is unable to enforce any but the most trivial standards on the winner, at the very least, bidding will drive the subsidy to the minimum cost of meeting these standards. Any larger subsidy would be marginally outbid by a provider who intended only to meet the standards, and offer no real services. If holding the contract has no direct benefit, competitive bidding will reduce the cost of the subsidy but will produce no, or minimal, services and quality.

On the other hand, if holding the contract offered some possibility of positive profits, bidding might produce a negative subsidy or fee. This would eliminate bids by those who did not intend to provide any services. Positive profits (when the subsidy is zero) are only possible when clients pay significant fees. A high quality provider with an established reputation will attract more clients for a broader range of conditions than a provider with a lesser reputation. Attracting more clients means higher revenues per unit and if there are economies of scale, the higher quality provider could outbid other providers. It is reasonable to assume that between two high quality providers, the one with the better reputation will earn higher revenues. However, we have no evidence that high quality providers earn more revenues than providers who specialize in selling medication and observable skills. The ‘race to the bottom’ observed among private health care providers attracts entrants and appears to be profitable. There are many ways to run a hospital, and we cannot presume that high quality care for a full range of services is more profitable than niche markets. Furthermore, in each of the cases discussed in Gilson et al. (1997), the contractors supplemented hospital income and government subsidies with their own funds. Even after subsidies, therefore, contractors were not making profits in these hospitals. The government is providing a subsidy to NGOs that is contingent on service expansion, but NGOs clearly have some intrinsic organizational interest in expansion. Bidding retains its use
in this context if it identifies least cost expansions, but the evidence suggests that NGOs are unlikely to pay a fee to expand. Unrestricted competitive bidding might produce a provider who intends to offer limited services.

If the government, on the other hand, can identify and restrict bidding to providers with established reputations for quality service, allow the winner to charge significant fees, and force it to publicly identify itself with the hospital in question, it is possible that competitive bidding could identify high quality providers at least cost. Restricting bidding to high quality providers eliminates organizations that might be tempted to offer no services or offer only niche products. The income derived at the original facility from the reputation built up over multiple years of privately subsidized service is the ‘stock of value’ which serves as a ‘bond’ for good quality service at the newly contracted facility. Publicly identifying the winner with the hospital eliminates the temptation for high quality providers to alter their pattern of provision or resell the contract, since such a decision would have an adverse impact on their reputation and would affect the services they sell at other health care locations. These restrictions reduce the number of potential bidders, and they require the government to prequalify bidders but they do not require ex post regulation and are therefore potentially implementable.

5. Conclusion

Markets for goods that suffer from asymmetric information require mechanisms in addition to prices in order to function properly, if at all. Health care is such a market. However, we do not observe one mechanism or institution governing health care markets throughout the world, but rather a great variety of institutions. Health care in Africa is characterized by an environment without functioning courts or effective regulation by either practitioners or governments. We have shown that this external environment severely restricts the potential for private markets. Without large individual investments in reputation, private practices are not likely to succeed across the full spectrum of the medical market. Instead, the empirical evidence suggests that they engage in a ‘race for the bottom’ and concentrate on the delivery of observables—pharmaceuticals and image-producing diagnostics. On the other hand, this lack of constructive environment highlights the existence of institutions, such as traditional medicine and child-birth ‘appreciation’ contracts that might not otherwise exist.

In this environment we find successful NGO health systems. NGOs differ from private providers in they can break the budget of value created when a patient is treated. This allows them to use implementable contracts even when outcomes are not verifiable. Implementation without verifiability is important because it allows contracts to exist without the backing of courts or other dispute resolution services. (As we have seen, such enforcement mechanisms do not exist in Africa.) NGOs also differ from government health services in their ability to use hierarchical control to reward and punish their practitioners. And, finally they have a stock of value to loose in the reputations they have built up as quality medical providers.

Empirical evidence shows that the features of NGOs that matter most to quality outcomes are their decentralized flexibility in dealing with staff. Patients choose to visit NGOs in a manner that is consistent with knowledge of the role of asymmetric information in the consumption of health care as well as the fact that NGOs use an effective mechanism to insure unobservable
medical effort. NGO medical facilities therefore have a set of attributes that make them effective in solving the asymmetric information problem in Africa. Privatization of health care on the continent is likely to make rapid progress and be of acceptable quality only if it builds on NGOs rather than for-profit providers. In this article we have used the tools of the New Institutional Economics and combined them with our seven-country field work to demonstrate why. The prominence of NGO medical facilities in Africa is not simply an historical anomaly; they overcome important imperfections in the market for health services on the continent.

Notes

1. The details of these contracts were related to the author by G. Njiru.
2. In the general literature, these mechanisms are referred to as Vickery–Groves mechanisms (Clarke, 1971; Groves, 1973).
3. This mild propensity to tell the truth has bothered other writers (it is a weakly dominant strategy) and other solutions have been proposed (see Moore, 1992 for a summary). The reality of the contracts actually observed makes these theoretical issues less important.
4. They are the mechanism whereby most real professional self-regulation and quality enhancement takes place in the United States (Freidson, 1970).

Acknowledgments

This paper owes a significant intellectual debt to David K. Leonard and the other contributors to the research project from which this paper grew. The framework for the paper was developed with David Leonard while he was editing the volume cited within and the work on decentralization in particular is heavily influenced by his experience. I am grateful to William Jack for his extensive and helpful comments on an earlier draft, as well as participants at the 1999 ABCDE conference at the World Bank.

References


Gilson, L. et al. (1997). Should African Governments contract out clinical health services to church providers? In Bennett et al. (Eds.), Chapter 17.


McPake, B. (1997). The role of the private sector in health service provision. In Bennett et al. (Eds.), Chapter 2.


