Moral Hazard in Health Care: An Empirical Investigation in Rural Cameroun*

Kenneth L. Leonard†

Columbia University

September, 2000

*This work was funded by NSF grant number 94-22768 and a University of California Rocca Fellowship.

†Department of Economics MC 3308, 420 W 118th Street, Columbia University, New York, New York 10027 KL206@columbia.edu

Abstract

Africa in general and rural Cameroun specifically offer a unique opportunity to investigate the impact of moral hazard on markets for health care. In Mbonge sub-district of South West Cameroun patients can choose between the government health system, church–operated (mission) health facilities and, importantly, traditional healers. Traditional healers provide health services on an outcome–contingent basis whereby patients pay a small fixed fee up front and a much larger fee only if they are cured. Both government and mission facilities, in contrast, are paid on a fee–for–service basis. We show evidence that patients behave as if they are aware of the different incentives to provide unobservable diagnostic effort at these different providers. Patients choose when to visit certain types of practitioners in ways that are consistent with an understanding of the comparative advantage of each contract type.

These features, namely variation in contract, and sophisticated patterns of choice based on this variation allow an investigation into the nature of moral hazard, its costs, and the manner in which contracts reduce and patient behavior mitigates its costs. We show that in the absence of moral hazard patients would increase the utility they get from health care by at least 170%. In a common health seeking pattern, patients are observed to frequently bypass facilities in order to seek care at much more distant facilities. We show that, by exercising choice in this manner patients increase their average expected utility by 13% despite significant additional travel costs. We verify the conventional wisdom that mission services are of higher quality and show that mission clinics provide, on average, over 1 and a half times as much diagnostic effort per patient as their government counterparts. In addition we show that if government facilities were able to consistently provider higher levels of effort, patients would significantly benefit, even if they had to fully compensate clinicians for this additional effort.

Traditional healers in Cameroun are paid on an outcome–contingent basis, where payments are linked to the recovery of the patient. On the other hand, organizational providers (government clinics and hospitals and church–based clinics and hospitals) are paid a fixed fee at the time of consultation. Is this 'custom' of payment method at the traditional healer a response to a problem of imperfect information in the supply of medical care? Eswaran and Kotwal (1985) suggest that share–cropping is a response to imperfect information in the supply of factor inputs owned by land–lords and tenants. Because different crops require different levels of inputs, one form of contract might be particularly appropriate for some crops but not others. We suggest that contingent–payment contracts are appropriate for some health production technologies and that fixed fee contracts with practitioners regulated by their employers are appropriate for other technologies, where a technology in health care is the medical response indicated by a set of presenting conditions. Both contracts fail to achieve the full information solution but by choosing between the contracts appropriately patients can mitigate the costs of moral hazard.

There are distinct patterns in the types of diseases that are reported at traditional and organizational providers (Leonard 2000). Diseases reported at traditional healers are characterized by high returns to medical and patient effort. We fit a contractual model of health care demand to data on observed patterns of provider and contract choice from the South West Province of Cameroun. Effort exerted on behalf of the patient's health is unobservable and is therefore only delivered according to the incentives that exist within the implicit contract between patient and provider. Patients create an approximate market for medical effort by choosing between discrete contract types.

We extend the Grossman (1975) model of investment in health capital by assuming imperfect factor markets as developed in Eswaran and Kotwal (1985) Both provider and patient play a bilateral–effort principal–agent game as in Hölmstrom (1982). With this relatively simple specification of incentives we show that, because diseases require a different mix of patient and medical inputs in their treatment, different types of contracts are better for different diseases.

This paper focuses on patients' choices between the five most commonly visited types of providers. The government of Cameroun runs clinics and hospitals. In addition there are a variety of clinics and hospitals run by churches. Both government and mission hospitals are similarly staffed and equipped, and both types of clinics are similarly staffed and equipped. All government centers are similarly managed and mission centers are similarly managed. Hospitals and clinics differ according to skill and government and mission facilities differ according to management. The fifth choice are traditional healers who remain popular among all ages and classes of rural Cameroun. We interviewed traditional healers[1] and examined secondary sources for information about the practice of traditional medicine[2]. Traditional healers have very different incentives than other providers because they accept payment for services contingent on a successful outcome, whereas the other four providers only accept fixed payments. Incentives to provide effort at clinics and hospitals come in the form of penalties from employers when standards for care are not met. Mission centers have the potential to impose significantly higher penalties than do their government counterparts. Each provider offers a different mix of skill and incentives to provide effort.

Our data include the characteristics of the disease or illness conditions from which patients suffered as well as characteristics of the patient and the expected costs at each practitioner. By assuming that choices are made based on expected net utility we use a conditional logit structural estimation to recover the parameters of the production–of–health–investment function as well as the parameters of the contracts between patients and providers for the delivery of medical and patient

[1]By traditional healers we mean rural health practitioners who run practices that resemble health practices that existed before the spread of 'western' medicine into the rural areas. We do not imply that all traditional healers use herbal medicines, nor that no non-traditional practitioners use herbal medicines. Our distinction is by method of practice not by types of medicines used.

[2]For details of the interviews see Leonard (2000). Secondary sources were Korse et al. (1989), Baerts (1989), Edwards (1983), Oyenye and Orubuloye (1985), Lasker (1981), Staugård (1985), Gelfand et al. (1985) and Conco (1972).

effort.

The willingness of patients to travel large distances and pay significant fees and drugs costs to seek health care offers variation that allows us to estimate both utility. The fact that traditional healers are paid on an outcome–contingent basis allow us to model the relationship between the value of medical effort and its disutility.

These features allow an investigation into the nature of moral hazard, its costs, and the manner in which contracts reduce and patient behavior mitigates its costs. We show that in the absence of moral hazard patients would increase the utility they get from health care by at least 170%. In a common health seeking pattern, patients are observed to frequently bypass facilities in order to seek care at much more distant facilities. We show that, by exercising choice in this manner patients increase their average expected utility by 13% despite significant additional travel costs. We verify the conventional wisdom that mission services are of higher quality and show that mission clinics provide, on average, over 1 and a half times as much diagnostic effort per patient as their government counterparts. In addition we show that if government facilities were able to consistently provider higher levels of effort, patients would significantly benefit, even if they had to fully compensate clinicians for this additional effort.

This paper is organized as follows. In the following section we outline our model of health production. We develop an explicit characterization of the contract available at each provider and the levels of effort that the patient can expect at each provider. In Section 2 we discuss the data that we collected in Cameroun as well as the potential issues presented by the data. Section 3 presents the results of reduced form tests on this data. A structural estimation of the model is presented and the results are discussed in Section 4. Section 5 concludes.

1 Health Care with Asymmetric Information

We begin with an individual who has fallen sick from an unknown disease (but a known illness condition, where the illness condition is described by the symptoms of the patient). The given level of health is H. Health intervention might lead to a change in the level of health, ΔH. We simplify the idea of health intervention by assuming that 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. We think of $\bar{h}$ as being a full recovery and $\underline{h}$ as being no change in the health status.

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. 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 probability of full recovery of $\rho^\star$. If the diagnosis is incorrect the probability of recovery is $\rho^\emptyset$. The probability of failing to recover is 1-$\rho^\star$ with the 'true diagnosis' and 1-$\rho^\emptyset$ with the 'false diagnosis.' 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 should be and the patient does not recover. On the other hand patients frequently recover when nothing has been done for their health (or when incorrect actions have been taken).

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 ($0 \le e \le 1$) where

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

Thus the 'best' technology has a 100% chance of correct diagnosis and leads to a chance of recovery of $\rho^\star$, and the 'worst' technology has a 100% chance of choosing among the incorrect diagnoses and leads to a chance of recovery of $\rho^\emptyset$.[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 (e). e is generally a function of patient effort, patient skill, practitioner effort and practitioner skill. Unobservable efforts imply that the patient does not ever observe e, only whether the outcome was $\bar{h}$ or $\underline{h}$. Since both outcomes are possible with all e the patient can never impute physician effort even if she knows her own level of effort, her own skill and the practitioner skill. Thus, patients can only expect incentive compatible effort which varies according to the means of physician compensation. Therefore, patients can only affect e through their own level of effort.

1.1 The Value of Health

Utility from health can be modeled in a variety of different ways. We follow the basic model of Grossman (1975) and consider health as increasing the hours of time available to consume work and leisure as well as augmenting utility directly. Thus $U = (H, I(H), c(p))$, where p is patient effort, $c(p)$ is the disutility of patient effort and $I(H)$ is the income potential at health level H.

The expected value of health is

$$\begin{aligned} EU = e\rho^\star \bar{U} + e(1-\rho^\star)\underline{U} + (1-e)\rho^\emptyset \bar{U} + (1-e)(1-\rho^\emptyset)\underline{U} \\ \bar{U} = U[\bar{h}, (I(\bar{h}) - C), c(p)] \\ \underline{U} = U[\underline{h}, (I(\underline{h}) - C), c(p)] \end{aligned} \tag{2}$$

C is the total cost of a visit. Of interest to the patient is the change in expected

[3]We deliberately based this description of ΔH on the Spanning Condition of Grossman and Hart (1983) and the Linear Distribution Function Condition of Hart and Hölmstrom (1987), which will allow us to characterize incentive compatibility constraints as first order conditions or relaxed incentive compatibility constraints.

utility. We choose as a natural comparison the utility when no health care is sought. We assume a separable utility form such that $U = U'[H, I(H)] - C - c(p)$. 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 ($U'[H, I(H)]$) depends on the outcome, not on the effort exerted. Costs and disutility have a linear relation to utility. 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

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

Using the separable utility function the expected utility of seeking care is

$$EU = \left(e(\rho^{\star} - \rho^{\emptyset}) + \rho^{\emptyset}\right) \bar{U}' + \left(1 - \rho^{\emptyset} + e(\rho^{\star} - \rho^{\emptyset})\right) \underline{U}' - C - c(p) \tag{4}$$

The change in the expected utility is

$$\Delta EU = e(\rho^{\star} - \rho^{\emptyset}) \cdot \left(\bar{U}' - \underline{U}'\right) - C - c(p) \tag{5}$$

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, $\bar{h} \cdot \bar{w}$ (where $\bar{w}$ 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 $\omega = \bar{w} + w$.) Thus

$$\Delta EU = e(\rho^{\star} - \rho^{\emptyset})\omega\bar{h} - C - c(p)$$

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 $\bar{h}$.

$$\Delta EU = (\rho^{\star} - \rho^{\emptyset})\omega h - C - c(p) \tag{6}$$

For simplicity we will refer to ΔEU as U. We end up with a functional form that is functionally equivalent to the simplest form we could try and write down, with the exception of the $\rho^{\star} - \rho^{\emptyset}$ term. In particular h is continuous. However, by beginning with the assumption that there are only two health outcomes and that utility derived from income and disutility from costs are additively separable we allow the use of relaxed incentive compatibility constraints and avoid risk aversion, respectively.

1.2 The Health Production Technology

The health production technology (h) is viewed as a search for the proper treatment regime. This search 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.

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. Expected utility for the patient is

$$\Delta EU = \omega(\rho^{\star} - \rho^{\emptyset})h - cc - fc - dc - tc - c(p) \tag{7}$$

Costs are composed of contingent costs (cc), fee costs (fc), drug costs (dc) and travel costs (tc). Contingent fees (which are used only at traditional healers) are endogenously determined and depend on patient, practitioner and illness condition characteristics. 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. The fixed cost is a function only of the practitioner. The income of the practitioner is the sum of all payments by the patient, not including travel costs.

1.3 Incentives at Traditional Healers

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. Both 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.

1.4 Organizational penalties

Government and church–operated clinics and hospitals charge a fixed fee for consultation and charge for all drugs administered. Both government and mission health centers operate to serve the health of their clients; they are not profit making entities. Thus, though the practitioner does not have a direct incentive to exert effort, his employer has an incentive to induce effort. The employer of the practitioner does not observe the outcome of health care, but does observe other outcomes that give information about the effort of the provider. Practitioners produce both health for the patient and what we call *organizational quality*. This second output is observed by the employer. Records are kept of the various activities that go into producing health. Typically a selection of records are examined during a site visit. The patients' symptoms and complaints are part of all records and therefore procedures and records should follow protocols developed for each set of complaints. If a particular record or collection of records is determined to be in violation of standards the practitioner is punished in accordance with the gravity of the deviation. This method of ensuring quality is what we refer to as a penalty–based scheme.

In practice, centers with stronger incentives use discretionary bonuses, the threat of termination and salaries levels to encourage the provision of effort. Mliga (2000) reports that, in Tanzania, where he studied 4 different health care provision systems, those organizations that had the power to use these forms of incentives provided significantly superior quality of care, as judged by other clinicians. We use the notion of penalties as a simpler modeling method but it should be thought of as capturing elements of all of these practices.

When an organization can force a practitioner to produce high organizational quality it is also forcing the practitioner to exert medical effort, even though the patient's health is never observed. The goal of protocols is to find the correct diagnosis; to increase e. This system of incentives is different from that of traditional healers. Because the employer does not observe the outcome of the treatment, the decision of whether or not to punish, or by how much, is independent of the effort of the patient.

The probability of being visited and observed, or of a record or set of records being examined, is fixed within and varies between organizations. Once the data from a particular consultation is observed, the organizational quality, $Q(m)$, is known with certainty. This is then compared to the required quality, Q^* and the punishment is proportional to this difference. The expected value of the penalty is

$$g(m) = v \cdot f\left(Q^* - Q(m)\right) \tag{8}$$

v (visit) is the probability, for any given organization, of a record being examined and f is the baseline penalty (forfeiture), for any particular organization. v and f

cannot be identified separately thus we refer to the product as F.

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.

$$E(U^m) = \underbrace{r\omega(\rho^\star - \rho^\emptyset)h(m,p)}_{\text{contingent fee}} + \underbrace{dc}_{\text{drugs}} + \underbrace{fc}_{\text{fixed fee}} - \underbrace{g(m)}_{\text{penalty}} - \underbrace{d(m)}_{\text{disutility}} \tag{9}$$

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.

Penalties are important in our analysis because they provide the incentive to exert effort. We hypothesize that the penalty basis at mission centers is larger than at government centers (where all practitioners are protected from the most severe penalties because they are civil servants) and therefore practitioners at mission centers exert more effort for every condition than their government counterparts.

1.5 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. Note that all actions must be incentive compatible, even the actions of the patient. Thus for patient who visits a traditional healer the problem can be represented as

$$\max_{r,m,p} EU = (1-r)\omega(\rho^\star - \rho^\emptyset)h - fc - tc - c(p) \quad \text{(P. utility)} \tag{10}$$

$$m^* \in \operatorname*{argmax}_m r\omega(\rho^\star - \rho^\emptyset)h + fc - d(m) \quad \text{(I.C. medical)} \tag{11}$$

$$p^* \in \operatorname*{argmax}_p (1-r)\omega(\rho^\star - \rho^\emptyset)h - fc - tc - c(p) \quad \text{(I.C. patient)} \tag{12}$$

$$r\omega(\rho^\star - \rho^\emptyset)h + fc - d(m) = V_m \quad \text{(I.R. medical)} \tag{13}$$

where 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 patient'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:

$$\max_{r} EU = \omega(\rho^{\star} - \rho^{\emptyset})h - tc - c(p) - d(m) - V_m \tag{14}$$
$$r\omega(\rho^{\star} - \rho^{\emptyset})(\frac{\partial h}{\partial m}) - \frac{\partial d(m)}{\partial m} = 0 \tag{15}$$
$$(1-r)\omega(\rho^{\star} - \rho^{\emptyset})(\frac{\partial h}{\partial p}) - \frac{\partial c(p)}{\partial p} = 0 \tag{16}$$

For organizational providers the patient faces the following optimization

$$\max_{r,m,p} EU = \omega(\rho^{\star} - \rho^{\emptyset})h - fc - dc - tc - c(p) \quad \text{(P. utility)} \tag{17}$$
$$m^* \in \operatorname*{argmax}_{m} dc + fc - g(m) - d(m) \quad \text{(I.C. medical)} \tag{18}$$
$$p^* \in \operatorname*{argmax}_{p} \omega(\rho^{\star} - \rho^{\emptyset})h - fc - dc - tc - c(p) \quad \text{(I.C. patient)} \tag{19}$$
$$fc - g(m) - d(m) = V_m \quad \text{(I.R. medical)} \tag{20}$$

Note that the objective function for the practitioner does not contain any information about the patient, only the illness condition. Thus the patient will take medical effort as fixed and optimize utility with respect to her effort. The major difference between the incentives at organizational providers versus those at traditional healers comes from the fact that the employers of practitioners at organizational providers never observe the outcome of treatment, whereas traditional healers do.

1.6 Joint determination of patient and medical effort

Under the contract offered at traditional healers the patient can be sure that she is getting more effort than if she paid a flat fee. However, if both medical and patient effort are unobservable and important in the production of health she will not be able to obtain the full information solution. On the other hand, if the employer of the practitioner judiciously chooses the level of medical effort for each condition and for each patient the full information solution can be achieved. This will, in practice, not be the case for the following reasons. First, the employer monitors quality after the event not during the consultation. As a result he will have difficulty collecting information about the characteristics of the patient or the level of effort exerted by the patient. Furthermore, since illnesses are rarely cured during consultation, the regulator, or employer, will in general never know the outcome of treatment. The penalty or reward cannot be contingent on the outcome, the effort of the patient, or the characteristics of the patient.

If the optimal level of medical effort does not depend on these two things there is no cause for concern. However, if medical and patient effort are complements (and we propose that they are) the optimal level of medical effort should depend on the level of effort of the patient. Furthermore, since the effort of the patient depends on her individual characteristics (her opportunity cost of healthy time and efficiency at transforming health inputs into health) medical effort should depend on these as well.

Having observed the mechanism used to reward and punishment practitioners we know that patient information is not collected. Furthermore, we can infer from the fact that practitioners pay no attention to patient effort that they are not properly reacting to it. This then opens a crucial difference between the contracts available at traditional healers and at organizational providers: medical effort at traditional healers correctly reacts to the responsiveness to patient effort where as medical effort at modern providers does not.

If our assumptions about the nature of organizational provision are correct, patients are most likely to visit traditional healers when they suffer from conditions that are responsive to both medical and patient effort. This hypothesis is more rigorously developed (using the functional forms for health assumed below) in Leonard and Zivin (2000).) We will show evidence that this hypothesis is born out in the data in section 3. However, for the purposes of the structural estimation we do not assume any particular form for organizational quality and allow the estimation to describe organizational quality. The results of this estimation show exactly the form we have described – the elasticity of medical effort with respect to responsiveness to medical effort decreases as the responsiveness to patient effort increases. An optimal reaction would dictate that it increase.

2 Data

Data on patient behavior in the face of illness were collected in Mbonge sub-division, in the South-West province of Cameroun. The sub-division is entirely rural. This area was chosen because of the presence of a German aid project that insured a consistent, reasonably-priced drug supply in all government health centers and hospitals.

40 villages were randomly chosen and 20 randomly selected households from each village were interviewed. Data were collected on all members of the household. There were 681 illness episodes reported within the 1 month recall period out of 4,489 individuals represented. Of these, 548 visited one of the five types of providers we are studying. In addition, 53 people reported not seeking care at any health care provider. We have complete and consistent data for 584 visits. All reported visits to providers are first contacts.

2.1 Data Issues

We have developed a model in which the value of health care depends on the opportunity cost of healthy time, the difference in the probability of a cure with the correct versus the wrong diagnosis, the responsiveness of a condition to medical and patient effort and skill, the cost of drugs, fees and travel, and the disutility incurred by patients. Our data contains information about patients, how far they live from each potential provider, the symptoms and other characteristics of the condition from which they suffered, the location they chose and the costs they incurred. There are two particularly difficult obstacles to surmount with the data. The first is that individual characteristics can play a role in the illnesses from which patients suffer as well as the opportunity cost of time. The second is that the characteristics of the illness condition affect both the expected value of health care at a given practitioner and the expected cost of drugs at any practitioner. We address each of these in turn.

The correlation between individual characteristics and illness conditions. Children suffer from different conditions than adults and have a different opportunity cost of time. We know that children are disproportionately represented at government clinics in this sample area, and in general throughout Africa. Is this because children have a lower opportunity cost of time or because they suffer from different illness conditions? Empirical analysis which controls for age and symptom, for example, will be unable to separate these two affects because dehydration in a child is fundamentally different from dehydration in an adult. This particular example would require (at the very least) an interaction term between age and symptom. Our data include 21 symptoms that were observed in at least 10 cases; 70% of episodes involved two or more symptoms, 29% involved 3 or more and 6% involved 4 or more symptoms. Interacting individual characteristics like age, gender and wealth with all or even most symptom pairings is beyond the capacity of this data set.

The problem is not unsolvable. The illness from which a person suffers is still exogenous given individual characteristics. People who live near mission clinics cannot choose to become ill with conditions that are particularly well treated at that type of clinic. The problem we face is to reduce the dimensions of the data without loosing information. We do this by introducing a series of continuous variables and assigning values to these variables for each unique condition using information about symptoms, other characteristics of the illness and salient individual characteristics. This works because these codings are the output of an expert medical evaluations of each case, as explained below. Furthermore we will show, in section 3, compelling evidence that this solution reduces the dimensions without sacrificing too much information. As a result we can use the codings to infer information about the expected outcome of health care and individual characteristics to infer information about the opportunity cost of health time.

Correlation between expected health outcomes and drug costs. The cost of drugs at a modern provider is correlated with the symptoms presented. In general we can assume that the greater the costs patients are willing to bear, the greater the potential benefit. But the opposite is not true. Inexpensive medicines can mean the difference between life and death and we would be wrong to assume that low drug costs imply low expected returns. We cannot know whether a provider was not visited because patients did not expect he could cure their condition or because high drug costs reduced the net benefit sufficiently to deter the visit.

If we knew the expected cost of drugs at each potential provider we could control for this but in our data we only know observed costs at the provider chosen. We need to estimate expected costs at all providers. The best way to do this is to estimate expected costs using information about the symptoms and other characteristics of the illness. We could then use expected costs for every provider as a variable in the estimation. If we were using symptoms to determine both the expected outcome of health and the drug costs we would have an identification problem, however, as explained above we will be using specially constructed variables to determine the expected outcome of health, freeing us to use symptoms to estimate the cost of drugs.

The solution to both of these problems is a nonlinear translation of discrete symptoms into continuous illness characteristics. The non-linearity is achieved by the expert evaluation rather than some complex functional form, but the effect is the same.

2.2 Variables

We now turn to a more detailed look at variables in the survey data, the variables estimated from this data, and the variables created from expert analysis of this data.

Illness Condition Data (β, α, π, σ) We collected information from respondents on the characteristics of the episode from which they suffered: all symptoms they experienced; the self–declared severity of the disease; the number of days sick before seeking care; and the number of those days the patient was bedridden. With these characteristics, the age and sex of the individual and information about endemic diseases in the area, two doctors and one nurse (all experienced in tropical medicine) independently score all the cases using the following definitions. All coding was blind of provider chosen and diagnosis.

Responsiveness to Medical Effort ($\tilde{\beta}_k$) is the degree to which outcomes depends on the effort of the practitioner.

Responsiveness to Patient Effort ($\tilde{\alpha}_k$) is the degree to which outcomes depends on the effort of the patient.

Benefit of skill and capacity ($\tilde{\pi}_k$) There are three levels of skill and capacity between which patients can choose. The benefit at a higher level facility is always at least as high as the benefit at lower level facilities

Informally trained personnel are people who are in the health care profession and may have practiced for many years, but who never completed a formal medical training program. Their experience can be sizable but it will not be based on a foundation of western medical training. Common drugs are available to them. *Formally trained personnel in a clinic* have some formal training and work in a facility that has a basic drug supply, beds, IV equipment and a delivery room but not much more. *Highly trained personnel in a hospital* have advanced formal training and practice in a hospital with a much greater supply of drugs and equipment for surgery, long term care, etc. General practitioners would also be expected to have reasonable access to the services or advice of specialists.

Illness Severity (σ) The potential severity of the illness condition.

In addition to these three sets of scores, we created scores for each case using basic medical references (Griffith 1985, Strickland, ed 1984, Werner 1977). This last set of scores is more consistently correlated with the behavior of patients and we will in general focus on this set of scores.

Patient level data Of interest to the structural estimation is the opportunity cost of healthy time and the patient's efficiency at converting health inputs into health. Neither is observable in the data set and we substitute for these two variables a set of variables that we do observe and hypothesize is related to the opportunity cost or efficiency. ω and η will be estimated within the structural estimation as the combination $\omega \cdot \eta$; the efficiency–adjusted value of health. The fact that the functional specification includes the term $\omega \cdot \eta$ is useful because we have little hope of separating out the different effects of the instruments. To estimate the opportunity cost of time we include variables like occupation, education and observed wages but these are clearly determinants of the efficiency of transforming health into health care as well.

The variables that will be used include the reported wage of the patient converted to a weekly rate, the log of instrumented family wealth, age, age squared, gender and education. Family wealth is derived in the following way. The total yearly earnings of each household are regressed on the occupations and education level of all its residents, the ownership of durable and consumption goods, ownership of animals and the construction materials used in the house. The log of predicted total family income is then used as a measure of wealth. In this manner retired individuals and families with remittance income are not qualified as poor. In addition the age, gender, education level and occupation dummies of the individual who looked after the patient for episodes involving children and the old or infirm. When a patient looked after herself these variables reflect her own characteristics. We do

Table 1: Mean Illness Condition Characteristics by Provider Visited

variable	no care	gov cln	mis cln	gov hos	mis hos	trad heal
Medical References						
β	3.97	4.63	4.96	5.51	5.44	4.93
α	3.67	3.81	3.77	3.98	4.18	4.61
$\alpha \cdot \beta$	15.29	18.97	19.27	22.72	22.21	24.79
π_h	4.13	4.64	4.81	5.23	4.92	3.83
$\pi_{c,k}$	3.92	4.19	4.02	4.33	3.87	3.43
$\pi_{t,k}$	3.19	3.27	2.97	3.18	3.02	2.95
σ	3.43	3.88	4.10	4.60	3.75	3.04
Coding 1						
β	5.10	5.64	5.48	5.93	5.85	5.41
α	4.63	5.21	5.14	5.27	5.45	5.38
$\alpha \cdot \beta$	24.31	30.18	28.76	31.87	32.17	30.30
π_h	6.12	6.62	6.69	7.06	7.08	6.62
$\pi_{c,k}$	5.45	5.94	5.93	6.17	6.15	5.79
$\pi_{t,k}$	4.31	4.47	4.50	4.59	4.45	4.07
σ	3.75	4.47	4.69	5.15	5.13	4.48
Coding 2						
β	6.10	6.07	6.05	6.75	6.55	6.64
α	5.73	5.40	5.36	5.35	4.87	5.33
$\alpha \cdot \beta$	35.98	33.21	33.13	36.58	32.72	36.16
π_h	7.25	7.34	7.55	7.79	7.62	7.74
$\pi_{c,k}$	5.67	5.31	5.23	5.17	4.92	4.57
$\pi_{t,k}$	3.63	3.14	2.83	2.85	3.62	3.48
σ	4.73	4.62	4.67	5.07	4.13	4.79
Coding 3						
β	7.42	7.65	7.67	7.58	7.40	7.41
α	4.71	4.90	4.99	4.98	4.40	4.21
$\alpha \cdot \beta$	35.13	37.82	38.77	38.02	33.43	31.25
π_h	7.42	7.66	7.69	7.57	7.40	7.41
$\pi_{c,k}$	6.19	6.19	5.90	5.66	5.13	5.85
$\pi_{t,k}$	1.71	1.82	1.85	1.96	2.04	1.97
σ	1.94	2.50	2.47	2.36	1.89	1.98

this because it is possible that the efficiency is not an inherent biological feature but an ability to understand medical advice, read labels, etc. Proficiency in the dispensing of medicine does not depend on the literacy of a 2 year old, but that of her mother. Summary statistics of the key variables are contained in Table2

Table 2: Mean Individual Characteristics by Provider Visited

variable	no care	gov cln	mis cln	gov hos	mis hos	trad heal
family wealth	-0.13	0.02	0.19	0.15	0.43	0.11
patient						
income	1.29	1.29	3.09	1.94	9.34	2.10
age	18.98	17.69	19.72	27.02	43.04	26.56
female	0.54	0.49	0.45	0.50	0.49	0.54
education	2.31	2.35	2.04	2.88	2.38	3.28
care giver						
age	35.92	38.31	38.70	40.58	42.42	38.57
female	0.52	0.53	0.51	0.50	0.42	0.39
education	4.62	4.32	4.41	5.00	4.26	4.62
student	0.04	0.00	0.02	0.02	0.04	0.02
cocoa farmer	0.25	0.33	0.31	0.35	0.26	0.39
other farmer	0.56	0.55	0.57	0.55	0.55	0.46
merchant	0.10	0.08	0.09	0.09	0.13	0.10
other business	0.04	0.03	0.05	0.01	0.02	0.08

Provider level data Fixed costs for organizational providers are known from the data and since they do not vary by disease or individual we can assign the known fixed cost to every center.

Drug Costs (dc) We cannot use the reported cost of drugs as the expected cost of drugs because we do not observe the cost at centers not visited and drug costs are not the same across centers. The cost of drugs at mission centers is higher than the cost at government centers because the government subsidizes drugs. The cost of filling a prescription at hospitals is higher than at clinics because there are more drugs and diagnostic tests available. Note that at traditional healers the drug cost is always zero whereas the same disease would require drug purchases at any other center.

In order to estimate the cost of drugs at each center we hypothesize that the drug cost is a function of characteristics of the disease (as well as the age and gender of the patient) and that costs are related across centers by a fixed ratio.

$$dc_{jk} = \Delta_k \cdot dc_{GC,k} = \Delta_k \cdot \prod_n z_{kn}^{\beta_n} \cdot \prod_l e_{il}^{\alpha_l} \tag{21}$$

dc_{jk} is the cost of drugs at practitioner j for illness k, and $dc_{GC,k}$ is the cost at government clinics for illness k, and serves as our numeraire drug cost for each illness condition. The cost at each center is related to the numeraire cost by a ratio Δ_k, ($\Delta_{GC} = 1$) and the numeraire cost is a function of individual characteristics (e_i) and disease characteristics (z_k). With a large set of observed drug costs we can estimate $dc_{GC,k}$ and Δ_k and thus estimate the drug cost at each provider for each illness condition. Results of the log-linear regression are reported in Appendix A.

Travel Costs (tc) We have data for each village on the distance and total taxi cost to most major hospitals and clinics. Taxi costs in this area are based on per kilometer charges, depending on the quality of the road. Thus, using a large set of known taxi costs we have established a taxi cost from every village to the closest of each of the four organizational providers. Although we have data on distances traveled to visit a traditional healer, we do not know the location of every traditional healer in the sample area so we cannot reliably calculate the cost of the counter-factual visit. The average distance traveled to visit a traditional healer is small so we normalize this distance to zero. Note that, though we use taxi costs, not every patient took a taxi.

3 Preliminary Data Analysis

The regression reported in this section use data on visits to the 5 types of providers that we are studying and do not include illnesses that did not result in a visit to any provider. We would like to test that the 4 variables created to characterize illnesses contain useful information and can reasonably replace the set of discrete variables representing the presence of a symptom. To this end we present a model and empirical specification we call the 'kitchen sink,' – so called to convey the idea that everything (every variable) is used without thought to a formal model or hypothesis. The log likelihood is formed as follows.

$$\log L = \sum_{i=1}^{N} \sum_{j=1}^{5} \delta_{ij} \log P_{ij} \qquad \text{where} \qquad P_{ij} = \frac{\exp(\gamma_j' x_i + \rho' z_{ij})}{\sum_{m=1}^{5} \exp(\gamma_m' x_i + \rho' z_{im})} \tag{22}$$

where $\delta_{ij} = 1$ when patient i chose provider j and 0 otherwise. This is a mixed McFadden conditional logit and standard multinomial logit. The z_{ij} can include travel costs, drug costs, and skills and are a vector of data that varies by provider and has only one coefficient for all providers, ρ. x_i can include illness condition symptoms, illness codings, other illness characteristics, or patient characteristics. It does not vary by provider, but the coefficients, γ_j do vary by provider.

To show that illness codings contain useful information and can be thought of as a substitute for the symptoms we perform the following test. We run a kitchen sink regression with symptoms, other illness condition characteristics (self declared severity, days sick, ability to work, and days bedridden), and the illness codings as x variables and estimated cash costs (drug costs plus fees) and travel costs as z_j variables. We then drop all of the symptom variables. A likelihood ratio test of the hypothesis that the additional information contained in the symptom data is equivalent to noise is rejected at the 5% level but not the 1% level.[4] On the other hand, if we test the restriction that the coding variables contain no information, a likelihood ratio test is rejected at both the 5% and the 1% level[5] We cannot conclude that the coding variables make the original symptoms redundant (in terms of ability

[4] Full log likelihood = -604.42, restricted = -660.64, 21 variables by 4 vectors of coefficient = 84 restrictions, $1 - \chi^2(111.16,84) = 0.0253$.

[5] Full log likelihood = -604.42, restricted = -640.94.64, 11 variables by 4 vectors of coefficient = 44 restrictions, $1 - \chi^2(73.04,44) = 0.0039$.

to predict location visited) but we can conclude that the coding information is not redundant. The fact that we cannot reject the redundancy of symptoms at the 1% level, however, is evidence that we have captured a great deal of the information contained in the symptom data.

3.1 Test of reduced form hypotheses

We have suggested that if organizational quality does not take into account patient effort and therefore does not respond fully to increases in the responsiveness to patient effort we should expect illnesses characterized by a high responsiveness to both medical and patient effort to result in an increased probability of a visit to a traditional healer. This hypothesis can be tested directly. We set up the following model. The regression is similar to the mixed conditional/multinomial logit above with the exception that we restrict the coefficients on illness characteristics to be the same at each institution and include skill as a conditional vector. Thus, in terms of the effect of the responsiveness to medical and patient effort and the illness severity, government clinics and hospitals are the same, and mission clinics and hospitals are the same. Hospitals (both mission and government) face the same responsiveness to skill and clinics face the same responsiveness to skill.

The results of this regression are shown in table 3. Included is an interaction term between responsiveness to medical and patient effort. The traditional healer is the base choice and therefore a negative coefficient signifies that an increase in the variable increases the probability of a visit to a traditional healer. We find that the interaction term is positive and significant for three of the four data sets created. When the responsiveness to patient effort is high an increase in the responsiveness to medical effort increases the probability of a visit to a traditional healer. This is exactly as we hypothesized. Furthermore the log likelihood for the first data set is far higher than the other three and we will use this data in the structural estimation.

Table 3: Preliminaries: Conditional Logit Coefficients

data set	Medical Ref		Individual 1		Individual 2		Individual 3	
variable	coef	z-test	coef	z-test	coef	z-test	coef	z-test
Government clinics and hospitals								
α	**0.353**	3.30	0.177	1.55	**0.424**	3.01	0.008	0.02
β	0.040	0.32	**0.483**	3.44	0.196	1.81	-0.011	-0.12
$\alpha \cdot \beta$	**-0.085**	-4.48	**-0.065**	-2.62	**-0.055**	-2.61	0.045	0.95
σ	**0.520**	4.31	0.038	0.34	0.016	0.23	0.032	0.33
Mission clinics and hospitals								
α	**0.516**	4.49	**0.367**	2.99	**0.574**	3.89	-0.288	-0.72
β	**0.302**	2.28	**0.503**	3.33	**0.347**	2.97	0.159	1.70
$\alpha \cdot \beta$	**-0.125**	-5.74	**-0.100**	-3.71	**-0.083**	-3.65	0.07	1.37
σ	**0.421**	3.38	0.147	1.27	-0.026	-0.35	-0.04	-0.39
Conditional Logit Coefficients								
skill	**0.206**	2.71	0.040	0.44	0.027	0.82	0.013	0.31
travel cost	**-0.880**	-11.09	**-0.823**	-10.47	**-0.841**	10.67	**-0.838**	-10.63
individual characteristics (income, household wealth, schooling and adult) controlled for but not reported								
log likelihood	-714.02		-735.29		-739.09		-732.71	

4 Structural Estimation

We return now to the model that we have developed in order to estimate the unknown parameters using patient behavior in the face of moral hazard. In order to advance we specify functional forms. The health production technology is represented as a Cobb-Douglas production function.

$$h = \pi \eta p^{\alpha} m^{\beta} \tag{23}$$

where π is the skill of practitioner at curing the illness condition, η is the efficiency of the patient at turning health inputs into health, p is the patient effort, α is the elasticity of output with respect to patient effort, m is medical effort and β is the elasticity of output with respect to medical effort. We assume decreasing returns to scale ($0 < \alpha < 1$, $0 < \beta < 1$ and $0 < \alpha + \beta < 1$.) For simplicity of notation we will refer to the product of skill, efficiency, the value of health and the differences in probabilities of cure ($\omega(\rho^{\star} - \rho^{\emptyset})\pi\eta$) as A. We assume that disutility of effort is a linear function of the effort such that $c(p) = p$ and $d(m) = m$.

Recall the specification of the penalty, $g(m) = F\left(Q^* - Q(m)\right)$. We chose the functional form of $Q(m)$ in anticipation of a simple functional specification for m^*. The quality factor (ζ) is a scalar that varies by illness condition.

$$Q(m) = \ln(m) \cdot \zeta$$

Thus the penalty is decreasing (and utility is therefore increasing) in m and ζ, and exhibits decreasing marginal returns to effort.

For visits to the traditional healer with the preceding functional forms we can reduce the optimization by assuming a Nash non-cooperative equilibrium and get

$$m^{\star} = r\beta\left(A((1-r)\alpha)^{\alpha}(r\beta)^{\beta}\right)^{\frac{1}{1-\alpha-\beta}} \tag{24}$$

$$p^{\star} = (1-r)\alpha\left(A((1-r)\alpha)^{\alpha}(r\beta)^{\beta}\right)^{\frac{1}{1-\alpha-\beta}} \tag{25}$$

$$E(U^p) = (1-(1-r)\alpha - r\beta)\left(A((1-r)\alpha)^{\alpha}(r\beta)^{\beta}\right)^{\frac{1}{1-\alpha-\beta}} - V_m - tc \tag{26}$$

A closed form solution for the optimal share, r^*, is not obtainable, but the solution can be shown to depend only on β, α.

$$r^* \in \underset{r}{\text{argmax}}\,(1-(1-r)\alpha - r\beta)\left((1-r)^{\alpha}r^{\beta}\right)^{\frac{1}{1-\alpha-\beta}} \tag{27}$$

If patients did not exert unobservable effort, r^* would be 1. If practitioners did not it would be 0. The optimal share is increasing in β and decreasing in α.

For organizational providers:

$$m^* = F \cdot \zeta \tag{28}$$

$$p^* = \left(A\alpha^{\alpha}\left(F\zeta\right)^{\beta}\right)^{\frac{1}{1-\alpha}} \tag{29}$$

$$E(U^p) = (1-\alpha)\left(A\left(F\zeta\right)^{\beta}\alpha^{\alpha}\right)^{\frac{1}{1-\alpha}} - fc - dc - tc \tag{30}$$

$m^\star$ does not vary with p or A, but can vary with α, and will clearly vary with β.

Efficiency–Adjusted Value of Health and Probability of Cure In order to capture the strong possibility of decreasing marginal returns in the relationship between patient characteristics and efficiency–adjusted value of health and to force all values to be greater than 0, we use the following specification.

$$\omega(\rho^\star - \rho^\emptyset)\eta = \frac{B_1}{1 + \exp\left(-1 \cdot \left(\sum_m E_m \cdot e_{im} + \sum_l T_l \cdot t_{kl}\right)\right)} \tag{31}$$

e_{im} are the individual characteristics of the patient discussed above. t_{kl} are variables that relate to the probability of a cure and include dummy variables for the ability to work, the self-declared severity of the illness, the medically evaluated severity of the illness and the days both sick and bedridden. B_1, $\vec{E}$ and $\vec{T}$ are parameters to be estimated.

Skill The skill of a practitioner is restricted to be greater than one. We use the information on skill from the medical references as the primary source of this variable. However, we allow information from the three other codings as well, obtaining our estimate of skill in the following manner.

$$\pi = 1 + \frac{1}{1 + \exp\left(-1 \cdot \left(P_0 + \tilde{\pi}_{k\mathrm{MR}} + P_1 \cdot \tilde{\pi}_{k1} + P_2 \cdot \tilde{\pi}_{k2} + P_3 \cdot \tilde{\pi}_{k3}\right)\right)}$$

$\tilde{\pi}_{k\mathrm{MR}}, \tilde{\pi}_{k1}, \tilde{\pi}_{k2}$ and $\tilde{\pi}_{k3}$ are the estimated skills from the medical references and the three codings respectively. The use of the tilde differentiates this inputs from the π derived from the estimation. $\vec{P}$ are to be estimated.

Responsiveness to Medical and Patient Effort and Illness Severity As with the skill variable we have 4 estimates of the responsiveness to medical and patient efficiency and the illness severity. We use the estimate derived from medical references as the primary estimate, allowing the estimation to determine the additional weights given to the additional three codings. In addition the values of α and β are restricted to be between 0 and 1, and their sum cannot exceed 1. The following specification allows for flexibility within such constraints.

$$\alpha = \frac{B_4}{1 + \exp\left(-1 \cdot \left(R_{\alpha,0} + \tilde{\alpha}_{k\mathrm{MR}} + R_{\alpha,1} \cdot \tilde{\alpha}_{k1} + R_{\alpha,2} \cdot \tilde{\alpha}_{k2} + R_{\alpha,3} \cdot \tilde{\alpha}_{k3}\right)\right)}$$

$$\beta = \frac{1 - B_4}{1 + \exp\left(-1 \cdot \left(R_{\beta,0} + \tilde{\beta}_{k\mathrm{MR}} + R_{\beta,1} \cdot \tilde{\beta}_{k1} + R_{\beta,2} \cdot \tilde{\beta}_{k2} + R_{\beta,3} \cdot \tilde{\beta}_{k3}\right)\right)}$$

B_4 and $\vec{R}$ are to be estimated.

For the severity of the illness there are no sign restrictions and we use the simpler form,

$$\sigma_k = \tilde{\sigma}_{k\mathrm{MR}} + S_1\tilde{\sigma}_{k1} + S_2\tilde{\sigma}_{k2} + S_3\tilde{\sigma}_{k3}$$

$\vec{S}$ are to be estimated.

Provider level data (r_k, F_j, fc_j, V_m) The share for the traditional healer (r^*) is determined endogenously in the empirical estimation, and is a function of the illness condition only. The fixed costs of organizational providers are known. The base penalty for government and mission centers are determined in the estimation. F_{GC} is assumed to be 1, F_{GH}, F_{MC}, F_{MH} are estimated in the structural estimation. The reservation utility of the traditional healer, V_{TH} is estimated in the structural estimation, and fees charged at traditional healers can be inferred from this value.

Organizational Quality Factor We estimate ζ_k as follows:

$$\zeta_k = \frac{B_2}{1 + \exp\left(-1 \cdot \sum_{n=1}^{N} Z_n \cdot z_{kn}\right)}$$

z_k include β α and σ (as derived above), quadratic terms for each and interaction terms. The goal of this is to seek the least restrictive form for ζ that allows for decreasing marginal impact of illness characteristics. B_2 and $\vec{Z}$ are to be estimated.

The standard deviation of utility is B_4. We observe $\tilde{\beta}_k$, $\tilde{\alpha}_k$, $\tilde{\pi}_k$, $\tilde{\sigma}_k$, e_i, t_k, z_k, fc, dc, tc, and δ(the choice of provider). Using maximum likelihood we estimate the following parameters; B_1, B_2, B_3, B_4, F_{GH}, F_{MC}, F_{MH}, $\vec{E}$, $\vec{T}$, $\vec{Z}$, $\vec{R}$ and $\vec{S}$. r^* is endogenously determined from α and β.

The log likelihood is formed as follows.

$$\log L = \sum_{i=1}^{N} \sum_{j=1}^{6} \log \left(P_{ij}\right)^{\delta_{ij}} \qquad \text{where} \qquad P_{ij} = \frac{\exp(U_{ij})}{\sum_{l=1}^{6} \exp(U_{il})} \tag{32}$$

where $\delta_{ij} = 1$ when patient i chose provider j and 0 otherwise. Note that, in order to get a good estimate of the net utility of seeking health care we include those observations of individuals who did not seek care, making a total of 6 choices.

Using a Newton-Raphson algorithm we maximize the log–likelihood. The asymptotic covariance matrix of the estimated parameters $(\hat{B})$ is estimated from the variance matrix of the first derivative vector.

$$\left[\hat{I}\left(\hat{B}\right)\right]^{-1} = \left[\sum_i \hat{g}_i \hat{g}_i'\right]^{-1} \text{where} \qquad \hat{g}_i = \frac{\partial\,(l_i)}{\partial(\hat{B})} \tag{33}$$

Where l_i is the log likelihood for each observation. We approximate $\hat{g}_i$ by finite differences.

4.1 Results

The estimated coefficients, their standard errors and their corresponding z-tests are reported in Table 4 and 5. The coefficients in Table 4 do not have a direct interpretation. The endogenous variables they imply are of greater interest and these are reported in Table 6.

The coefficients represented in Table 5 have more direct meaning than the previous table. The individual characteristics of import are family wealth, the age of the patient, her education level, the gender of the care taker and whether or not the

Table 4: Structural Coefficients I: log likelihood = -853.07

variable		coeff	std err	z-test
$\omega(\rho^{\star}$-$\rho^{\emptyset})$ η scale factor	B_1	4.4	1.4	3.10
ζscale factor	B_2	6.8	2.9	2.38
weight on α and β	B_3	0.78	0.45	1.74
reservation utility	V_{TH}	4.4	1.3	3.50
standard error	B_4	5.3	0.62	8.60
α	$R_{\alpha,0}$	2.1	0.65	3.21
Coding 1	$R_{\alpha,1}$	0.054	0.02	2.66
Coding 2	$R_{\alpha,2}$	0.095	0.026	3.65
Coding 3	$R_{\alpha,3}$	-0.064	0.016	-4.13
β	$R_{\beta,0}$	0.0028	0.44	0.01
Coding 1	$R_{\beta,1}$	-0.068	0.015	-4.54
Coding 2	$R_{\beta,2}$	-0.021	0.013	-1.59
Coding 3	$R_{\beta,3}$	-0.088	0.024	-3.75
π	P_0	-6.7	3.5	-1.92
Coding 1	P_1	4	1.5	2.71
Coding 2	P_2	-1.3	0.46	-2.86
Coding 3	P_3	1.9	0.68	2.83
F	F_{GH}	2.7	0.53	5.09
	F_{MC}	1.82	0.27	6.74
	F_{MH}	3.6	0.85	4.24
σ				
Coding 1	S_1	0.13	0.057	2.35
Coding 2	S_2	0.2	0.062	3.21
Coding 3	S_3	-0.21	0.08	-2.67

care taker is a cocoa farmer or other farmer.

4.1.1 Endogenous Variables

There are two endogenous variables that we claim to be able to identify; the utility of the patient and the disutility of effort provided by the practitioner. The values of these two variables are reported in Table 6. The utility of no care are no reported since these are zero by definition. The column titled sample gives the average utility of the whole sample given patterns of choice. Two sets of average values are reported for each provider. The first is the average utility and effort exerted over the whole sample. The second reports the average weighted by the probability of visiting that

Table 5: Structural Coefficients II

variable		coeff	std err	z-test
$\omega \cdot \eta$	family wealth (ln)	0.098	0.042	2.36
patient	income	0.0016	0.0027	0.61
	age	0.024	0.013	1.78
	age^2	-0.02	0.015	-1.34
	female	-0.14	0.083	-1.63
	education	-0.026	0.012	-2.12
	infant (age $<$ 5)	-0.1	0.21	-0.49
	child (5 $<$ age $<$ 15)	-0.21	0.16	-1.31
care giver	age	0.0026	0.0028	0.94
	female	0.17	0.084	2.02
	education	0.014	0.01	1.36
	student	0.05	0.13	0.38
	cocoa farmer	-0.27	0.14	-1.93
	other farmer	-0.17	0.1	-1.65
	merchant	0.19	0.17	1.13
	other business	-0.37	0.34	-1.11
	njangi member	0.17	0.1	1.61
	meeting member	0.061	0.085	0.73
ζ	α	-4.2	1.9	-2.25
	β	-10	2.9	-3.62
	σ	16	3.9	4.10
	α^2	0.58	1.3	0.46
	β^2	7.5	2.6	2.89
	σ^2	19	4.8	4.02
	$\alpha\beta$	4.9	2.7	1.82
	$\alpha\sigma$	-12	4.4	-2.68
	$\beta\sigma$	-30	8	-3.76
	$\alpha\beta\sigma$	-7.4	2.6	-2.84
$\rho^{\star}$- $\rho^{\emptyset}$	could not work or restricted	-0.035	0.098	-0.36
	some work or difficulty working	-0.058	0.083	-0.70
	self-declared severity	0.35	0.11	3.19
	self-declared as not severe	0.29	0.18	1.59
	medically evaluated severity	-0.073	0.028	-2.65
	days sick (ln)	0.11	0.042	2.65
	days bedridden (ln)	0.11	0.04	2.79

provider.

Support for the argument that we can identify the disutility of medical effort. We can calculate the expected payments to traditional healers using the medical effort provided and the reservation utility. By these measures we estimate that the average visitor to a traditional healer expected to pay 7,200 CFA. In reality no patient will pay this amount; if cured they will pay more and if not cured they will pay less. However, over a large sample, the expected payment should be equal to the average observed payment. The observed average is 7,800 CFA. A priori we expected the expected payment to be larger than the observed average (since we do not believe that all payments had been made at the time the survey was administered) these variables are very close. Note that we did not use information of payments to traditional healers in the estimation, so this represents an independent confirmation.

Table 6 helps to illuminate the role of choice in the model. If patients were to visit the government clinics for all conditions they would have an average utility of 10,200 CFA. However, by bypassing clinics and seeking other providers, they achieve 13% greater utility, despite the greater additional costs. Note that the variation in the level of utility from visiting a traditional healer and mission hospital is particularly great. The variation on the part of the traditional healer comes from the variation in the level of effort provided by traditional healers. Variation at mission hospitals comes from the very travel costs, drug costs and fees required to visit a mission hospital. Few illness conditions merit this quality, but some benefit greatly from it. If patients chose to visit traditional healers for every condition they would do worse, on average, than if they visited another type of provider like a government clinic. But patients know when they should visit healers or mission hospitals and when they should not and this greatly increases the value they get from these options. Note that the value of visiting a government clinic falls when we account for selection, reflecting the fact that patients choose to visit government clinics.

The differences between providers in the level of effort provided come from the

Table 6: Endogenous Variables in the Estimation

variable	trad heal	gov cln	gov hos	mis cln	mis hos	sample
Utility						
average	2.8	10.2	10.8	11.1	4.3	12.2
maximum	147.8	76.5	135.9	106.1	150.7	148.8
weighted average	5.6	6.0	18.1	12.5	33.9	12.2
Effort						
average	0.625	4.61	12.4	8.37	16.4	
maximum	20.7	6.82	18.3	12.4	24.2	
weighted average	0.967	4.564	13.924	8.879	19.563	

All units 1,000 CFA (approx 2 USD)
weight is the probability of visiting that provider

different base penalties used at each institution. We set the penalty at government clinics to be one. Table 4 reports that the estimated penalty base at government hospitals is 2.7, that at mission clinics is 1.8, and that at mission hospitals is 3.6. This is a measure of the effectiveness of the incentives that are provided at these varying institutions. Note that, although traditional healers do not offer high levels of average utility or effort, even after weighting for visit patterns, the maximum levels of utility and effort are almost as high as that of mission hospitals. This reflects the fact that the outcome–contingent contract offered by traditional healers is better than regulation at adapting to the different needs of illness conditions.

Table 7 reports the values of α and β obtained in the estimation. The levels of α are higher than those of β. This does not imply that patient effort is more important in health care than medical effort since patient effort has no measurable units. The range of β is greater than the range of α implying that variations in the responsiveness of illness conditions to medical effort are of greater importance that variations in the responsiveness to patient effort. The levels of α and β imply that the optimal share varies between 25 and 31% of the value of health care.

In addition, the level of α and β at the maximum efforts for organizations and traditional healers are reported. Note that the maximum level of effort is different at each organizational provider, but effort exhibits the same relationship to α and β

at all of these providers. The maximum levels of effort at the traditional healer are, by construction, at the maximum joint levels of α and β. However, the maximum level of effort at organizational providers is at levels below these maximums. The comparative advantage of traditional healers comes when the responsiveness to both medical and patient effort are simultaneously high. In the reduced form analysis we found support for the hypothesis that the advantage of traditional healers comes when the responsiveness to both medical and patient effort is high. This is supported again in the structural estimation.

Table 7: Effort and the responsiveness to medical and patient effort

variable	min	mean	max	max effort	
				trad heal	organizations
α	0.6005	0.6312	0.6622	0.655	0.6622
β	0.1594	0.1953	0.2488	0.2245	0.2382

4.2 Full Information

The structural estimation allows us to estimate the disutility of effort, the relationship between effort and utility and the level of utility. Using the model that we derived we can estimate the level of utility that could be achieved in the absence of moral hazard. Table 8 reports the estimated utilities under full information. In this regime the patient compensates the practitioner for all disutility of effort and pays fees, drug costs and travel costs in addition. The gains are large. We report the gains for the full sample from visiting any given provider as well as the gains weighted by the probability of visiting that provider under the moral hazard regime. We do not calculate the probability of visiting any given provider under full information because the changes in patterns are almost absolute; patients no longer visit government hospitals, mission hospitals or mission clinics. Full information is not introduced as a suggestion of a possible regime, but rather because it allows some estimate of the costs imposed by moral hazard.

Table 8: Full Information Solution

variable	trad heal	gov cln	gov hos	mis cln	mis hos
average	32.7	71.8	24.6	67.4	14.2
weighted average	60.0	20.9	27.5	32.2	86.4
difference	29.8	61.6	13.8	56.3	9.9
weighted difference	54.4	14.9	9.4	19.7	52.6

All units 1,000 CFA (approx 2 USD)
weight is probability of visiting in base run not full information

Again it is instructive to note the role of choice in mitigating the costs of moral hazard. The difference in utility under moral hazard and full information falls when weighted by the pattern of visits for visits to government clinics, government hospitals and mission clinics, but rises at traditional healers and mission hospitals. The level of effort provided at institutional providers under moral hazard is based on penalties and is not highly responsive to the characteristics of the illness. Essentially the regulator has a crude tool with which to enforce quality and chooses to ensure relatively even levels of effort. Under moral hazard visitors to government clinics and hospitals and mission clinics suffered from illness conditions that were relatively less responsive to medical effort. Since effort levels are relatively constant, they were approaching full information levels of effort in this manner. Full information increases the level of effort they receive, but not by as much as it increases the levels of effort provided to the average patient. Taking into account selection reduces the estimates of increased utility.

On the other hand, visitors to mission hospitals and traditional healers tend to suffer from conditions that are highly responsive to medical effort. The additional effort provided under full information has a much larger benefit for them that for the average patient.

The traditional healer uses a different institution to provide medical effort, one in which the product of effort is shared by both participants. The outcome–contingent contract of the healer is a solution to problems of moral hazard, however it is an imperfect solution when both patient and practitioner provide unobservable medical effort. The large gains in utility when traditional healers operate under full information give some indication of the degree to which these contracts are an imperfect solution.

4.3 Government Policy and Incentives

Full information is an interesting bench mark but is obviously unobtainable. However, the estimation shows quite clearly that mission clinics and hospitals are superior in the provision of effort to their government counterparts. The major drawback of these providers, from the point of view of residents of our sample area is that travel costs to these providers is much higher than to government clinics or even hospitals. A government policy in which the incentives to provide effort were increased would not only increase the utility of visitors to those providers it would also offer those who would otherwise have traveled to mission facilities a way to save the expenses on travel.

Table 9 reports the utilities that would be derived if the gap between base penalties at government clinics and hospitals and mission clinics and hospitals were reduced by 90%. Importantly, the cost of any increase in effort over the original level provided is passed directly on to the patient. Thus, patient get higher levels of effort, but pay higher fees. We do this to insure that the policy can reasonably be considered self-financing. Table 9 reports the average and weighted average of utility under three different regimes. The first regime is the base case, the utility under the current regime. In the second case the government increase the incentives at both clinics and hospitals but patients continue to visit providers exactly as they did before. Note that only the utility at government clinics and hospitals changes under this regime and that the average utility of the sample does not change.

In the third regime the same change in quality is observed, but patients can adapt their patterns of visits, choosing to visit government clinics and hospitals more. Now the levels of utility for those visiting these providers increases significantly and the

Table 9: Incentive Based Policy

variable	trad heal	gov cln	gov hos	mis cln	mis hos	sample
current policy						
average	2.8	10.2	10.8	11.1	4.3	12.2
weighted average	5.6	6.0	18.1	12.5	33.9	12.2
increase incentives						
fixed pattern						
average	2.8	11.7	10.7	11.1	4.3	12.2
weighted average	5.6	5.8	18.8	12.5	33.9	12.2
variable pattern						
average	2.8	11.7	10.7	11.1	4.3	12.7
weighted average	4.0	8.8	28.8	9.7	19.8	12.7

All units 1,000 CFA (approx 2 USD)

utility of the sample as a whole increases as well. Again this suggest the power of selection on the patients part.

5 Conclusion

Moral hazard is a significant cost in health care. The unique element of a traditional healer combined with informed choices on the part of patients has allowed us to estimate its costs as well as the benefits offered by different contractual solutions and coping mechanisms. The behavior of patients is very important in reducing the costs of asymmetric information. In addition traditional healers are offering a contract form that, despite its drawbacks, offers great benefit to patients under given circumstances. The results of this investigation offer strong support to the author's previous work on traditional healers which suggests that their continued success depends more on the contracts they offer than the medicine they deliver.

In addition the conventional wisdom that mission health services provide higher quality care is supported by this work. Despite their greater costs and significant additional travel costs to patients in this region mission clinics and hospitals are frequently visited and it appears that patients benefit greatly by being able to choose these facilities when they feel it is appropriate.

By comparing mission facilities to government facilities the structural estimation allows us to examine the potential impact of policies that increase the incentives that practitioners have to exert unobservable effort. We find that patients benefit greatly from such increased incentives even if they have to compensate practitioners directly for the additional effort.

References

Baerts, Martine, *Guerisseurs et plantes medicinales de la region des cretes Zaire–Nil au Burundi* number 18. In 'Annales Science Economiques.', Tervuren, Belgie: Koninklijk Museum voor Midden–Afrika, 1989.

Conco, W. Z., "The African Bantu practice of medicine: Some preliminary observations," *Social Science and Medicine*, 1972, *6* (3), 283–322.

Edwards, S.D, *Indigenous African healers and western trained psychologists: a comparative study of interview methods* number 38. In 'Publication series of University of Zululand, series B.', Kwa Dlangezwa, South Africa: University of Zululand, 1983.

Eswaran, Mukeshi and Ashok Kotwal, "A Theory of Contractual Structure in Agriculture," *American Economic Review*, 1985, *75*, 352–367.

Gelfand, M. et al., *The Traditional Medical Practitioner in Zimbabwe: His principles of practice and pharmacopoeia*, Harare, Zimbabwe: Mambo Press, 1985.

Griffith, H. Winter, *The Complete Guide to Symptoms, Illness and Surgery*, Los Angeles, CA: The Body Press, 1985.

Grossman, Michael, "On the Concept of Health Capital and the Demand for Health," *Journal of Political Economy*, 1975, *80*, 223–255.

Grossman, Sanford and Oliver Hart, "An Analysis of the Principal Agent Problem," *Econometrica*, 1983, *51* (1), 7–45.

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.

Korse, Piet et al., *Sorcellerie, maladie et chasseur de sorciers chez les Mongo (Rep. du Zaire)* number 103. In 'CEEBA, series 11.', Bandundu, Zaire: Publications (CEEBA), 1989.

Lasker, Judith N., "Choosing among therapies: illness behavior in the Ivory Coast," *Social Science and Medicine*, 1981, *15A* (2), 157–68.

Leonard, David K., ed., *Africa's Changing Markets for Human and Animal Health Services*, London: Macmillan, 2000.

Leonard, Kenneth L., "African Traditional Healers and Outcome–Contingent Contracts in Health Care," Discussion Paper Series 9900-02, Columbia University 2000.

– **and Joshua Zivin**, "How to Compensate Physicians when both patient and physician effort are unobservable," Discussion Paper Series 9900-01, Columbia University 2000.

Mliga, G.R., "Decentralization and the Quality of Health Care." In Leonard, ed (2000) chapter 8.

Oyenye, O.Y. and I.O. Orubuloye, *Some aspects of Traditional Medicine in Bendel State of Nigeria: An exploratory study* number 14. In 'Niser monograph series.', Ibadan, Nigeria: Nigerian Institute of Social and Economics Research, 1985.

Staugård, Frants, *Traditional Medicine in Botswana*, Vol. 1, Gaborone, Botswana: Ipelegeng Publishers, 1985.

Strickland, G. Thomas, ed., *Hunters' Tropical Medicine*, 6 ed., Philadelphia: W.B. Saunders Co., 1984.

Werner, David, *Where There is No Doctor; A village health care handbook*, Palo Alto, CA: The Hesperian Foundation, 1977.

A Drug Price Determination

In order to estimate the cost of drugs at each center we hypothesize that the drug cost is a function of characteristics of the disease (as well as the age and gender of the patient) and that costs are related across centers by a fixed ratio.

$$dc_{jk} = \Delta_k \cdot dc_{GC,k} = \Delta_k \cdot \prod_n z_{kn}^{\beta_n} \cdot \prod_l e_{il}^{\alpha_l}$$

$dc_{j,k}$ is the cost of drugs at practitioner j for illness k, and $dc_{GC,k}$ is the cost at government clinics for illness k, and serves as our numeraire drug cost for each illness condition. The cost at each center is related to the numeraire cost by a ratio Δ_k, ($\Delta_{GC} = 1$) and the numeraire cost is a function of individual characteristics (e_i) and disease characteristics (z_k).

Table 10: Regression of Log of Real Cost

variable	Coef.	Std. Err.	t
constant	.583764	.1424384	4.098
nocare	-.8930396	.1550042	-5.761
govhs	1.025169	.1129373	9.077
miscl	.3464014	.1072761	3.229
mishs	1.314765	.1630441	8.064
s1	.7730967	.1254969	6.160
s2	.4119151	.1095111	3.761
s3	.4298704	.1853261	2.320
infant	-.2680316	.0928182	-2.888
ldays	.1060058	.038108	2.782
bedridd	-.0013809	.0004632	-2.981
lbed	.1489192	.0446116	3.338
rnose	-.6621452	.30628	-2.162
fever	.2353869	.0945969	2.488
fracture	.4652542	.2840258	1.638
abdpain	.3602139	.2015079	1.788
extswel	-.3501687	.2985265	-1.173
dpcut	.5924908	.2590559	2.287
int0203	-.4875229	.1396483	-3.491
int0218	-.348287	.1901768	-1.831
int0708	.716706	.3957439	1.811
int0324	.9085881	.409071	2.221
int0207	-.3870165	.3162465	-1.224