

# Active patients in rural African health care: implications for research and policy

Kenneth L Leonard\*

Agricultural Economics, 2200 Symons Hall, University of Maryland, College Park, MD 20742, USA

\*Corresponding author. Agricultural Economics, 2200 Symons Hall, University of Maryland, College Park, MD 20742, USA.  
E-mail: kleonard@arec.umd.edu

---

**Accepted** 14 December 2012

We introduce the ‘active patient’ model, which we claim is a better way to describe health-seeking behaviour in low-income countries. Active patients do not automatically seek health care at the closest or lowest cost provider, but rather seek high-quality care (even at higher cost) when they estimate that such care will significantly improve outcomes. We show how the active patient can improve his or her health even when access to adequate quality care is insufficient and that the empirical literature supports this model, particularly in Africa. Finally, we demonstrate the importance, in analysing health care policy, of recognizing patients’ efforts to improve health outcomes by seeking quality care.

**Keywords** Agency in health care, rational peasant, rural health care, Africa, asymmetric information

---

## KEY MESSAGES

- Households in low-income countries are active consumers of health products over which they have little control.
- The active patient model is informed by empirical evidence on health-seeking behaviour.
- Policy makers should recognize the activities of patients in seeking health care and design programmes that take advantage of these efforts.

## Introduction

Healthcare policies in low-income countries are primarily focused on improving technology, expanding access and increasing the quality of health service delivery. This focus on the current shortcomings in the supply of health care means that less attention is paid to patient demand and health-seeking behaviour. Ignoring patients’ expectations and behaviour, however, is equivalent to assuming a particular model of patient behaviour, what we call the ‘passive patient’ model. The passive patient is assumed to seek health care at the nearest primary healthcare facility when sick and to follow the advice received. When patients do not visit the nearest facility, it is assumed that high costs are preventing them from seeking care. We argue that this view of patient behaviour, whether implicit

or explicit, does not fit the well-documented empirical reality. Research on health-seeking behaviour has made great strides in modelling and understanding household behaviour. Households and patients are generally not passive; they choose healthcare providers from among a wide variety of possible sources, and they collect and use information on these choices that improves their healthcare outcomes. Importantly, by recognizing the significant effort patients exert in seeking proper health care (rather than going only to the closest or least expensive facility), policy makers can substantially improve the impact of supply-driven policies.

This article outlines an empirically based view of patients, drawn from the literature on health-seeking behaviour in low-income countries, as active consumers of products over which they have little control. ‘Active patients’ do not have the

power of voice; they cannot directly improve the quality of services available to them in the current healthcare services market. However, they do have the power of exit; no force compels them to seek health care or to seek it at the 'accessible' provider, at the free or subsidized provider, or at any kind of provider that states or international agencies would expect them to visit. Although patients cannot force any healthcare practitioner to provide better care, they can improve their health by actively choosing to visit better providers.

Four assertions about health care and health-seeking behaviour drive the active patient model. First, patients care about health, not health care, and patients are only willing to pay for health care when they believe it will help them return to health. Thus, when deciding to seek care, patients compare the cost to their potential long-term gain from returning to healthiness to their current level of income or savings. Second, each illness has a unique production function for healthiness, and treating it requires a different combination of healthcare inputs, such as skill, equipment and medicine. In other words, some illnesses may require little in the way of skill to treat, whereas others require significant skill. Third, markets for health care suffer from asymmetric information, where purchasers are unable to assess the value of goods sold and it is, therefore, difficult to directly purchase the things patients need and want. Because patients cannot evaluate the quality or quantity of care they receive, they cannot directly purchase the inputs they value. Fourth, in all developing countries, patients are seeking this valuable input in an environment in which regulatory and institutional guarantees of quality differ greatly from organization to organization; quality care is not guaranteed, but it is not universally low either.

Thus, the active patient can choose among a range of healthcare practitioners providing varying qualities of care. By considering what may be needed to treat the specific illness from which he or she suffers, the active patient can choose providers in a way that is superior to visiting the nearest provider or even the 'best' provider in general terms though perhaps not for the patient's particular illness.

In important aspects, this model is similar to and draws from some of the sophisticated empirical models of health-seeking behaviour in developing countries (see, e.g. Dor *et al.* 1987; Gertler *et al.* 1987; Gertler and van der Gaag 1990; Bolduc *et al.* 1996). These papers also see patients as actively choosing among a range of healthcare providers. They develop a model of health-seeking behaviour that we refer to as the 'perfect patient' model, in which patient choices collectively drive the healthcare market towards efficient supply. In contrast, the active patient model reaches different conclusions about the responsiveness of supply by focusing on the importance of asymmetric information and institutional features of healthcare providers.

In the following section, we outline our view of the active patient, differentiating its implications from those of both the passive and perfect patient models. In 'Evidence of active patients' section, we introduce the empirical evidence in support of the active patient model. 'Implications for healthcare policy' section discusses the policy implications of this model and 'Conclusions' section offers conclusions.

## The active patient

The active patient model builds on three observations about health seeking in low-income countries: patients are usually passive when at the health facility; they expend significant energy on seeking care beyond their local facility; and they often fail to secure adequate quality care despite this energy. From these observations, we build a model in which (1) patients value health not health care, (2) different illnesses require different levels and types of quality treatment, (3) patients cannot observe or evaluate many key elements of health care and (4) regulatory and institutional guarantees of quality are inadequate.

Patients are often passive in the presence of medical personnel [see, e.g. Kabakian-Khasholian *et al.* (2000)]. This passivity makes sense when patients do not know whether the medical provider's diagnosis and procedures are the best choices. Even if they know what should be done, it is not clear what action they could undertake to improve anything. The passive patient model assumes that this passivity extends to all aspects of seeking care. The active model, on the other hand, assumes that patients are not passive when they seek care. Thus, the active model focuses patients' behaviour when they fall sick rather than when they arrive at a health facility.

For a person who has fallen sick, the goal is to maximize the probability of a good outcome (usually recovery) through treatment. The effectiveness of treatment is the degree to which it increases the probability of a good outcome when compared with no treatment. The value of treatment is a combination of effectiveness and the value of recovery.<sup>1</sup> Patients value treatment because the probability of being cured is greater with treatment than it is without treatment, even though it is possible to not get better with good treatment and to get better without treatment.

As treatment increases the probability of a good outcome, the value of health care is based on the value of the good outcome. The same healthcare service can be worth different amounts depending on what illness the patient thinks they are suffering from. If the cost is the same, the value may exceed the cost for some illnesses but not for others. Furthermore, as the value depends on the effectiveness of the treatment provided (quality) patients would be more willing to pay for high-quality care than for low-quality care. Thus, the affordability of health care (value compared to cost) is a function of the quality of care provided. Inexpensive low-quality care can be unaffordable, whereas expensive and high-quality care is affordable.

In addition, active patients seek health care while understanding that quality can mean different things for different illnesses. Each illness requires different levels and combinations of health inputs. Treatment is the combination of a series of inputs, such as diagnostic effort, medical capacity, medicines and equipment. These inputs increase the probability that a patient will experience the best possible outcome; however, the role of each input varies according to the illness condition. Therefore, two patients with different illnesses may seek different levels and combinations of inputs. More generally, the same patient may seek all high-quality inputs for some illnesses and low-quality inputs for other illnesses. Thus, both the relative value of treatment as a whole and of individual inputs depends on the illnesses from which patients suffer.

Our third assertion focuses on the fact that patients cannot directly purchase many of the inputs even when they know they are valuable. If patients could purchase each input at an individual cost, then two patients who needed different levels of inputs could visit the same provider and ask for only the inputs they needed. However, health care is not an ordinary good and patients cannot observe or evaluate the levels and quality of many types of inputs. It matters whether the input is observable or not. Observable inputs are those that patients should be able to assess: the presence of staff, availability of medicines and politeness of care providers are features that can be observed and assessed by patients. Unobservable inputs are those that patients cannot easily see or assess. Diagnostic quality and skill, in particular, are almost impossible for patients to evaluate; even when they observe the activities of the clinician, they do not have the medical training to know if the right things were done. Note that observable and unobservable are not the same thing as important or not important. Quality nursing care, e.g. is important and mostly observable. Thus, services such as maternity care, which require frequent interaction with doctors and nurses, have many observable inputs. In addition, patients may have traditionally treated medicines as observable; you know if you received a medicine and even if it is branded or generic. However, counterfeit medicines are common now, and patients may begin considering medicines as unobservable inputs.

Importantly, the stochastic nature of health outcomes means that patients cannot always accurately infer the level of treatment by looking at the outcome: some patients are cured despite poor diagnostic care or counterfeit drugs, whereas some are not cured despite the best diagnostic care or branded drugs. Thus, not only is it difficult for patients to assess the inputs provided at the time of service, they cannot properly assess them after the fact.

Our fourth assertion in the active patient model is that regulation and organizational management do not ensure that patients get what they need for each illness. The level of inputs provided to individual patients is a decision made at the facility or organization level without asking the patient what she wants or values. Some providers, such as small private organizations, are more likely to be responsive to the collective demands of patients, offering the services demanded by most patients. However, most large organizations—including those in the public sector in particular—offer healthcare inputs driven by their own assessment of what is required and are not responsive to patient demand.

To aid our conversation about the difference between what the patient does and does not know about the levels of healthcare inputs, we introduce a few terms:

- The ‘input elasticity of health gain’ is the change in the probability of a good outcome from a 1% change in the supply of a particular input. This reflects the fact that quality will be more important for some illnesses than for other illnesses. The input elasticity of medicine is high for pneumonia, e.g., whereas the input elasticity of diagnostic skill is high for stomach pain (Leonard *et al.* 2002).
- The ‘optimal input quantity’ is the optimal level of a particular input for a particular illness and patient.

- The ‘actual input quantity’ is the amount of each input provided by a particular provider for each illness.
- The ‘estimated input quantity’ is the patient’s estimate of the input levels provided by a particular provider for a particular illness.

Active patients pursue the ‘optimal input quantity’ but for many inputs, they cannot know what level of inputs they received. The ‘actual input quantity’ is not necessarily equal to the ‘optimal input quantity’ because regulation is not perfect, and the ‘estimated input quantity’ is not necessarily equal to the ‘actual input quantity’ because patients cannot easily assess the inputs they receive. The purest form of the passive patient model assumes that the actual input quantity is equal to the optimal input quantity, based on the assumption that doctors do what is best for their patients. In contrast, the perfect patient model assumes that all three levels are the same (optimal equal to estimated equal to actual). The equality in this model is not driven by the assumption of perfect markets but by the dual assumptions of well-informed patients and perfect markets. These well-informed patients can drive bad doctors out of the market and force those who remain to provide exactly what their patients demand (which is also what they need).

The active patient model asserts that patients have little control over the inputs that are provided by any practitioner and many practitioners do not provide the optimal levels of inputs. Unlike perfect patients, active patients cannot negotiate for higher levels of unobservable inputs. This leaves patients with two strategies for improving their health. First, they can try to get better care from individual healthcare providers. In the case of observable inputs, this can be straightforward: payment for services. In the case of unobservable services, patients and healthcare providers can use implicit or explicit contracts to ameliorate the problems with unobservable inputs. The second strategy available to patients for improving their health is to use their understanding of contracts and other available evidence on quality to estimate the levels of inputs provided by each practitioner and facility. Using these estimates, they can then seek the best available quality for a particular illness.

## Evidence of active patients

In this section, we examine (1) the way that patients use implicit contracts to improve the quality of care, (2) evidence that patients consider quality in relation to their particular illness when choosing providers and (3) evidence of how households learn about the quality of care that is provided at facilities from which they can choose.

### Contracts for the provision of inputs and outcomes

There is significant evidence that healthcare providers in many low-income countries do not provide optimal levels of inputs (Das *et al.* 2008). Particular attention is being paid to the ‘know–do gap’—the gap between what health workers know how to do and what they actually do for their patients (Rowe *et al.* 2005; WHO 2005; Das and Hammer 2007; Maestad and Torsvik 2008; Maestad *et al.* 2010). This gap

is significant: de Savigny *et al.* (2004) report that 78% of children in Tanzania who died of malaria sought care from a modern healthcare provider before they died. This high percentage indicates that benevolent and skilled doctors are not common. Thus, regulators do not assure quality and patients must seek other ways to assure quality care.

When inputs are observable, patients can buy them directly from practitioners. For example, patients should be able to buy the presence of a doctor, access to medicines and even certain procedures. Even when the organization that employs the practitioner sets the level of inputs, patients are often able to use informal payments (bribes) to get what they want (Lewis 2007). However, because patients cannot buy unobservable inputs directly, they may pay a fee for the unobservable input without knowing the input's quality or even if the doctor actually provided it. Most bribes and informal payments, therefore, buy access but not quality. Lindkvist (2012) finds that clinicians who accepted bribes did not provide higher overall effort. Most settings in developing countries give many opportunities to pay bribes and many health workers are willing to take them, yet these conditions are not associated with high quality.

On the other hand, the theoretical literature suggests that contracts can work to help ensure the quality of unobservable inputs where direct payment fails. We examine three types of contracts: the pay-only-if-cured contract, the hierarchical supervision contract and the reputation contract. The pay-only-if-cured contract has long been recognized as an ideal in health care (Arrow 1963; Dranove and White 1987). If patients paid their doctors only if they are cured, or paid more if they are cured, the economic cost of unobservable inputs would be reduced. This 'ideal' contract is exactly the method by which traditional healers in Africa do business. Patients in the Southwest Province of Cameroun pay traditional healers more than twice as much if they are cured (Leonard 2003) than they pay if they are not cured. Healers receive an initial payment and negotiate with the patient over a payment to be made in the future. In all the cases studied, if the treatment did not result in improvement of the condition, the patient paid nothing beyond the initial payment. This system does not guarantee that the patient is cured, nor does it guarantee that the healer is only paid if he does the right thing. Sometimes healers are lucky and the patient recovers despite getting poor care; sometimes they are unlikely and the patient is not cured despite getting high-quality care. However, the healer is more likely to be paid if he provides high-quality care, which gives healers strong incentives to provide quality care even if the patient cannot evaluate or observe this quality.

At modern providers, patients pay a fee for services and the fee does not depend on the outcome; the fee does not encourage high-quality care. However, most healthcare providers use implicit contracts with their patients common to many goods and services provided in developed as well as developing settings.

One example of an implicit contract is the hierarchical supervision contract (Leonard 2002), whereby the patient pays the institution for services, and the institution ensures that high-quality care is provided by paying the practitioner for quality not for services provided or outcomes. If patients

experience a bad outcome, they can punish the institution by refusing to return (or even refusing to pay), but this cost is born by the institution, not the physician. Thus, in theory, the patient pays the institution for good outcomes and the institution pays the physician for high-quality inputs. This type of contract requires that organizations (supervisors and managers) can observe quality when patients cannot. However, because the link between what the patient pays and what the physician receives is broken, this means that when high-quality care is provided but a patient is not cured, the patient can punish the institution at the same time that the institution rewards the physician. The necessity of a third party means that this contract is only possible in hierarchal systems (Hölmstrom 1982; Ellis and McGuire 1990).

Small private providers in the modern sector cannot use either the pay-only-if-cured contract or the hierarchical supervision contract because contracts are only useful if both parties believe they are enforceable. Pay-only-if-cured contracts only work when patients are willing to return to the provider after being cured and honestly report the outcome—patients will do this for traditional healers either because they live in the same community (where outcomes are visible to everyone) or because they believe false reporting will cause the disease to return. In the past, doctors could use credit to achieve much the same result in small communities where outcomes were clearly visible. If patients were not cured, practitioners knew collecting on the debt would be difficult, creating a link between outcomes and payment (Starr 1982). The hierarchical supervision contract, on the other hand, only works if there is a third party involved. But for most small private providers, the only possible contract for unobservable services is the reputation contract on the other hand, whereby the provider loses or gains future business with each outcome experienced. This is different from the pay-only-if-cured contract because the private providers only earn the benefit of a good outcome from future business, whereas a traditional healer earns the benefit in an immediate (and often large) payment. It is different from the hierarchical supervision contract because the private provider is not paid a salary for high quality, whatever the outcome. As the patient's outcome experience is not the same thing as the quality of care provided, the success of the reputation contract depends critically on how households infer quality as they experience outcomes over time.

The reputation contract is also described as trust (Gilson 2003, 2005; Bloom *et al.* 2008). When households are not sure whether a practitioner provides high-quality care, they are often unwilling to take significant risks and may react strongly to bad outcomes. On the other hand, once households trust a provider, they are willing to pay higher fees and accept the occasional bad outcome, in essence rewarding initial high-quality care with later business.

The contract available at traditional healers shows that indigenous institutions clearly recognize the problems facing active patients. For the other contracts, it is not their existence but the way patients use them that support the view of active patients. Leonard (2009) examines household health-seeking behaviour in the face of the contracts available to them and shows that households' choices reflect an understanding of the value of these contracts. The fact that solutions to market

failures in the delivery of health care exist in indigenous, traditional institutions are evidence of both an understanding of these market failures on the part of traditional healers, and a willingness to seek these solutions on the part of patients.

### Selection of providers

The health-seeking literature clearly demonstrates that patients actively make choices in their search for quality care. Particularly informative is the phenomenon of bypassing, whereby patients go past one or more healthcare facility to visit another facility. Many studies show that these choices are guided by the character or severity of the illness (Mwabu 1986, 1989a,b; Mwabu and Mwangi 1986; Sauerborn *et al.* 1989; Mwabu *et al.* 1993; Leonard, 2003). Studies of bypassing that implicitly test or support the active patient model include some element of quality available at two or more facilities from which patients can choose (Akin and Hutchinson 1999; Leonard *et al.* 2002; Sahn *et al.* 2003; Hanson *et al.* 2004; Lindelow 2004; Grobler and Stuart 2007; Kruk *et al.* 2009; Gauthier and Wane 2011). These studies overwhelmingly document a willingness to pay significant additional costs to visit higher-quality facilities.

Since patients should be able to buy observable inputs without having to travel additional distances, we expect patients to be particularly active in the search for quality when they need unobservable inputs. To this end, Hanson *et al.* (2005) show (in a discrete-choice experiment) that patients value thoroughness of evaluation, staff attitudes and drug availability, in that order. Note that this order is the reverse order of these inputs' observability. Mariko (2003) reports that estimates of the willingness to pay for quality are significantly understated when the model only considers structural quality (availability of medicines and equipment). Patients are seeking and paying for process quality even though structural quality should be easier to observe. This is because observable inputs are easier to find in more places, and unobservable inputs are therefore more valuable. For example, if patients need medicine and diagnostic quality, they are more likely to go where they expect to find diagnostic quality because they can find medicine more easily elsewhere.

In the active patient model, the value of unobservable inputs also varies with the illness and patients with different illnesses will have greater or lesser degrees of preference for particular unobservable inputs. Leonard *et al.* (2002) show that the willingness to travel extra distance to obtain a higher level of a particular input varies by the illness condition, suggesting not only that patients know where to seek inputs but also when certain inputs may be more or less valuable to them.

### Learning about the quality available at facilities

How do patients assess the various levels of inputs available at different facilities given that many of the inputs they seek are unobservable? Leonard (2007) examines the patterns of bypassing in relation to the average tenure of clinicians in each facility and shows that household willingness to seek certain inputs at a facility increases when the clinician at that facility has been present for longer. In essence, a household's estimate of the level of inputs increases as a clinician has been present for a longer time, even when the actual level of inputs has not

changed. This pattern shows that households are learning about the levels of inputs over time.

Leonard (2011) uses data from medical professionals who estimate 'the input elasticity of health gain' for illnesses reported by households and examines the average input elasticity at each facility over time. The proportion of cases that are self-limiting (do not require any inputs) does not change significantly over time. The proportion of illnesses that are urgent (requiring immediate attention) increases somewhat, but only if the clinician at the facility is above average in quality. Importantly, the proportion of cases that respond well to effort changes significantly over time. In particular, for clinicians who have above average diagnostic skills, the proportion of illnesses that respond well to quality inputs increases steadily for the first 4 years the clinician is present at the facility. This pattern means that, as patients learn that a particular clinician is good, they are more willing to visit that facility with illnesses that require the input they have learned this clinician can actually provide.

This evidence is consistent with patients who observe outcomes from their own health episodes and those who observe the outcomes of other households around them and then infer the levels of inputs. Clearly, households cannot immediately learn about the levels of unobservable inputs provided by health facilities but active patients can slowly assess unobservable inputs provided by facilities over time and use this information to seek high-level inputs when they have a high probability of being valuable for their particular illness. Leonard (2007) shows that patients respond almost immediately to improvements in inputs that should be observable (the number of malaria prescriptions that involve an injection, e.g.) and that it can take up to 4 years before patients fully respond to increases in unobservable inputs (consultation and prescription quality). The response of patients to observable inputs could be seen as patients responding to changes they can observe, or to patients demanding what they value. However, the second explanation does not explain the fact that we also see patients avoiding facilities with low levels of unobservable inputs. Why would they avoid such a facility if they could demand what they value? The active patient model suggests they cannot demand what they value (even if it is unobservable), they can only seek it when they know that it is present.

### Implications for healthcare policy

Active patients are able to improve health outcomes by seeking practitioners with implicit contracts for the provision of high-quality unobservable inputs, bypassing facilities with low levels of quality when the patients estimate that higher quality is valuable for a particular illness. To do this, they gather information about the quality of care that is provided at existing healthcare providers. When households cannot know the quality of unobservable inputs at facilities and implicit contracts do not guarantee quality, households can learn about quality, but the process takes time. This delay in patient learning means that practitioners who use reputational contracts or organizations that seek to improve the quality of unobservable inputs by exercising their regulatory authority will not see immediate benefits. This model of health-seeking

behaviour has implications for policy. In this section, we look first at the general set of policies currently in place to increase access to quality care and then at three specific examples (social accountability, private health care and performance pay) of policies.

### Access to quality health care

When patients get adequate health care at lower costs (either in fees or travel costs), this demonstrates an increase in access to health care. Traditionally, policies that improve access focus on building new facilities that may be nearer to patients, upgrading existing facilities (increasing either equipment or staff qualifications) or lowering fees. In the passive patient model, improvements in access lead to improvements in health outcomes because patients are more likely to visit their local facility when it is closer or less expensive. Sometimes better outcomes occur simply because the patients happen to get better care at the facility they had already chosen to visit. However, once we understand that patients are active and that changes in access can come from a wide variety of policies, we can see that the potential benefits of changes in access are quite variable.

Active patients bypass inadequate facilities when they need quality, so if there are improvements in the facilities that are closer to them, they might gain from reduced travel costs even if there are no changes in health outcomes; they were getting high-quality care by travelling and now they do not need to travel to get the same quality care. In addition, the active patient might gain when quality is improved at facilities that are far away, even if this now means they will travel long distances to get to the higher-quality facility. Thus, active patients have a different perspective of 'access' that makes it useful to replace the idea of a 'local service provider' with that of a 'local care portfolio'. For the active patient, the key variable is the cost to obtain quality care when it is necessary. This cost includes fees, but it also includes the travel cost to find a high-quality provider. In rural areas, travel costs are often much larger than any other costs.

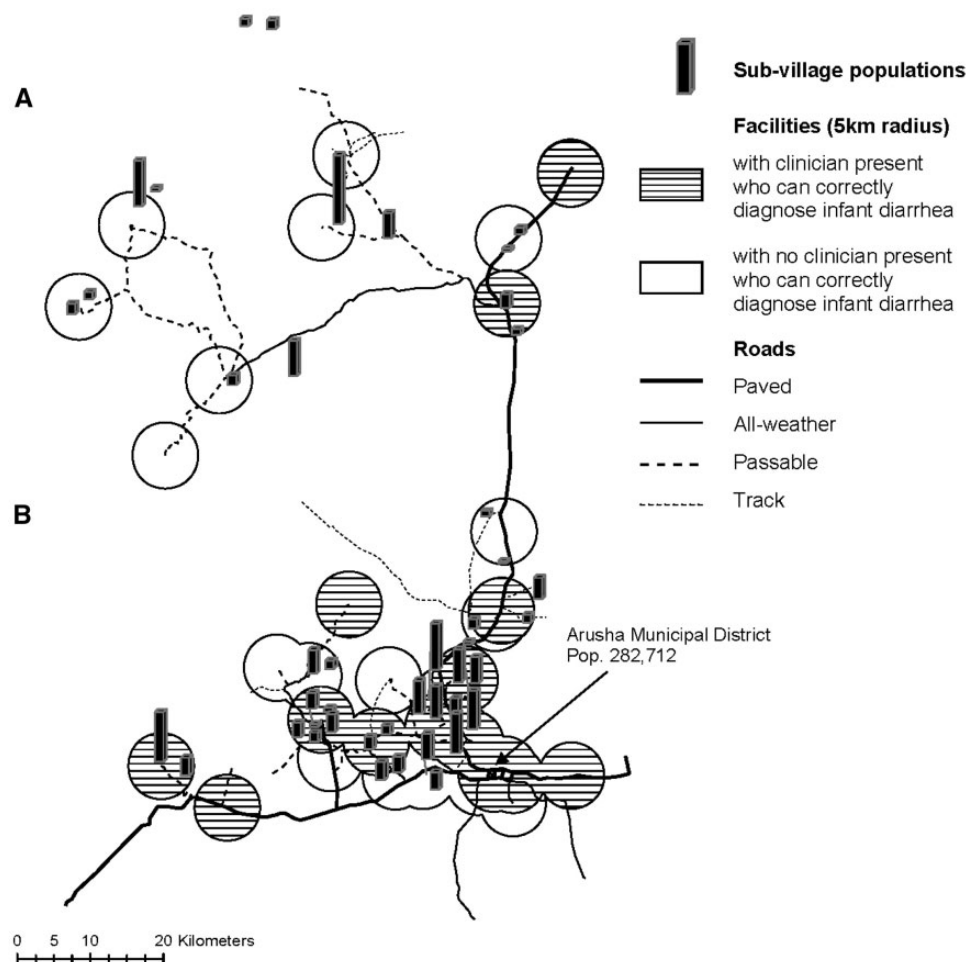
Figure 1 illustrates the active patient view of access in data from the Arusha region in Tanzania. This map shows the location and population of sub-villages sampled in a household survey, all health facilities, roads and road conditions and one measure of clinician quality. The 5-km radius around each health facility shows the standard view of access, based on the passive patient model, which assumes that if a household lives within such a circle, it has adequate access. According to this standard view, most rural households in this region have reasonable access to facilities. However, as the figure indicates, not all clinicians are competent. For example, the shaded circles indicate facilities with at least one staff member competent to diagnose the causes of infant diarrhoea. Viewed in this light, the average rural household in the region clearly has poor access to adequate care. If patients always visit the nearest facility, then households in locations A and B face a similar, unacceptably low level of access. On the other hand, if patients know where to find quality and travel to reach acceptable facilities, then households in location B have significantly better access than households in location A. Location B has the same poor local service provider, but a better local care portfolio.

From a policy perspective, it is not necessary to improve the quality of the nearest local care provider to improve the quality of the local care portfolio in the active patient model. Instead, the active patient would view improving any nearby facilities or reducing travel costs (by improving roads) as significant improvements in access. Where density of population and facilities is greater, access to quality care is seen as significantly higher in this model, even if most of the facilities are of low quality, because patients can avoid the low-quality facilities and get to the higher-quality ones when they need better care.

Klemick *et al.* (2009) further differentiate the passive and active patient views by comparing the predicted outcomes for three different policy scenarios under two sets of assumptions. In the first, patients always visit the nearest healthcare facility, whereas in the second, patients make tradeoffs in quality and distance for each illness, choosing the best facility in their portfolio. The policies considered include improving capacity by upgrading the staffing levels, upgrading roads to reduce travel costs and improving performance of existing health workers without improving capacity. Improving performance, in this setting, makes quality more homogeneous because under normal circumstances quality varies significantly across facilities. Improving capacity raises the performance at all facilities while maintaining the variability. If patients always visit the nearest facility, the gain from upgrading the staffing levels is large, whereas the gains from improving roads and performance are smaller. However, once we take into account the fact that patients actively seek quality by travelling to other facilities the implications are different. In this case, improving capacity by adding staff at all facilities is less valuable to patients because they were already seeking high-performing facilities when it mattered. On the other hand, the gains from improving roads and healthcare performance are much larger for active patients because (1) bypassing poor facilities is less expensive on better roads and (2) there are more high-quality facilities that do not need to be bypassed.

From the point of view of an organization that provides health services, active patients allow for policy to provide a double benefit. First, patients visiting a facility that increases quality will benefit, and second, new patients will be attracted to this facility. Patients who had previously incurred significant travel cost to avoid the facility by visiting facilities owned by other organizations will benefit from increased quality and decreased avoidance costs. In addition, when health facilities collect (and retain) fees or health workers earn bonuses that are tied to the number of patients at the facility, increases in quality will increase the earnings of the staff at the facility. Mariko (2003) estimates that authorities in Burkina Faso could double user fees and still see an increase in utilization if the fees were directed at quality improvement. Litvack and Bodart (1993) document such a project in Cameroun where they found that everyone (including the poor) increased their use of the improved facilities.

On the other hand, it is also possible to improve access by lowering fees at some facilities. Such policies are generally considered for public facilities, meaning that fees are reduced at a large set of facilities but remain unchanged at facilities run by other organizations. Clearly, demand for lower-cost services increases when such a policy is enacted, but not all the new



**Figure 1** Access to health facilities in Arusha, Monduli and Arumeru districts, Tanzania.

clients are people who otherwise would not have obtained care. Many of the new patients will have previously travelled to other facilities before the fees were reduced; although they have switched providers, they have not gained new services (Sahn *et al.* 2003).

Unless the lower fees are accompanied by increased quality (or at least stable quality), the case mix of the impacted facilities will be significantly altered. Unfortunately, evidence shows that ignoring quality concerns while lowering fees will lead to lower quality because more patients seek the same, limited resources (Meessen *et al.* 2011). Note that people may still visit facilities even if the quality of care decreases; however, only patients with illnesses that do not require quality will visit. In that case, patients needing quality care who used to visit the facility and pay higher fees will now avoid that facility and seek care elsewhere (at unreformed facilities). Thus, it is possible that more people will visit a facility even while quality falls. At the same time, people who must get quality care will go elsewhere. Indeed, Meessen *et al.* (2011) report that in some countries, the elimination of user fees at public facilities has led to an increase in expenditures on private healthcare services.

To the degree that policy makers focus on caseload as a measure of progress at facilities where fees have been lowered,

they will fail to understand the true outcome of their policies. At the very least, policy makers must monitor changes in case mix and caseload. Ideally, they should monitor quality at reformed facilities as well as caseload and case mix at other facilities in the local area.

### Social accountability

Social accountability is an attempt to give local communities more power over the services provided in their area. It is particularly useful where political accountability fails (World Bank 2003). Proponents of social accountability would find at least two aspects of the active patient model familiar: first, patients have little voice in their face-to-face encounters with providers, and second, patients can share and utilize information about what is happening in health facilities. Social accountability also suggests that informed patients can collectively pressure health providers to improve healthcare services.

Most social accountability programmes in health care have proven unsustainable or were never properly evaluated, but one programme from Uganda stands out as a success (Björkman and Svensson 2009). Unlike many others, this intervention programme gathered information about the quality of care provided by facilities and then disseminated it to households.

Active patients would normally learn about changes in quality only slowly; therefore, direct provision of information, such as in this case, could be useful to households and could lead to immediate changes in health-seeking behaviour. The information aspect of this programme is common to many programmes that do not address accountability. Many examples of information campaigns from the family planning literature (which has long recognized that women do not automatically seek care) and programmes such as the Gold Star Quality Program have relied on disseminating information about the quality of health care in addition to telling people what they should expect. Although such information campaigns assist active patients in choosing facilities, they do not change the fundamental relationship of the household to the health system. In particular, there is little evidence that they enable individuals to bargain directly with providers or that they increase the quality of care available.

The evidence from the experiment in Uganda, however, does suggest that the programme did more than provide information. In theory, social accountability should give households collective power in choosing health providers. However, it is not clear how collective pressure can force health facilities to change the quality of care they provide in a public health system that previously relied on hierarchical supervision. Björkman and Svensson report that the facilities in Uganda did improve the quality of care provided. Some of the inputs were observable (absenteeism), but others are less so (effort). In some cases, the link between accountability and quality is clear; patients learned about services that health workers were required to provide free and did in fact demand more of these services. In addition, it is possible that pressure from the implementers led to increased quality. In that case, patients were learning more about quality that was also increasing and what was designed to increase the accountability of providers to their local communities was in fact a new (and better) form of the hierarchical supervision contract. Clearly, information is necessary in any programme that addresses social accountability, but more work can be done to show that new forms of contracts are available to communities.

### Private health care

In most developing countries, private health care represents different types of services sold to different groups of individuals. In many countries, the private sector delivers observable services (such as ultrasound and maternity care) and provides high-quality care to the wealthy (Bennett *et al.* 1997). Whereas the lower end of the quality scale (e.g. itinerant drug salesmen and other peddlers) is well populated with private providers, they generally do not sell services associated with asymmetric information, such as consultation and diagnosis. In a corrupt health system, health workers who are trained and paid by the state sell private services to customers during or after their normal work hours. Even with this implicit public subsidy of private services, there is little evidence that these practitioners provide quality diagnostic care. In some countries, bribes are the only way to receive adequate nursing care and medical attention for inpatient services, but these services are observable and therefore it is not surprising that quality can be purchased. Corrupt public-sector practitioners should face

similar economic incentives as private providers (they want to attract more patients who are willing to pay for their services). Thus, the fact that corruption does not lead to increased quality calls into question the potential role that private health care can play in providing quality services to poor people.

To succeed in selling high-quality unobservable services to the general population, private practitioners must attract active patients by offering explicit or implicit contracts. As discussed earlier, individual private for-profit practitioners providing modern medicine are not in a position to utilize the informal contracts used by either traditional healers or public facilities with hierarchical supervision models. This is because the contracts used at traditional healers are difficult to enforce (particularly to doctors who provide care to people outside of their local community) and because the hierarchical supervision contract is based on a model that is not applicable to the private practitioner. Thus, the only contract currently available to them is the reputation contract. However, it is difficult for the individual practitioner to earn rewards for delivering high-quality unobservable health inputs, such as diagnostic services, because developing a good reputation takes considerable time when patients learn about quality slowly. This leads to a significant lag between when the practitioner begins to provide high-quality care and when patients will be willing to pay for this higher quality—a period difficult to bridge financially for most individual practitioners. Thus, private practitioners could easily fail financially before they can establish their reputation despite the high patient demand for quality.

This institutional gap suggests two possible future lines of investigation on the prospects of privatization. First, private practitioners can seek new contract types that signal quality to active patients. Private practitioners in many developing countries have traditionally been sole practitioners in separate practices, but this does not have to be the case. If practitioners can become part of a larger hierarchy, they can potentially gain from the types of contracts used in other organizations. There is evidence from South Africa that exactly such a network of private practitioners can provide adequate quality care to the poor (Palmer *et al.* 2003). Alternatively, practitioners could become active members in a medical association or part of a network based on mutual referrals. However, in the absence of structures that allow for the formation of contracts with active patients, practitioners cannot immediately sell high-quality services to their patients because their patients have no assurances that they are receiving high quality until enough time has passed that they can observe the benefits.

The possibility of reputation overcoming the market failures that arise from asymmetric information offers a second line of inquiry concerning the prospects of privatization. Because reputation is the only contract available, the active patient will learn only very slowly about the quality provided by a single private healthcare practitioner. This suggests a role for organizations that collect and publish information about quality. Private practitioners might be willing to submit themselves to evaluation if that information was made publicly available. In fact, well-designed accreditation schemes should be most attractive to small private providers who cannot otherwise guarantee their quality of care. Unlike the



information campaigns mentioned earlier, accreditation programmes usually report whether a facility has met the standards for accreditation and do not examine a range of facilities. However, this information is less useful to patients than information on the level of inputs at a number of facilities between which they can choose.

### Performance pay

The goal of performance pay is to give health workers direct incentives to increase the quantity and quality of health care they provide. In many ways, it is similar to a fee-for-service health insurance system in which the insurance company reimburses the expenses of patients. Basinga *et al.* (2011) report on the positive experiences of a performance pay plan in Rwanda, where health facilities receive funds tied to increasing the quantity of health services provided. Unlike traditional fee-for-service, in the Rwanda implementation, the payment depends on externally verified quality measures.

If pay for performance is going to work, changes in the quality of care must elicit a response from patients. The active patient model suggests that patients will respond immediately to changes in the provision of observable inputs, so reducing absenteeism, eliminating informal payments (in particular bribes for access to care), ensuring medicines are in stock, and even being polite are all ways to attract patients to a facility. The level of services in Rwanda before the programme reported in Basinga *et al.* was low enough that demand increased because of improvements in observable measures and the new patients should begin to notice the impact of higher quality unobservable services. However, in the long run and in countries with higher baseline measures of quality, attracting more patients will require providing higher-quality unobservable services. The active patient model suggests that increases in the level of unobservable services will change the caseload or case mix only after patients can learn about quality. Thus, it is overly optimistic to expect that increases in quality will immediately lead to increases in quantity (and therefore greater payments to health facilities that can translate into pay for performance for workers). In P4P programmes without immediate rewards for quality, health workers have few incentives to improve quality of unobservable services.

In countries where healthcare performance is very low, pay for performance can achieve significant results precisely because patients are active and respond to higher levels of observable quality. The long-run success of pay for performance in settings above minimum levels of quality will depend on giving clinicians tools with which to commit to higher quality and ways of signalling this commitment to patients. A tool such as paying facilities when they attract more patients is similar to a reputation contract, which has not been widely successful to date. Another tool for committing to higher quality, which is more likely to work, is a system in which the external implementer can directly encourage high-quality care and simultaneously signal to patients that quality has improved. The facility in this system has an immediate reason to improve quality and, if the external implementer is credible and known to the community, the active patient will immediately respond to the promise of higher-quality care. Once patients know that quality is higher, their additional willingness to pay for this

quality will make the programme sustainable and may even replace the incentives that were provided by external implementers.

## Conclusions

The active patient model summarizes much of what has been learned from the study of health-seeking behaviour in developing countries over the past 20 years. Patients seek health care in a rational and complex manner. They understand that health care can increase the chances of their returning to health and that quality is important in this transformation. They understand that different illness conditions require different levels of inputs. They cannot immediately assess many of the inputs that are provided to them, but they do understand that different facilities and organizations provide different levels of inputs. Most frequently, they improve outcomes by actively choosing among providers that offer differing quality of health care.

The literature demonstrates the significant energy patients exert in their search for health care. Patients learn about quality, how it applies to different illnesses and how symptoms relate to various illnesses. They travel to seek quality care when necessary, even bypassing closer facilities. When patients are not cured, they travel even farther in search of cures. Ideally, these energies can be harnessed and turned into productive forces in the healthcare system.

When patients are viewed as passive, these energies are ignored, and well-intended policies often work against patients. For example, when fees are raised at government clinics, ministry officials may observe that many patients are no longer attending and conclude that they are no longer seeking care. If the fees are then dropped in response, the reforms that could have been paid for with the extra revenue must be shelved. In fact, patients may simply have turned to other sources of health care providing similar quality at lower cost. If quality reforms had been implemented at the government clinics, evidence suggests that most patients would have returned and welfare would have increased (Litvack and Bodart 1993).

On the other hand, when fees are reduced or eliminated, new patients will come to public facilities either from other facilities or from self-care. This may be taken as evidence of high price elasticity for health care when in fact patients are simply switching from one facility to another. If quality falls because the added caseload leads to insufficient resources, existing patients will be worse off than before the fees were lowered or eliminated even if new patients gain. In this case, the health facilities will be busy with self-limiting illnesses, and patients who suffer with more severe illnesses will be forced to travel greater distances in search of care. In the end, the cost savings from lower care fees at the nearby facility will be transferred to transportation costs or lost in hours of walking to find better care.

Because patients are active care seekers and have difficulties observing the quality of care, policies that simultaneously increase quality and signal these increases to patients will attract new patients more quickly than policies that only increase quality. Where incentives for health workers are properly designed, this increasing caseload can be an additional stimulus to health worker income. The demand for quality

services does not automatically create a supply of high-quality care because high-quality care is composed of both observable and unobservable services and not all providers have access to contracts that ensure the delivery of unobservable care. However, properly designed institutions can exploit the self-reinforcing discipline that comes from incorporating the significant energies expended by active patients.

## Funding

This paper was written with financial support from the Knowledge Product, Human Resources for Health Group at the World Bank.

## Endnote

<sup>1</sup> The value of a particular treatment  $T$ , with cost  $C$ , is  $T(\rho^* - \rho^\phi) (V^G - V^B) - C$ .  $T$  and  $C$  are features of the healthcare provider chosen,  $(\rho^* - \rho^\phi)$  is the change in the probability of being cured with the best treatment ( $\rho^*$ ) and no treatment ( $\rho^\phi$ ) and  $(V^G - V^B)$  is the difference in the value to a patient of being cured compared with not being cured.

## References

- Akin JS, Hutchinson P. 1999. Health-care facility choice and the phenomenon of bypassing. *Health Policy and Planning* **14**: 135–51.
- Arrow KJ. 1963. Uncertainty and the welfare economics of medical care. *American Economic Review* **53**: 941–73.
- Basinga P, Gertler PJ, Soucat A, Sturdy J. 2011. Effect on maternal and child health services in Rwanda of payment to primary health-care providers for performance: an impact evaluation. *Lancet* **377**: 1421–8.
- Bennett S, McPake B, Mills A (eds). 1997. *Private Health Providers in Developing Countries: Serving the Public Interest?* London and New Jersey: Zed Books.
- Björkman M, Svensson J. 2009. Power to the people: evidence from a randomized field experiment on community-based monitoring in Uganda. *The Quarterly Journal of Economics* **124**: 735–69.
- Bloom G, Standing H, Lloyd R. 2008. Markets, information asymmetry and health care: towards new social contracts. *Social Science and Medicine* **66**: 2076–87.
- Bolduc D, Lacroix G, Muller C. 1996. A comparison of medical providers in rural Benin: a comparison of choice based models. *Journal of Health Economics* **15**: 477–98.
- Das J, Hammer JS. 2007. Money for nothing, the dire straits of medical practice in Delhi, India. *Journal of Development Economics* **83**: 1–36.
- Das J, Hammer JS, Leonard KL. 2008. The quality of medical advice in low-income countries. *Journal of Economic Perspectives* **22**: 93–114.
- de Savigny D, Mayombana C, Mwangeni E *et al.* 2004. Care-seeking patterns for fatal malaria in Tanzania. *Malaria Journal* **3**.
- Dor A, Gertler P, van der Gaag J. 1987. Non-price rationing and the choice of medical care providers in rural Cote D'Ivoire. *Journal of Health Economics* **6**: 291–305.
- Dranove D, White WD. 1987. Agency and the organization of health care delivery. *Inquiry* **24**: 405–15.
- Ellis RP, McGuire TG. 1990. Optimal payment systems for health services. *Journal of Health Economics* **9**: 375–96.
- Gauthier B, Wane W. 2011. Bypassing health providers: the quest for better price and quality of health care in Chad. *Social Science and Medicine* **74**: 540–9.
- Gertler P, van der Gaag J. 1990. *The Willingness to Pay for Medical Care: Evidence from Two Developing Countries*. Baltimore, MD: Johns Hopkins University Press. Published for The World Bank.
- Gertler P, Locay L, Sanderson W. 1987. Are user fees regressive? The welfare implications of health care financing proposals in Peru. *Journal of Econometrics* **36**: 67–8.
- Gilson L. 2003. Trust and the development of health care as a social institution. *Social Science and Medicine* **56**: 1453–68.
- Gilson L. 2005. Editorial: building trust and value in health systems in low- and middle-income countries. *Social Science and Medicine* **61**: 1381–4.
- Grobler C, Stuart IC. 2007. Health care provider choice. *South African Journal of Economics* **75**: 327–50.
- Hanson K, McPake B, Nakamba P, Archard L. 2005. Preferences for hospital quality in Zambia: results from a discrete choice experiment. *Health Economics* **14**: 687–701.
- Hanson K, Yip WC, Hsiao W. 2004. The impact of quality on the demand for outpatient services in Cyprus. *Health Economics* **13**: 1167–80.
- Hölmstrom B. 1982. Moral hazard in teams. *Bell Journal of Economics* **13**: 324–40.
- Kabakian-Khasholian T, Campbell O, Shediak-Rizkallaha M, Ghorayebc F. 2000. Women's experiences of maternity care: satisfaction or passivity? *Social Science and Medicine* **51**: 103–13.
- Klemick H, Leonard KL, Masatu MC. 2009. Defining access to health care: evidence on the importance of quality and distance in rural Tanzania. *American Journal of Agricultural Economics* **91**: 347–58.
- Kruk ME, Mbaruku G, McCord CW *et al.* 2009. Bypassing primary care facilities for childbirth: a population-based study in rural Tanzania. *Health Policy and Planning* **24**: 279–88.
- Leonard KL. 2002. When states and markets fail: asymmetric information and the role of NGOs in African health care. *International Review of Law and Economics* **22**: 61–80.
- Leonard KL. 2003. African traditional healers and outcome-contingent contracts in health care. *Journal of Development Economics* **71**: 1–22.
- Leonard KL. 2007. Learning in health care: evidence of learning about clinician quality in Tanzania. *Economic Development and Cultural Change* **55**: 531–55.
- Leonard KL. 2009. The cost of imperfect agency in health care: an empirical investigation in rural Cameroun. *Journal of Development Economics* **88**: 282–91.
- Leonard KL. 2011. *Improving Health Outcomes by Choosing Better Doctors: Evidence of Social Learning about Doctor Quality from rural Tanzania*. Maryland, USA: University of Maryland.
- Leonard KL, Mliga G, Mariam DH. 2002. Bypassing health facilities in Tanzania: revealed preferences for observable and unobservable quality. *Journal of African Economies* **11**: 441–71.
- Lewis M. 2007. Informal payments and the financing of health care in developing and transition countries. *Health Affairs* **26**: 984–97.
- Lindelov M. 2004. Understanding spatial variation in the utilization of health services: does quality matter? *Technical Report 212*, The Centre for the Study of African Economies.
- Lindkvist I. 2012. Informal payments and health worker effort: a quantitative study from Tanzania. *Health Economics*. In press.
- Litvack JI, Bodart C. 1993. User fees plus quality equals improved access to health care: results of a field experiment in Cameroun. *Social Science and Medicine* **37**: 369–83.
- Maestad O, Torsvik G. 2008. *Improving the Quality of Health Care when Health Workers are in Short Supply*. Bergen, Norway: Chr. Michelsen Institute.

- Maestad O, Torsvik G, Aakvik A. 2010. Overworked? On the relationship between workload and health worker performance. *Journal of Health Economics* **29**: 686–98.
- Mariko M. 2003. Quality of care and the demand for health services in Bamako, Mali: the specific roles of structural, process, and outcome components. *Social Science and Medicine* **56**: 1183–96.
- Meessen B, Gilson L, Tibouti A. 2011. User fee removal in low-income countries: sharing knowledge to support managed implementation. *Health Policy and Planning* **26**: ii1–ii4.
- Mwabu GM. 1986. Health care decisions at the household level: results of a rural health survey in Kenya. *Social Science and Medicine* **22**: 315–19.
- Mwabu GM. 1989a. Nonmonetary factors in household choice of medical facilities. *Economic Development and Cultural Change* **37**: 383–92.
- Mwabu GM. 1989b. Referral systems and health care seeking behavior of patients: an economic analysis. *World Development* **17**: 85–92.
- Mwabu GM, Mwangi WM. 1986. Health care financing in Kenya: a simulation of welfare effects of user fees. *Social Science and Medicine* **22**: 763–7.
- Mwabu GM, Ainsworth M, Nyamete A. 1993. Quality of medical care and choice of medical treatment in Kenya: an empirical analysis. *Journal of Human Resources* **28**: 838–62.
- Palmer N, Mills A, Wade H, Gilson L, Schneider H. 2003. A new face for private providers in developing countries: what implications for public health? *Bulletin of the World Health Organization* **81**: 292–7.
- Rowe AK, de Savigny D, Lanata CF, Victora CG. 2005. How can we achieve and maintain high-quality performance of health workers in low-resource settings? *Lancet* **366**: 1026–35.
- Sahn DE, Younger SD, Genicot G. 2003. The demand for health care services in rural Tanzania. *Oxford Bulletin of Economics and Statistics* **65**: 241–59.
- Sauerborn R, Nougara A, Diesfeld HJ. 1989. Low utilization of community health workers: results from a household interview survey in Burkina Faso. *Social Science and Medicine* **29**: 1163–74.
- Starr P. 1982. *The Social Transformation of American Medicine: The rise of a sovereign profession and the making of a vast industry*. NY: Basic Books.
- WHO. 2005. Bridging the “Know-Do” gap: meeting on knowledge translation in global health. *Technical Report*. WHO/EIP/KMS/2006.
- World Bank. 2003. *World Development Report 2004: Making Services Work for Poor People*. Washington, DC: The World Bank and Oxford University Press.