2/3/2013

THE WORLD BANK GROUP

MEASURING THE QUALITY OF HEALTH CARE IN CLINICS

Alaka Holla

February, 2013

CONTENTS

February, 2013

ACKNOWLEDGMENTS

The author would like to thank (without implicating) the following people for helpful comments on an early draft: Aneesa Arur, Jishnu Das, Ariel Fiszbein, Gyuri Fritsche, Benjamin Loevinsohn, Son Nam Nguyen, Santhosh Srinivasan, and Christel Vermeersch.

February, 2013

INTRODUCTION

Until quite recently, governments and aid organizations have tried to improve health outcomes by investing in infrastructure, medical equipment, and drug supplies and by ensuring access to a certain number of qualified health workers. Underlying these investments is the assumption that these inputs will then translate into quality health care for patients when they walk into a clinic – that is, once these are in place, providers will be able to examine, diagnose, and treat patients appropriately. While these inputs are of course necessary for these functions, they are by no means sufficient. The presence of stethoscope is no guarantee that a medical provider will use it when necessary. Similarly, a regular stock of modern drugs does not tell us if providers are prescribing tranquilizers for a simple cold or whether a provider can diagnose if a patient even has a cold.

A number of recent studies in both high- and low-income settings have made progress on measuring this type of medical advice available in clinics, where the focus is on the interaction between the patient and provider during the clinical encounter. They have quantified a number of trends such as low adherence to medical guidelines (Hutchison et al, 1991; Leonard and Masatu, 2005; Das et al, 2012), a divergence between what providers know and what they put into practice (Rethans et al, 1991; Thamlikitkul, 1991; Igun, 1994; Grad et al, 1997; Leonard and Masatu, 2005; Das and Hammer, 2007; Ozuah and Reznik, 2007), and sizeable differences between the public and private sectors (Das and Hammer, 2007; Das et al, 2012).

This note provides a brief overview of the surveys designed to measure the quality of care within clinics, focusing in particular on the quality of examination, diagnosis, and treatment offered to patients in the primary care setting. Part 1 describes the various survey options available to a research team, focusing on the research questions that each instrument can address along with their advantages and disadvantages relative to the other instruments. Part 2 outlines the basic steps required for developing and implementing these surveys and aims to provide a realistic sense of the time and personnel requirements associated with each different instrument. This section draws heavily from my experiences from the Medical Advice, Quality, and Availability in Rural India (MAQARI) study, which was funded by the Global Health Program of the Bill & Melinda Gates Foundation and which I have worked on as part of a consortium of researchers investigating the availability and quality of health care in India.[1]

Partly due to the logistical and ethical difficulties inherent in observing patient-provider interactions in an acute care setting, this kind of measurement has been mostly limited to primary care services in much of the recent literature. Thus, the descriptions and guidance contained in this note are unlikely to transport well to secondary and tertiary care settings.

This note also draws heavily from surveys implemented in low-income countries, although, in this case, much of the experiences can inform quality measurements in settings with more resources. These surveys have been implemented as part of research studies, rather than incorporated into routine monitoring of service delivery or into a health management information system. Recent experiences with pay-for-performance schemes in clinics (for example, Basinga et al, 2011) demonstrate how these kinds of metrics can be extended to policies aiming to improve quality of care.

[1] This team includes Brian Chan, Jishnu Das, Veena Das, Michael Kremer, Manoj Mohanan, Karthik Muralidharan, Diana Tabak, the Center for Policy Research (New Delhi), and the Institute of Socio-economic Research on Development and Democracy (India).

PART 1: SURVEY OPTIONS

THE INSTRUMENTS AND WHAT THEY MEASURE

This section briefly describes four commonly used instruments to measure the quality of advice in clinics: medical vignettes, provider observations (also known as direct clinical observations), patient exit interviews, and unannounced standardized patients (also known as incognito or simulated patients). For examples of the data that each instrument can generate, see recent reviews of vignettes, provider observations, and exit interviews in low-income countries (Gertler and Das, 2007) and standardized patients in both high (Rethans et al, 2007) and low (Madden et al, 2007) income counties. For examples of the survey instruments themselves, see the Human Development Network's website on Quality of Care and the website of the MAQARI project (www.healthandeducationinindia.org).

MEDICAL VIGNETTES

A medical vignette is a hypothetical medical case posed to health care providers, designed to elicit what they think is an appropriate course of medical history-taking and examination, their diagnostic ability, and proposals for treatment. Typically, two enumerators sit with a provider and play two distinct roles – one acts like the patient and responds to any history questions that the provider might pose. The second plays the role of the omniscient narrator – he or she reveals the outcomes of any examinations the provider suggests.

Take an unstable angina case, for example. The omniscient narrator describes a patient and his symptoms - for example, " A 45 year old male walks into your clinic, complaining of severe chest pain" – and assures the provider that the patient will comply with all of his suggestions for diagnostic testing and treatment. The provider must then indicate how he would proceed with the case, and the omniscient narrator notes down on a structured form the specific history-taking questions and examinations that the provider says he would complete and any articulated diagnoses or prescribed treatment that he mentions. If, for example, he asks the enumerator acting like the patient, *Do you feel pain anywhere else?,* the enumerator acting like the patient would respond "Yes, down my arm too." If the provider then says that he would take the patient's blood pressure, the omniscient narrator would respond with "130/90." In fact, for each question and examination the doctor could possibly mention, whether medically relevant or not, the team has a prepared answer so that the provider can use that information for the rest of the case and so that all interviewed providers can see the same exact case.

Given that we are only asking the provider what he *would* do when faced with a similar case and that we are giving him ample amount of time to consider the details of the case and his treatment options, this instrument can only measure a provider's *competence* or medical knowledge and what he would do in the best-case scenario. To what extent this measured competence translates into actual *practice* is an empirical question – an empirical question that can be answered using some of the other techniques described here. Existing evidence from both high and low income countries, however, already suggests that competence and practice do diverge (Rethans et al, 1991; Thamlikitkul, 1991; Igun, 1994; Grad et al, 1997; Leonard and Masatu, 2005; Das and Hammer, 2007; Ozuah and Reznik, 2007) and that provider *effort* is also an important determinant of the quality of care received in clinics. [2] In

[2] An exception is Peabody (2000) who finds little difference between doctors' scores on his copyrighted version of medical vignettes (Clinical Performance Vignettes) and what could be obtained from standardized patients from a study implemented in

the Arusha region of Tanzania, for example, one study compared the history questions and physical examinations mentioned during vignettes to what providers actually did during provider observations for three presenting conditions (fever, cough, and diarrhea) that had specific checklists of what providers had to do to make a differential diagnosis and treat the patient appropriately. There was only a 53 percent chance that a correct question or examination elicited during medical vignettes translated into a correct question or examination during provider observations (Leonard and Masatu, 2005). Similarly, in urban Delhi, public sector doctors asked an average of 29 percent of necessary questions for diarrhea and cough with fever during vignettes, which fell to 8 percent when they were observed in practice (Das and Hammer, 2007).

PROVIDER OBSERVATION

The simplest way to observe how a provider practices is to sit in a clinic and see how he behaves with his real patients. This is essentially what happens during provider observations, sometimes also called direct clinical observations. An enumerator sits in a clinic for an entire day or sometimes one shift and notes down on a structured form information similar to what is obtained during medical vignettes - history taking, examinations, diagnoses, and treatments – in addition to any fees charged in the clinic. The survey could also capture if the provider cares to see the patient in a follow-up visit and the quality of communication – for example, if providers mention a diagnosis, explain what it means, and provide dosage information about medicines.

While this direct observation technique could encourage providers to perform better than they would in the absence of an observer in his/her clinic (i.e. we might worry about bias from Hawthorne effects), research in clinics in Tanzania suggests that provider effort spikes after an observer enters the clinic and returns to normal after the first 10 to 15 patients (Leonard and Masatu, 2006).

There are other issues of concern, however, that need to be considered when using this instrument. Because patients can walk in with any case and because observers (including possibly the provider) do not know what ailment the patient is suffering from, for most primary care conditions, it is generally not possible to note the identity of particular questions – rather we can document more general attributes of history-taking, such as the number of questions and which systems were asked about - nor can we evaluate diagnoses and treatment. Instead, we are limited to estimating the average frequency of more process-oriented measures – such as articulation of any diagnosis, number of physical examinations, injections, antibiotic prescriptions, etc. Thus, while this exercise can help us describe the typical patient's experience when they walk into a clinic, we cannot determine whether or not they are leaving with the appropriate medical advice and treatment.

The variety of patients and cases seen in real clinical interactions also prevents us from comparing quality across doctors or types of doctors (e.g. public vs. private or qualified vs. unqualified). Some providers may not only get particular types of medical cases (e.g. cardiovascular disease, digestion problems) but also particular types of patients (e.g. patients with a long delay between the onset of symptoms and the clinic visit , patients with poor communication skills, patients from a particular ethnic community). Thus, some providers may need to ask more questions or perform more exams, not because they are better doctors but because they have more complicated cases or needier patients. This makes it difficult to compare providers' quality.

---

two Veterans Affairs medical centers in the United States. A CPV differs from the medical vignettes described above. They are self-administered paper questionnaires, in which more and more details of the case are revealed throughout the questionnaire.

Relying on real patients also makes it difficult to assess how providers manage less prevalent, but important, illnesses and how they treat patients from minority populations. A frontline provider, for example, may only receive one unstable angina patient a month, but given the fatality rates for this condition if not managed appropriately and given patients' inability to triage themselves adequately to secondary or tertiary care, we might want to know how a primary care provider would handle such a case. Similarly, we might have to wait around for days or weeks to see a sufficient number of minority patients to detect disparities in treatment with any degree of statistical precision.

PATIENT EXIT INTERVIEWS

Enumerators can note down their observations of patients as well, but beyond approximations of age and wealth and possibly presenting symptoms, it is very difficult to collect detailed data on patients during provider observations without disrupting a provider's practice. We can, however, interview them immediately after they exit the clinic and ask about their interaction with the provider. This kind of survey allows us to fill in some of this missing detail about patients, their illnesses, and ascertain their satisfaction with the type of care they received. While we still cannot determine with much certainty what the patient was suffering from and therefore cannot evaluate providers' diagnostic ability or treatment prescriptions, we can assess the extent to which patient characteristics vary across providers.

In some cases, however, we do know what the doctor should have done, and we can ask patients about examinations or any information provided, as long as they can recall them easily. Take prenatal visits, for example, and the measurement of uterine height. While a patient might not be able to say with certainty whether a provider has taken her pulse or checked her ankles for swelling, she should be able to say whether the size of her belly was measured.

Some studies have interviewed patients in their homes. While this might enable the interviewer to ask more questions and perhaps elicit more candid responses compared to an interview outside the clinic, recall could be a constraint. In one study in Delhi, for example, patients found it difficult to remember that they had gone a doctor 30 days after the visit (Das, Hammer, and Sanchez 2012).

STANDARDIZED PATIENTS

A standardized patient is an individual trained to consistently portray a medical case and all of its physical and psycho-social aspects and to accurately recall his/her interactions with providers.[3] After the clinical interaction, details of the visit are recorded using a structured form similar (or identical) to the form used to describe interactions during the vignettes, which documents whether or not history questions and exams were completed, what diagnosis was offered (if any), and what treatments were prescribed. Alternatively, if the clinical circumstances permit it, the interaction can be recorded with a hidden microphone or camera and coded afterwards.

[3] Simulated patients need not be restricted to live actors. Doctors can view video recordings (e.g. Shulman et al, 1999) or perform on anatomical models designed to simulate certain medical situations, such as the NeoNatalie and MamaNatalie simulators developed by Laerdal Global Health, designed to train birth attendants in proper obstetrics, newborn care, and neonatal resuscitation. The quality of care measured using these kinds of simulated patient, however, may not capture providers' incentives to perform when faced with real patients.

Most OECD countries use standardized patients in medical education and licensing exams, but when used in quality audits, unannounced standardized patients can circumvent many of the limitations posed by the alternative methods to measure provider practice discussed earlier. First, providers do not know the identity of the standardized patient and therefore will treat him/her just like any other patient. This takes care of the potential Hawthorne effects (observation bias) of provider observations. Second, since the standardized patients are portraying specific medical cases, we know what the correct diagnosis and form of treatment should be and can therefore make direct comparisons between a provider's performance on the vignettes and provider's behavior recorded from their interactions with standardized patients. Finally, the fixed cases and the standardization of case presentation across "patients" ensure equivalent case and patient mixes across providers. This makes it easier to compare the effort of different providers (or types of providers, such as public and private doctors).

While they are the gold standard for measuring the quality of care patients actually receive when they enter a clinic (Rethans, 2007), standardized patients do come with their own set of limitations. First, and most obviously, the cases that the standardized patients present must be easily simulated. It would be very difficult, if not impossible, to simulate a bone fracture or wound. The one-time visit also precludes a proper assessment of providers' performance in managing chronic illnesses, such as HIV, tuberculosis, diabetes, or heart disease, although there is nothing that would prevent a standardized patient from visiting a provider multiple times under the same guise. This restriction to particular cases poses a challenge for extrapolating the quality measurements generated by standardized patients to the providers' normal patient pool. If health care providers' quality differs across illness conditions (for example, a doctor might treat tuberculosis appropriately but would be useless for a case of unstable angina), or if providers have specialized in certain conditions, then the circumscribed set of conditions that a standardized patient would portray might not give an accurate picture of a given provider's quality.

Second, and perhaps less obviously, any examinations triggered by the case must not endanger the standardized patient. In India, for example, we tried to choose cases that would minimize the probability that a provider would check a patient's temperature because providers in our sample clinics often did not disinfect their thermometers between patients. Therefore, we had to avoid all communicable illnesses.

Third, in many instances – for example, very remote, rural areas or providers that do not accept "walk-ins" – designers of the study must also construct a ruse for standardized patients to enter a clinic when a provider expects to know all of his patients. In India, our standardized patients had to be traveling through the area – to visit relatives, for example, or to conduct a tree census, depending on what would be more believable in the sample village – and thus had to be familiar with the names of roads and nearby villages.

Finally, providers, especially medically qualified doctors, do not like standardized patients or audits more generally. Sometimes they do not acknowledge that their quality can be evaluated by non-doctors (which does not occur with standardized patients anyway, since standardized patients are simply recalling whether certain questions or exams were completed rather than rating the providers' quality directly). Sometimes, however, they have legitimate ethical concerns, such as potential negative externalities on their real patients if they are devoting time to "fake" patients that do not really require medical attention.

RECORD REVIEWS

Reviewing medical records kept on each patient in the clinic could be another option for measuring the quality of clinical interactions, especially if providers' diagnoses, treatments, and management strategies can be extracted from the charts rather than just basic information about the patient and presenting symptoms. In most low-income settings, however, these kinds of records contain very little information, if they exist at all. Besides, even in high-income settings, medical charts contain much less information than what they are meant to document. One study in the United States examining the quality of care received by adults, for example, found that only 15 percent of the charts of elderly patients indicated that they had received an influence vaccine, but 85 percent of the same patients reported that they had received one during a subsequent interview (McGlynn et al, 2003) . In another U.S. study, process scores were 10 percentage points lower when the medical records corresponding to standardized patients' visits were used to evaluate doctors than when quality was measured from the reports of standardized patients (Peabody et al, 2000).

PROVIDER MAPPING

These clinic-based instruments all give us a picture of what kind of care patients *could* receive when they enter a clinic. What kind of care they ultimately *do* receive depends on which providers they end up visiting for their health care needs. If a very popular provider exhibited very poor quality, we might be much more concerned than if that provider only saw two patients per week. Moreover, we might not want to waste time measuring quality among the infrequently visited providers and make sure we target the providers with the most patients. The market share of each provider, however, is often unknown. A provider mapping and household census (or survey) could remedy this and generate a sampling frame from which providers could be selected for quality surveys.

Ideally, the government or local medical association has a list of all providers practicing in the areas of study. In low-income settings with little regulation of the private sector, frequent public-sector transfers, not to mention frequent provider absence, these lists are unlikely to be accurate. In this case, the study team may have to conduct a provider mapping.

In a provider mapping, enumerators first visit a village, neighborhood, or any other geographically defined population of interest and ask them about the geographic locations (e.g. adjacent villages) where they seek primary care services. Then, within these mentioned locations, enumerators must identify *all* providers offering primary care services (or whichever health service is under study), regardless of whether those providers had been mentioned by name by people in the original area (Part 2 discusses techniques to ensure that all providers are mapped). They assign these providers a unique identification number, and interview them briefly. Ideally, they are also literally placed on a map with GPS coordinates. Then households in the original village are asked about the providers they have visited in the previous 30 days, and enumerators can exactly identify the providers they mention, allowing us to calculate the share of visits captured by each provider.

CHOOSING INSTRUMENTS

To choose among these instruments, it is first useful to pinpoint the exact measurement goal and to determine what would be feasible in the context under study. This section discusses a number of possible research objectives and the combination of surveys and survey protocols that could be used to meet them.

MEASURING AVERAGE QUALITY

Instruments: ***provider mapping + quality instrument***

In many contexts, the quality of health care delivered in clinics has never been measured, and the purpose of the exercise is to perform a basic assessment of quality or to obtain a baseline measure prior to an intervention that aims to improve quality. If this is the case, then a reasonable starting point would be to decide whether the exercise should measure quality of the average health care *provider*, quality of the average *public provider*, or quality of care available to the average *household or individual*, as these could be three very different things, depending on which providers households tend to visit more frequently. In India and Cambodia, for example, most households – even the poorest - visit unlicensed primary care providers in the private sector who have very little (often no) medical training. In rural Madhya Pradesh, for example, the private sector accounted for 91 percent of all primary care visits; 76 percent of all visits were to providers with no medical training whatsoever. According to a recent report, the private sector also captures a large share – 50% on average – of the market for primary care services in the Africa region (World Bank, 2011). Often, however, studies focus on public clinics or licensed doctors that have been recognized by the government, a medical board, or a doctors' association. While measuring quality of public providers would provide some indication of the efficiency of government spending on public health workers, it will not tell us about what kind of clinical encounters the average person experiences when they seek medical attention in such settings. Thus, to measure quality for the average patient or household, the research team would need to take into account the market shares of the providers in their sample.

Ensuring that quality is measured in the entire health care market is especially important in the context of an impact evaluation since the intervention in question could trigger changes in people's choice of provider. Say, for example, most patients visit private providers, and an intervention has been designed to increase outreach or decrease out-of-pocket payments in the public sector. If this draws some patients out of the private sector and if the private sector exhibits much higher levels of quality to begin with, then we have just lowered the average level of quality obtained. Of course, patients might be paying a lower price now, and a full analysis of the intervention's welfare implications would take this into account, but in terms of quality, patients would be worse off after the intervention. Thus, knowing the market shares of providers *and* their relative quality would help predict whether or not an intervention would improve the quality of care experienced by the average person in a healthcare market.[4]

Once it has been decided *who* will be in the sample, the research team will need to specify *what* exactly they mean by quality – for example, whether they want to measure what providers know about certain illnesses, what knowledge providers ultimately translate into practice once patients enter their clinics, whether providers adhere to national guidelines and regulations, and what role patients play in determining quality. The following sections describe how the survey instruments described earlier can be deployed to measure these various dimensions of quality.

[4] The same argument would go through if we only cared about the quality of care accessed by the poor. To predict what would happen to the quality of care they receive following an intervention that seeks to bring them into certain health clinics, one would need to know where the poor currently seek care and how the quality of care there compares with quality in the targeted clinics.

DISTINGUISHING BETWEEN WEAK KNOWLEDGE AND WEAK PERFORMANCE

Instruments: ***provider observations or standardized patients + medical vignettes***

When we observe low levels of care experienced by real patients or standardized patients, the source of this shortfall is not immediately clear - do providers, for example, not know how to care for patients or certain illnesses, or do they know but just fail to translate their knowledge into practice?

Distinguishing among such mechanisms would be important as they each suggest the need for a different policy response to improve quality. If providers simply do not know how to diagnose or treat certain illnesses, then provider training might help. If older providers exhibit lower knowledge, then continuing medical education could be the optimal training to start with. On the other hand, if providers know what to do but do not translate this into the effort required to implement what they know, then training is likely to be ineffectual.

Indeed researchers in Holland have documented a divergence between doctor's competence (what a doctor is capable of doing) and performance (what he does in his daily practice). They sent unannounced standardized patients presenting 8 common illnesses to general practitioners and then followed up five months later with a test using the same cases, only this time the doctors knew that the patients were being used for evaluation purposes. Doctors' competence scores measured when they knew they were being evaluated turned out to be 49 percent higher than their performance scores measured when they were examining an unannounced standardized patient (Rethans et al, 1991). Something similar occurred in a hospital study in the United States – communication skills in the context of an asthma case were much stronger among announced standardized patients than in unannounced standardized patients (Ozuah and Reznik, 2007). In Canada, a standardized patient study focused on elderly patients also found evidence of a know-do gap, as doctors' knowledge of appropriate drug use correlated poorly with the therapeutic management standardized patients obtained when they visited these providers in their offices (Grad et al, 1997).

These findings are not unique to high income countries. In Tanzania, for example, the history taking, physical examinations, and health education that providers mention during vignettes for patients presenting with fever, cough, or diarrhea do not correlate well with what they do when tending to real patients presenting with the same symptoms (Leonard and Masatu, 2005).

This also seems to be the case among public sector doctors in India. Das and Hammer (2007) found that public doctors in urban Delhi knew how to diagnose and treat 5 illnesses better than private providers (both trained and untrained). Their vignettes scores were much higher. When it came to tending to real patients, however, provider observations revealed that public providers did much less. Das, Holla, Kremer, and Muralidharan (ongoing) provide even stronger evidence of a divergence between knowledge and practice. Standardized patients visited doctors in both their public practices and their private clinics and presented the same cases. The doctors asked many more questions and performed more physical examinations, however, when they were visited in their private clinics. In their public clinics, they performed much worse than unqualified private providers.

Thus, fielding a number of different instruments among the same population of providers would be helpful for revealing which dimensions of quality are the source of the problem.

ADHERENCE TO NATIONAL GUIDELINES OR REGULATIONS

Instruments: ***provider observations or standardized patients***

These investigations into clinical practice can also be used to assess adherence to medical guidelines or national laws and regulations. Because they are typically implemented in the framework of an audit study, unannounced standardized patients are an ideal method for uncovering compliance to rules.

A standardized patient study in Canada, for example, found that that the preventive care guidelines of the Canadian Task Force on the Periodic Health Examination had not been completely integrated into clinical practice (Hutchison et al, 1998). In the standardized patient study in India, we found that providers implemented only 34 percent of a very parsimonious *subset* of the history questions and examinations that the Government of India expects from its doctors in rural areas (thus, the fraction completed of the entire set of guidelines is likely to be much smaller) (Das et al, 2012).

In the Indian study, we are also able to shed light on a number of other governance related issues. Take staffing of public clinics, for example. All health personnel examining, diagnosing, and treating patients in public health clinics should be qualified – either as a doctor or as a nurse. Previous research had already established high rates of official health worker absence in these clinics in India (and in 5 other countries) (Chaudhury et al, 2006). What standardized patients could tell us, however, was who was sitting there in their place. In one round of data collection, we had instructed the standardized patients to receive care from whoever was acting like the doctor in public clinics. This turned out to be someone with no medical training whatsoever 63 percent of the time.

On the other hand, data from this study also revealed very few requests for under-the-table payments in these public clinics. In 98 percent of visits, these providers charged Rs. 5 ($0.09) or less, which is consistent with the schedule of fees in public clinics.

Most countries also prohibit (at least legally) discrimination in health care on the basis of gender or racial/ethnic affiliation. A study in the United States that used videos of standardized patients in a computerized survey found that both female and black patients had 60 percent lower odds of a referral for cardiac catheterization than male and white patients, respectively (Shulman et al, 1999), who presented with the same symptoms.

DEMAND-SIDE AWARENESS

Instruments: ***patient exit interviews + medical vignettes, provider observations, or standardized patients***

Can patients themselves observe and demand quality health care? Or are information asymmetries too great for them to choose high quality providers or to navigate clinical interactions so that they get the best possible care from the provider that they have decided to visit? The answers to these questions would presumably influence the policy solution we would turn to improve the quality of care that patients can access. If, for example, patients cannot distinguish between high and low quality providers, then a government or other interested party might want to provide them information (for example, through a report card) or even shut down the worst providers. If, on the other hand, patients can accurately assess provider quality but cannot access high quality providers because their prices are determined in a market and are too high, then a policymaker might want to consider subsidies or vouchers to make it easier to visit these providers.

The findings on patient awareness from recent empirical work have been mixed. In Uganda, for example, patient satisfaction, as measured in patient exit interviews, did not correlate significantly with the more medical aspects of

the clinical interaction that patients could recall, such as whether or not the provider had performed any physical examination of the patient (Lundberg, 2008). On the other hand, the market shares of the different types of providers in India that we obtained from provider and household censuses suggest that patients prefer to visit providers who exerted more effort when tending to both real patients in provider observations and standardized patients.

Regardless of whether or not patients accurately assess quality, a recent experiment suggests that when patients can signal that they have some information, providers treat them differently. In a standardized patient study in China, patients visited doctors in hospital clinics complaining of flu-like symptoms that did not require treatment with expensive antibiotics (Currie, Lin, and Zhang, 2011). Half of the patients mentioned to the doctor that they had learned from the internet that simple cold and flu patients should not take antibiotics. The other half said nothing. This one sentence made quite a difference. Providers assigned the more informed patient were 22 percentage points (or 35 percent) less likely to prescribe antibiotics. Informed patients' drug expenditures were nearly 40 percent lower as well.

PART 2: SURVEY DEVELOPMENT AND IMPLEMENTATION

This section outlines the basic steps required to design and field the quality instruments presented in Part 1. There are a number of excellent guides and manuals cited in the text that provide more detail and nuance for each method. In particular, the Impact Evaluation Toolkit of the Health Results Innovation Trust Fund also provides templates for terms of reference for the survey firms that would be implementing these kinds of surveys.

When these instruments are implemented together among the same providers, their sequencing could have implications for data quality in each subsequent round of data collection. The suggested order of the instruments below takes these spillovers into account.

INSTRUMENT 1: PROVIDER MAPPING

If a research team wishes to target a certain type of provider, regardless of whether or not they account for the majority of health care visits, or if exhaustive and reliable lists of all providers exist, then a provider mapping may not be necessary. Otherwise, it is a crucial first step both for defining the geographic boundaries of health care markets (which are a function of people's decisions to travel a certain distance to the provider of their choice) and for obtaining the full list of providers who can be sampled for a quality survey.

After primary sampling units (e.g. wards or villages) have been sampled for the study, a full-scale exercise would consist of the following steps:

(i) A *participatory rural appraisal* (PRA), which need not be confined to rural areas, in which households in the primary sampling units are asked about the providers whose case they seek for primary care services and about the locations of these providers.

(ii) A *provider census* in each location mentioned during the PRA, in which all providers in these locations – regardless of whether or not they themselves were mentioned during the PRA - are mapped (preferably geocoded), assigned a unique ID, and interviewed briefly.

(iii) A *household census* or large sample survey in the primary sampling unit, in which households are asked about their health-seeking behavior in the previous 30 days and the identity of any providers visited.

STEP 1: THE PARTICIPATORY RURAL APPRAISAL

In a PRA, the survey team should visit at least three geographically distinct sections of the primary sampling unit and gather people for an informal discussion about health care. In particular, they should define what they mean by primary care services in very simple terms (for example, where you would go for illnesses that do not require an overnight stay in a clinic or hospital) and ask people to name the providers they go to. If after three iterations of this, the exercise still yields providers that had not been mentioned in previous discussions, then the team should choose more locations within the primary sampling unit to conduct additional PRAs until they start to generate no new information. In each discussion, it is important to note down as many identifying details of the providers and their clinics as possible since providers may go by different names or operate more than practice (as is the case in India and Cambodia), and we want to make sure that we are not double-counting providers and/or undercounting clinic locations. For the providers that practice in adjacent primary sampling units – let's call these cluster units - surveyors should inquire about any nearby landmarks that could guide them to the clinic. After these discussions, it is helpful to create one consolidated PRA form and to assign IDs to the providers, clinics, and clusters mentioned during the PRA. This entire process lasts between half a day to a day in each primary sampling unit, depending on the number of cluster units and the number of providers found within the health care market.

STEP 2: PROVIDER CENSUS

Next, surveyors must visit all of the providers mentioned during the PRA to make sure they exist, geocode them, and administer a short facility survey. Within the primary sampling unit and each cluster unit, surveyors should ask all providers mentioned during the PRA, all pharmacists, and shopkeepers for information about other providers who may be practicing in the cluster. These providers are then added to the list of providers, geocoded, interviewed, and asked about the presence of other providers. Through this iterative process, all providers practicing in the primary sampling unit and all clusters mentioned during the PRA can be mapped and identified. Once the team is confident that they have tracked down all of the providers operating out of the primary sampling unit and its cluster units, they should create one final list of providers so that all survey teams can refer to one (and only one) set of IDs during the household census when they ask for the identity of the providers visited.

If people in the clusters are generally ignorant about health care providers practicing in the area, an alternative strategy would be to simply meander down every lane in the cluster and look for clinics. It is important to keep in mind, however, that this kind of exercise would also entail a great deal of consultation with people in the clusters, since clinics – especially those operated by unlicensed private providers – may not have a signboard advertising their medical services. In India, for example, nearly half of the providers in our sample operated without a signboard. Nearly 26 percent practiced from their residences.

This mapping stage is an ideal time to administer a facility survey. Not only does this yield additional information to use when asking households to identify the providers they use (e.g. *Is it the bald man with a big mustache* or *Is it the woman practicing out of the house behind the post-office*), but it also acclimates the providers to the enumerators and builds trust for subsequent quality instruments, such as medical vignettes or provider observations. This facility survey can be as brief or detailed as the study requires. The MAQARI study implemented a very parsimonious instrument, but the Service Availability and Readiness Assessment (SARA) instrument of the World Health Organization and the Service Provision Assessments (SPA) used by the Demographic and Health Surveys would be very comprehensive starting points for teams also interested in documenting more details of clinics and disease-specific inputs.

STEP 3: HOUSEHOLD CENSUS

Conducting a household census (or large sample survey) in the primary sampling unit is the final step in the mapping process and serves two main purposes - (i) matching households to providers to measure providers' market shares and (ii) ensuring that all clusters and providers have been covered during the Provider Census.

If the Provider Census suggests that there are very few providers in the market, then a full census of households might not be required; the research team can instead take a large sample. If, however, many providers operate in each market (as is the case in rural India and Cambodia), then sampling households is unlikely to lead to accurate estimates of providers' market shares. To see this, consider a village with 8 providers in its market, where one provider captures 25 percent of the market, and the remaining 7 providers evenly split the rest. It will be impossible to avoid small sample bias (not to mention imprecision) if we sample only 15 households (as is often the case in household surveys). Thus, the more providers we think have operating in the market, the greater is the need to complete a full household census.

Nevertheless, to complete the mapping process in as short of time as possible (recall, the ultimate goal is to sample relevant providers for the quality surveys), it is important to resist the temptation to ask households very detailed questions about their healthcare seeking strategies (you can sample a subset and do a longer survey with them later).

Table 1: Basic steps of a provider mapping

Step	Purpose	Development requirements	Time required per sampling unit	Relevant instruments
1. Participatory rural appraisal	Identify all locations where people seek care to define boundaries of the health care market	Most of the work is in the field.	½ day to 1 day	See PRA form and Master Code File from www.healthandeducationinindia.org.
2. Provider census	Identify all providers practicing in health care market and capture basic information Serves as universe to sample for quality surveys	Drafting of brief questionnaire, field-testing, and revisions.	1 to 2 days	Provider Census form

3. Household census	Match households to specific providers to calculate providers' market shares Double-check completeness of provider survey	Drafting of brief questionnaire, field-testing, and revisions.	2-3 days	Household Census form from www.healthandeducationinindia.org

INSTRUMENT 2: STANDARDIZED PATIENTS

When measuring quality of care through both unannounced standardized patients and medical vignettes, it is important to send the standardized patients into clinics first since the cases and their details are ideally identical across both instruments. If a provider is first administered the vignettes, (s)he could be primed for the cases when the standardized patients visit, which could either increase the likelihood that a standardized patient is detected or inflate our measure of quality.

This section outlines the basic steps in developing a standardized patient study. Chan et al (2012) and Siminoff et al (2011) provide more detail.

STEP 1: PERMISSIONS

Because of the deception design inherent in this type of study, the first step in developing this instrument would be to obtain clearance and support from the government and a local research ethics board. Not only will this ensure that all parties' interests are taken into account in the study design (e.g. the government will know that the identities of the providers under study will never be revealed to anyone outside of the study team), but it will also help during fieldwork if a provider detects a standardized patient and reacts in a hostile manner.

A close consideration of research ethics will guide the choice of three main study features – the health status of the standardized patients, the cases that will be simulated, and the type of consent obtained from providers. Some studies employ people who are actually suffering from the chosen case and who display symptoms at the time of the survey. The vast majority of studies, however, employ healthy individuals who will simulate the symptoms of the cases selected for the study. While real patients could increase the accuracy of case portrayal, the research team would also need to worry about the physical and mental fatigue associated with portraying their illnesses multiple times and the risk that this poses to their health.

The choice of cases to be simulated must consider clinic conditions and standards of care. Because we had observed that providers did not routinely disinfect their thermometers between patients in India, we selected cases that we thought were unlikely to trigger a temperature check. Similarly, we avoided presenting symptoms that might lead to a pelvic exam for our female standardized patients. In countries where patients presenting with diarrhea are typically given fluids intravenously, regardless of whether or not they need it, the research team might want to consider developing a proxy-case, where the standardized patient is presenting on behalf of an absent sick relative. In countries where the likelihood of injection is quite high regardless of the case, the research

team may want to reconsider the use of standardized patients or consider using prosthetics to simulate the injection site.

In most facility-based surveys, the research team obtains the informed consent of providers prior to administering their instrument. In this case, alerting providers to the possibility of a standardized patient visit could compromise the research design since it could increase the likelihood that the deception is detected, which is a frequent problem in many standardized patient studies (Rethans et al, 2007). Many research teams will seek consent and, without revealing the identity of the standardized patients or the details of their presenting symptoms, simply inform providers that a standardized patient will visit their clinic in the next few months. In emergency room settings, management must be informed of the exact days of these visits so that they can plan for additional staff on that day to ensure that the standardized patients do not take time that could be spent on real emergencies. In many of these studies, however, participation rates in the survey are predictably low – the majority of providers do not elect to receive standardized patients to assess the quality of their clinical practice.

An alternative to informed consent is a pure audit design, in which providers are not informed about the study at all. The United States Department of Health and Human Services, for example, will consider this kind of waiver of the informed consent principle when researchers are assessing the quality of public services. This is the approach we adopted in the study in rural India, after obtaining permission for the study from the central government and ethical clearance from a North American Internal Review Board.

STEP 2: CHOOSING CASES

Because providers may be more skilled at some cases than others, to get an accurate estimate of the quality of care they offer, they should receive multiple standardized patient cases. To keep survey development, recruitment, and fieldwork manageable, however, we found that the number of different cases selected for the survey should be no more than three.

To maximize accuracy and relevance and to minimize detection of the standardized patients, it is important to keep in mind three criteria when selecting the exact medical cases that the standardized patients will be portraying. First, and most obviously, the presenting symptoms must be easy to simulate by a healthy individual. Ideal cases would be those with no overt physical symptoms like non-specific acute lower back pain, for example, although it is not difficult to simulate the minor physical signs of cases such as a cold or asthma.

Second, since we do not want to put the standardized patient in an uncomfortable or dangerous situation in which (s)he has to reveal the deception, the cases should not trigger any invasive examinations or needle pricks or provoke undue anxiety in the provider. Standardized patients, for example, should not simulate a heart attack in front of the provider. This for certain would cause the provider unnecessary stress, not to mention tie up any emergency vehicles at his/her disposal.

Finally, the cases should be relevant for the context. Not only should incidence or prevalence of the condition be high, but there should be clear expectations for how the condition should be diagnosed and managed. It makes no sense, for example, to expect providers in remote resource-poor settings to diagnose rare forms of cancer or atypical presentations of stroke. Ideal cases would be those with established medical guidelines in the country under study. In India, for example, we chose cases in which the government had issued guidelines on what they expect public providers to do in rural health centers.

STEP 3: ENTRY PLAN INTO CLINICS

Before developing the content of the cases, it would be useful to first establish how the standardized patients will enter clinics sampled for the study. This aspect of the study design depends a great deal on context. In India, where health insurance schemes for the general public are still in their infancy and where appointments for primary care visits are rare, our standardized patients could simply walk into the clinics and expect to be seen by the provider. Because providers expect to know most of their patients in rural areas or are likely to ask patients they do not know about where they live, all of our standardized patients had to be "traveling through the area" and prepared to shoot the breeze with any curious people who might probe for additional details.

In other contexts, however, clinics might not accept walk-in patients or they may require their patients to demonstrate a certain type of insurance coverage. In these cases, researchers have relied on a number of strategies so that unannounced standardized patients could still enter sample clinics. Some have obtained fake insurance cards or ID cards (see, for example, Brown et al, 1998 and Burke et al, 2009). Others have registered the standardized patients as new patients and taken appointments, sometimes with the cooperation of the office manager (Rethans et al, 1991).

STEP 4: DRAFTING MEDICAL SCRIPTS

Once these preliminary steps have been completed, the substance of the work can begin. Each case needs a medical script, which is essentially an opening statement for the standardized patient and all of the responses – both verbal and physical – to any question or medical examination that the provider might pose to the standardized patient. These are the questions and examinations which the standardized patients must recall after their interactions with providers and which form a key component of the quality measurement. Thus, these scripts should also include all of the questions and examinations required to make a differential diagnosis.

The nature of the study's research objectives should determine what goes into the opening statement. Take the case of *cold*, for example. If the team wants to test whether or not the doctor would ask enough questions to distinguish between a virus and an infection, then they will need to leave out information about fever and body aches from the opening statement. At the same time, the doctor should take the case seriously and not wonder why a traveling patient cannot not wait until returning home to visit their normal doctor. In this case, a standardized patient presenting with a simple could say, *Doctor, for the past four days I haven't been feeling well. My throat is sore, and my nose is running*. Likewise for the *unstable angina* case in India, the opening statement did not include information about the radiation of the chest pain since that was something we wanted doctors to ask about. Instead, the standardized patient simply said, *Doctor, this morning I had a pain in my chest.*

A good place to start populating the list of questions and examinations that require standardized responses would be online differential diagnosis guidelines from medical schools, national guidelines, and doctors practicing both in a mature health system and in the context under study. This will likely lead to a rather lengthy list that then needs to be pared down to the most essential items since standardized patients cannot recall an infinite list of questions and examinations and since in practice most providers do not ask all of the questions they possibly could. At the same time, this list must also include the questions and exams that might be medically unessential for the case but that providers in the context under study are likely to ask or do.

A panel of doctors from a mature health system and the local area can help ensure that the scripts have the appropriate essential and unessential items and that the research team has identified the correct treatment protocol corresponding to the presenting symptoms and the responses prepared for the questions and exams in the scripts. For example, the team can request comments on the draft scripts and request the doctors to note if

any questions or exams essential for making a differential diagnosis have been omitted, if any questions or exams are unlikely to be completed in the primary care setting, and if any of the prepared responses should be modified to make the case more typical. This panel should also agree on what constitutes a correct diagnosis and treatment protocol.

This panel (especially if comprised of high quality doctors) may find it difficult to suggest unessential questions and exams. To populate the list of these items that nevertheless will require standardized responses, the research team will need to pilot the scripts as though they are medical vignettes with doctors from different regions of the study area who will not be a part of the sample. In short, this will be role-play exercise, where participating doctors are told the presenting symptoms of the patient and then requested to proceed with questioning and examinations as they would with a real patient. Any questions and examinations that the first panel neglected to mention but that multiple pilot doctors suggest should be added to the scripts, along with standardized responses. During this process in India, for example, we learned that our asthma patient was routinely being asked whether (s)he had a sore throat. Although this question was not essential for making a differential diagnosis, we needed to prepare a response that all asthma patients could use so that all doctors, regardless of the questions and examinations that they actually posed, had the potential to see an identical case.

STEP 5: RECRUITMENT

Once the cases and the medical content of the scripts have been finalized, the team can start the recruitment and screening process, which will continue throughout the training period. Given that standardized patients need very specific skills that are not very observable prior to training (even to the potential candidates themselves), the research team may need to initially recruit 2 to 4 times the number of people ultimately required. In order to be certain that idiosyncratic traits of the standardized patients are not driving the results, each case should be presented by more than one (ideally more than four) person.

While actors might be used to the idea of simulation and university students have flexible time schedules to make multiple visits to clinics, ordinary people recruited from the regions in the sample generally make the best unannounced standardized patients since the whole package - their appearance, style of speaking, and body language - would conform more naturally to providers' regular patient populations.

The cases themselves would imply a particular demographic background – for example, a middle-aged individual for a stable angina case - as would the entry plan into clinics. If, for example, this middle-aged person needs to be traveling through the area on work, then we might make the person a male construction worker. Regardless of the cases, however, all of the standardized patients should be healthy with no conditions that would complicate the case they would be presenting, should the provider detect any abnormalities while examining the patient (e.g. high blood pressure, diabetes, epilepsy, asthma, pregnancy.)

Specifying a minimum education threshold would also help ensure that the standardized patients understand the notions of simulation, standardization, and confidentiality required for the project. In India, we found that we could work with people who had completed their primary education.

Finally, and most importantly, the people recruited for this kind of exercise need to be trustworthy. They cannot reveal the cases or any of their details to providers in the sample, either directly or more indirectly by discussing the project with many people. They will also need to protect the identity of all the providers that they visit. Thus, it is probably best to work with people who are already familiar to the research team or survey firm organizing the work.

STEP 6: TRAINING

It is in the training phase of a standardized patient study that the difference between this kind of instrument and more common interview-based surveys is most obvious. First, the research team will need to assemble a team with particular skill sets to conduct the training sessions. This is not something that a random survey firm can carry out on its own. There are professional standardized patient trainers employed by most of the standardized patient programs in medical schools, who can be hired as consultants to draft and implement the training agenda. Given their years of experience, they can also advise candidates on ways to avoid medicines and invasive examinations in the clinic and to minimize detection when faced with unanticipated events. Equally important would be a doctor, especially a local doctor, who can provide continuous feedback on the accuracy of the case presentations and participate in mock-clinical encounters with the candidates.

Second, because this training not only imparts the skills required for survey work but also serves as the second stage of recruitment, it can last up to two weeks, and during this time, the standardized patient candidates will help develop the personas of the cases and the non-medical content of the scripts (e.g. what is your occupation, what did you eat for dinner last night, who are you visiting in the area), learn the appropriate body language for clinic entry and exit and for the examination stage of a clinical encounter, demonstrate and practice their ability to recall details of patient-provider interactions, standardize their presentations, and refine everything based on the evaluation of an external panel of doctors.

Mock interactions, in which SPs must perform their role as patients with someone else acting as the doctor, are the primary pedagogical tool during training, and the team can begin using them the very first day. They provide the best way for SPs to learn the details of the cases and, when a real doctor is portraying the role of the provider, for the research team to refine the medical content of the case. They also demonstrate areas of weakness to the training team, which can be used to screen the candidates further and isolate areas that require more practice. One common tendency of first-time standardized patients, for example, is to reveal too much information in response to a doctor's question. They become so excited about the character they are assuming and their mastery of the details of the case that they want to tell the provider everything about themselves, which could compromise the research team's ability to evaluate whether the provider is asking the right questions. If an asthma patient already reveals that her breathing difficulties are episodic rather than constant, then it will not be possible to determine if the provider would have asked that question. Another common consequence of the training is the use of medical language. In response to a question about pain location in an unstable angina case, for example, a response that uses the word "radiation" would surely arouse a provider's suspicion.

Video-recording equipment also plays a critical role in assessing candidates' ability to recall details of clinical interactions and in standardizing presentations. The accuracy of all of data collected using standardized patients depends on their ability to recall what happened in the clinic. While it might seem easy to recall all questions and examinations completed during a 10 minute interaction, this task is actually quite challenging in the context of a large audit study. First, there could be delays or distractions between the end of the clinical interaction and the time when details of the interaction are recorded on the exit interview form . Second, during real field work, standardized patients could be presenting their case to 25 to 30 different doctors whose questions and exams will not vary to a very large extent, and there is a danger that the details of each interaction could be confused with those of a previous one. A good way to test candidates' recall abilities in this context would be to record a mock-interaction in front of the candidates. Then when the camera is turned off, the candidates should observe another mock-interaction of the same case, which they as a group can critique and discuss. Afterwards, a written examination can test their ability to recall details of the first interaction (e.g. *Did the doctor ask about the location*

of the pain?), and the recorded interaction can serve as proof of how easy it is to forget questions and examinations and mix up two different interactions.

Recording just one aspect of the interaction – for example, the opening statement – for all standardized patients can also demonstrate how each portrayal differs.

After the team has fully developed the medical and non-medical aspects of the cases and screened candidates for their ability to consistently and accurately portray the scripts and recall clinical interactions, they should organize another panel of doctors unfamiliar with the study design and case scripts to validate the standardized patients' portrayals. In particular, these doctors should examine the patients as though they were real and provide feedback on what could make the presentations more realistic. If these doctors cannot diagnose the cases appropriately, then the cases must be reworked to make the presentations more straightforward. Once these weaknesses are addressed, the standardized patients can start making unannounced pilot visits. Any unanticipated questions or examinations that come up during the pilot must be flagged, and the group must work to standardize responses prior to fieldwork.

STEP 7: ASSIGNMENT OF CASES AND STANDARDIZED PATIENTS TO CLINICS

The number of cases that can be presented in each clinic will depend on context. A large clinic with multiple providers that rarely interact, for example, may be able to receive 6 standardized patients easily. In a smaller clinic, however, six different patients traveling through the area could be strange, even if each of the visits were separated by a couple of weeks. In India, each of our sampled providers saw three different cases. When a provider practiced out of two different clinics, however, he did not see the unstable angina case twice, since we thought that the rarity of such a case, combined with the traveling excuse, could increase the likelihood of detection.

While the training should ensure that the case presentations are standardized across all different standardized patients portraying the same case, inter-rater reliability – or the degree of concordance across two different standardized patients' portrayals and recollections of a case – could still be an issue. It might also be impossible to standardize all aspects of candidates such as their voices or attractiveness, which may or may not influence the way providers treat them. Randomized assignment of standardized patients to providers can mitigate these concerns and ensure that patient and provider characteristics are uncorrelated, so that, for example, most of the taller standardized patients are not seen older providers. Because the team will have multiple observations of the same standardized patients, they can also net out any systematic (fixed) variation that they observe for a particular patient in the analysis phase.

STEP 8: FIELDWORK AND DATA COLLECTION

The data from the clinical interactions between standardized patients and providers are recorded during a debriefing session that must occur within one hour of the interaction, in which a debriefer (senior enumerator) interviews the standardized patient and notes down on a structured questionnaire what exactly happened in the clinic. The information obtained through this structured debriefing interview constitutes the main data of a standardized patient study. At this time, the debriefer asks the standardized patient a series of questions (mostly with Yes/No responses) about the doctors' history questions, examinations, diagnoses, and treatment advice and their general impressions of the way they were treated by the doctors. Any results from diagnostic tests or prescription slips can be stapled to questionnaire to later confirm which tests were performed and the names and dosages of any prescribed medicines. Medicines dispensed in the clinic itself can be saved, and local pharmacists

can help identify the drugs if it is not obvious – for example, if individual pills are dispensed from opened packages. If the panel of doctors that advised on the medical scripts have specified an entire class of drugs in the cases' treatment protocols (for example, a blood thinner or an anti-platelet agent), then pharmacists can also help classify the medicines into these more analytically useful categories.

One may wonder why we need the extra debriefing step? Why can't the standardized patients themselves fill out this kind of form? First, since providers do not always use the same exact language as the doctors during training and could reveal a diagnosis in myriad ways, the standardized patients may need to discuss the interaction with someone else in order to sort out whether a certain question on the structured questionnaire was indeed asked and what exact diagnosis the provider made. For example, instead of asking an unstable angina patient if he has diabetes, the provider might ask about "sugar problems. " Or a provider might tell a patient presenting with a simple cold that he suspects his condition is a cold stemming from a bacterial infection. Although the doctor has said the magic word "cold," in this case, he has incorrectly diagnosed the patient with a bacterial infection.

Second, completing a visit as an incognito standardized patient can be a stressful experience. When they emerge from the clinic, standardized patients are likely to assume that they have been detected and may keep fretting about things they said during the interaction. Naturally, this state of mind is not conducive to accurate recall, so part of the debriefer's role is to give the standardized patient the sense that they are not alone in the field and to calm him/her down immediately after the interaction so that they can discuss its details in a relaxed frame of mind.

The debriefers require almost an equal amount of training as the standardized patients. Ideally, they would participate in the training from the very beginning so they could also learn the cases and fully understand the nature of the interactions that the standardized patients would be reporting on. Since there needs to be a sufficient amount of trust between the standardized patient and his/her debriefers, early involvement can also help lessen the impression that the debriefers are functioning as the standardized patients' supervisors.

STEP 9: ANALYSIS

The data that results from this exercise can provide a very rich description of the full clinical encounter from history-taking to diagnosis and treatment (see, for example, Das et al, 2012). Some variables can be summarized immediately without much post-coding. We can easily learn, for example, how much time a provider spends on average with a patient, the number of questions and examinations he completes, how these compare to what medical guidelines stipulate, and whether the provide offers the patient any information about the presenting illness or suggested treatment.

Other attributes, such as the correctness of the offered diagnosis or treatments, must be post-coded after fieldwork has been completed and the range of responses has been realized. Some of the coding will be straightforward. If a standardized patient presenting an unstable angina case receives a diagnosis of *Your liver is boiling*, for example, this can clearly be coded as an incorrect diagnosis. Likewise, if the provider treats this patient by waving a magic wand around his head, this is obviously not a correct treatment. Most of the time, however, the research team will need to rely on doctors and local pharmacists to accurately distinguish between correct and incorrect diagnoses and treatments.

For our work in India, we compiled all of the diagnoses uttered to our standardized patients for each case and asked a doctor in the United States to categorize them into three bins – correct, partially correct, and incorrect. We also needed to follow up with a local doctor to translate some of the more idiomatic expressions.

To deal with the thousands of medicine names that we found in the data, we had to follow a two-step process. First, we consulted doctors and management guidelines in the United States to determine which treatments and classes of medicines would be considered correct, helpful but not sufficient, and unnecessary/harmful. We sent them a compiled list all medicine names and asked them to identify the medicine classes of all of the ones that they recognized. We then attempted to classify the remaining medicines through internet searches. Next, we turned to local pharmacists to help us with the medicines that we still could not categorize.

Finally, the research team might want to summarize all of these quality indicators with one single index, especially if they want to rank providers on a distribution of quality – for example, if linking their pay to performance. One option would be an index, in which each question and examination receives equal weight or weights agreed upon by a panel of doctors. Such consensus, however, would be rare. An alternative strategy would be to use information from the data itself to determine the weights. Das and Hammer (2005), for example, score their vignettes using methods from Item Response Theory (IRT), in which each item is weighted by its degree of difficulty and its ability to discriminate among providers of different quality, as observed in the data, similar to the way standardized education tests are scored.

Table 2: Basic steps of a standardized patients survey

Step	Development requirements	Estimated time required	Relevant instruments or resources
1. Permissions	High Research ethics applications are lengthy, and a government may require considerable information before approving the study.	3 months Research ethics review boards often meet only on a bimonthly basis	
2. Choosing cases	Minimal	Minimal	
3. Entry plan into clinics	Medium This might need extensive piloting, issuance of fake identity cards, or collaboration with hospital or office managers	2+ months	
4. Drafting medical cases	High A panel of international and local doctors needs to give inputs on questions and examinations essential for making a differential diagnosis in an iterative process with the research team. Scripts should also be piloted with providers to populate list of non-essential but typical questions and examinations.	2 months	Medical Script for Standardized Patients from www.healthandeducationinindia.org Online differential diagnosis materials posted by medical schools

Step	Development requirements	Estimated time required	Relevant instruments or resources
5. Recruitment	Medium In addition to fitting demographic requirements of chosen cases, recruits need to be local and trustworthy with at least a primary education.	2 weeks Screening continues throughout training period.	Chan et al, 2012
6. Training	High Further screening of candidates, script development, validation, and piloting all occur at this stage and requires a team that includes a professional standardized patient trainer and at least one local doctor.	2 weeks	Chan et al, 2012 University of Toronto, Faculty of Medicine, Standardized Patient Program: http://www.spp.utoronto.ca
7. Assignment of cases and standardized patients to clinics	Minimal Assignment of standardized patients should be randomized to address potential concerns of inter-rater reliability.	2 days	
8. Fieldwork and data collection	Medium Clinical encounters are often shorter than most face-to-face surveys.	Depends on length of typical provider visit and geographic dispersion of sample clinics.	Exit form for Standardized Patients from www.healthandeducationinindia.org
10. Analysis	Medium Inputs required from doctors to categorize classes of treatments as correct and harmful and from pharmacists to sort medicines into these classes. Questions and examinations can be aggregated into an index that weights items by their ability to distinguish among providers	1 month	Das and Hammer, 2005 OpenIRT program for STATA: http://www.people.fas.harvard.edu/~tzajonc/openirt.html

INSTRUMENT 3: MEDICAL VIGNETTES

Since the role play of medical vignettes proceeds in the form of a structured interview, this instrument will provide a measure of providers' *knowledge* of differential diagnoses and case management, as opposed to the *practice* elicited by the standardized patients or provider observations, which would be a combination of knowledge and performance. To distinguish between knowledge deficits and underperformance, the cases presented by standardized patients should also feature in the roster of cases of the medical vignettes. The development of the

vignettes proceeds in an almost identical way, and much of the work for the overlapping cases would already have been completed when developing the standardized patient cases.

STEP 1: CHOOSING CASES

The same basic criteria that governed the choice of standardized patient cases apply to medical vignettes as well (Step 2 of Instrument 2). Because the vignettes cases do not require individuals to simulate them, however, they are not bound by some of the restrictions. Thus, conditions such as infection diseases and injuries can be developed as medical vignettes. Likewise, since each vignette takes around 15 minutes to administer and since all cases can be completed in one interview, a vignettes study can implement up to 5 different cases per provider.

STEP 2: DEVELOPING MEDICAL SCRIPTS

Script development for medical vignettes is identical to the process required for standardized patients (Step 4 of Instrument 2). In this case, however, the research team must also prepare responses for any examinations that the providers might suggest. For example, if the provider suggests that the patient gets blood work, an enumerator will reveal the results of the blood test so that the provider can use that information to diagnose and treat the case. Similarly, if the provider requests an x-ray or ECG, then the team could bring an x-ray or ECG reading to the interview and let the provider interpret it (of course, all providers must see the same x-ray or report).

As with cases used with standardized patients, if doctors known to be high quality to the research team cannot make an appropriate diagnosis or if they suggest alternative treatment options at the pilot stage, then the scripts must be reworked to make them more straightforward, classic presentations of the cases.

STEP 3: RECRUITMENT AND TRAINING

The implementation of the vignettes instrument resembles a typical survey interview more than the standardized patient visits, but it is important that the team selects only the best enumerators for this task. First, recording some responses – even though, they are typically of the form “Yes (or no), the provider asked about this” – requires some judgment. If the team has prepared a response for “Is there pain radiation,” in an unstable angina case, for example, then the enumerator must be able to recognize a slightly modified version of the same question in “Does the pain spread to the arm?” One might then think that only enumerators with medical backgrounds should participate in this exercise, but this leads to the second reason for selecting only the best and most experienced survey enumerators. The provider, especially if (s)he is unlicensed with limited medical training and/or formal education, must feel comfortable during the interview. In particular, the interview should not feel like an evaluation. Otherwise, providers might underperform, and the vignettes will eventually measure a combination of knowledge and test-taking skills. In the India study, we used the provider mapping exercise to vet enumerators based on their interview styles and error rates in completing questionnaires. Because these same enumerators had administered the short facility survey during the Provider Census a few months earlier, they already had established some rapport with the providers.

Training for this component of a quality study is also more intensive than what is required for a typical survey since during the interview, the enumerators must engage in role-play to create the sense of a real patient-provider interaction. To do this properly, they essentially need to memorize the medical scripts so that they are not looking down at the questionnaire when the provider asks a question or recommends an examination. The enumerator playing the patient also needs to coordinate well with the enumerator playing the omniscient narrator. As with the standardized patients, mock interactions and pilots are the best way to develop these skills.

STEP 4: FIELDWORK AND DATA COLLECTION

Although it mostly depends on how thorough the provider is or how long (s)he takes to consider the case, the average time required for each vignette is approximately 15 minutes. Thus, administering 5 vignettes could take an hour and a half. To avoid inconveniencing their patients and potentially embarrassing the provider, this interview would ideally take place when the clinic is not open. If the research team already sent in standardized patients presenting the same cases and is worried that they had been detected, it can use also this time to ask providers to describe any patients they had recently seen with similar symptoms and to furnish any documentation (e.g. charts or log book) that indicates the time and date of the visit.

The data of medical vignettes consists of indicators for whether or not a question was asked or an examination was suggested, the provider's diagnosis, and a list of the exact treatments that the provider advises. Enumerators can probe providers further about any medicine that does not have a clear label and ask a local pharmacist to identify its medicine class if the provider does not have this information.

STEP 5: ANALYSIS

The data that results from this exercise can be analyzed in exactly the same way as the standardized patients data (Step 9 of Instrument 2). In this case, however, the interpretation of the data naturally will be different, as the medical vignettes are not capturing actual practice, but rather provider knowledge.

Table 3: Basic steps of a medical vignettes survey

Step	Development requirements	Estimated time required	Relevant instruments or resources
1. Choosing cases	Minimal	Minimal	
2. Drafting medical cases	High A panel of international and local doctors needs to give inputs on questions and examinations essential for making a differential diagnosis in an iterative process with the research team. Scripts should also be piloted with providers to populate list of non-essential but typical questions and examinations.	2 months	Online differential diagnosis materials posted by medical schools
3. Recruitment	Minimal Only very experienced, high quality enumerators should be used.	Minimal	

Step	Development requirements	Estimated time required	Relevant instruments or resources
4. Training	High Enumerators must memorize questionnaire and must be able to coordinate across different roles (patient and omniscient narrator). They must also be skilled in developing an appropriate rapport with providers.	2 weeks	Training videos, posted on www.healthandeducationinindia.org
5. Fieldwork and data collection		Each vignette takes approximately 15 minutes to administer. Total field time also depends on geographic dispersion of sample clinics.	Vignettes Form from www.healthandeducationinindia.org Vignettes manual, posted on www.healthandeducationinindia.org
10. Analysis	Medium Inputs required from doctors to categorize classes of treatments as correct and harmful and from pharmacists to sort medicines into these classes. Questions and examinations can be aggregated into an index that weights items by their ability to distinguish among providers	1 month	Das and Hammer (2005) OpenIRT program for STATA: http://www.people.fas.harvard.edu/~tzajonc/openirt.html

INSTRUMENT 4: PROVIDER OBSERVATIONS

Provider observation is another data collection method that does not proceed as a structure d interview. In fact, survey enumerators should seek to be as unobtrusive as possible so that both providers and their patients forget about their presence. To establish this kind of indifference in the midst of observation, they should never speak with patients and limit their clarification questions to providers to the time between patients (for example, to ask the exact names and doses of medicines prescribed or dispensed in the clinics). The research team can also help with this by carefully designing their instrument and field protocols to minimize distractions and maximize the accuracy of the information that gets recorded.

STEP 1: PLAN FOR INFORMED CONSENT

Providers must obviously consent to having an observer in their clinic for an entire day. What about their patients? Almost all guidelines on the ethics of human subjects research would require that they provide their consent as well. The research team then needs to decide how to obtain this consent without disrupting the provider's practice, especially in situations where the typical clinical encounter does not last longer than 5 minutes. One option, which we used in India, would be to rely on the provider to mention the study to each patient and explain that they can request the enumerators to leave the room at any time during their visit. An alternative strategy would be to approach patients before they enter the clinic or before they are seen by the provider.

STEP 2: DRAFTING THE DATA COLLECTION FORM

Because enumerators will have to note down aspects of the clinical encounter as they occur in real time, they should not have to flip through pages of a questionnaire. Rather, they should be able to document what is going on without looking down at all. Thus, the data collection form for provider observations is ideally no longer than one page.

What exactly should go on this page? The overall structure of the interaction – history taking and examination, diagnosis, and treatment –can provide some guidance. Context, however, will determine what other important variation needs to be captured. In India, for example, we always documented whether or not an injection was administered with a sterile needle. In a place like rural Cambodia, one might want to note whether the provider injected the patient with solution taken out of an IV-drip bag or whether blood for a glucose measurement was drawn intravenously – not because these are general traits of any interaction but because these are commonly observed errors that we would want to quantify. The only way to get a sense of this before data collection is to make field visits and sit in a number of clinics for several days to observe the interactions between providers and their patients.

STEP 3: RECRUITMENT AND TRAINING

Because of the speed with which enumerators must document features of the interaction, the research team should once again choose only the best enumerators for this exercise. Some studies have employed individuals with a health background (Leonard and Masatu, 2005), while others have relied on survey enumerators (Das and Hammer, 2007) who have been trained to recognize history questions and physical examinations, such as checks of temperature, pulse, blood pressure, and to report on the sterility of any needles used.

Either type of enumerator will have to memorize the data collection form and practice intensively with mock interactions during training. In India, for example, we recorded real interactions between providers and their patients in clinics and asked enumerators to fill out the data collection forms while watching the video. This gave them a sense of the concentration and speed required to record real clinical encounters. Relative to other types of surveys, provider observations will also require much more piloting and training in the field, not just to make enumerators faster and more accurate in their observations but also to fine tune issues such as the optimal place to sit in the clinics to minimize disruptions and properly timing clarification questions about medicine names.

STEP 4: FIELDWORK AND DATA COLLECTION

Prior to sending enumerators into the field, the research team must first decide how long they are going to sit in each clinic. In some contexts – for example, the rural areas of India and Cambodia where there are many providers in market, each with around 10 to 15 patients per day, the enumerators may have to observe all day. In

environments with larger patient volumes, the research team may want to send in the enumerators only for a couple of hours. It would be important, though, to randomly choose the time slots (within normal clinic hours) for each provider in case quality of care differs throughout the day. Thus, if normal clinic hours are in the morning and evening, then the research team might want to randomly choose half of their clinics for provider observations in the morning and send enumerators to the remaining half in the evening.

If provider observations and patient exit interviews are being implemented in tandem, then the team will need to use tokens to track patients so that their data can be matched across instruments. At the end of each clinical encounter, the enumerator can give the patient a token with a number written on it and record this number alongside the patient ID. The patient can then give the token to the enumerator outside administering the patient exit interviews, who will also record the token number alongside the patient ID on the exit questionnaire.

Table 4: Basic steps of participation observations

Step	Development requirements	Estimated time required	Relevant instruments or resources
1. Plan for informed consent	Minimal	Minimal	Local research ethics board or medical ethics committee can provide guidance.
2. Drafting the data collection form	High Form should be one page and capture all important variation in clinical encounters. Research team should observe interactions in clinics prior to drafting form.	1 month	Participant Observation Form from www.healthandeducationinindia.org
3. Recruitment and training	High Only very experienced, high quality enumerators should be used. Enumerators must memorize data collection form. Training consists of mock interactions and multiple field visits.	1-2 weeks of training	Training videos, posted on www.healthandeducationinindia.org
4. Fieldwork and data collection		Depends on typical patient load. Enumerators might have to sit all day or for part of the day.	

INSTRUMENT 5: PATIENT EXIT INTERVIEWS

Relative to the other four instruments, patient exit interviews more resemble typical face-to-face surveys. In this case, however, the interview will be much shorter than a normal survey since patients most likely have somewhere else to go once they exit the clinic and did not factor an exit interview into their schedule. While the team could choose to interview only the patients that do have plenty of time to chat about their health and their interaction with the provider, this sample would lead to a biased perspective of the providers' client pool. Thus, as in the provider observations, the challenge is drafting a data collection form for the patient exit interviews that is brief yet deep enough to capture important variation. Again, this requires extensive field-testing.

It is also possible to follow-up with patients afterwards in their homes. The information that they provide, however, may differ from what they would reveal immediately after their interaction outside of the clinic. In Paraguay, for example, patient satisfaction elicited in patient exit interviews was higher than what patients reported in their homes up to one week later (Das and Sohnesen, 2006).

If the patients interviewed have also been observed in the clinic, then enumerators must coordinate with the enumerators of provider observations so that patients can be matched across instruments (Step 4 of Instrument 4).

SUMMARY

Table 5 presents the main instruments used for measuring quality of care within clinics, summarizing their strengths and weaknesses and the resource requirements for their development and implementation. As the examples from Part 1 demonstrated, this kind of measurement can yield valuable insights on what actually happens in clinics that gives us a different picture of quality than measurements of infrastructure, medical supplies, and provider qualifications. As we saw in Part 2, developing instruments for this kind of measurement is feasible, though they may require research teams to consult medical professionals for guidance and to train survey enumerators more intensively than for standard facility or household surveys.

Table 5: Summary of quality instruments

Method	What is measured	Advantages	Limitations	Development requirements	Personnel Requirements	Time requirements for fieldwork
PROVIDER MAPPING	Number of distinct providers and clinics operating in health care market and their market shares	Allows for representative sampling of providers, especially in presence of large informal sector	None	*Low*	*Low*	*Medium* Low in rural areas (1 day). Could be time consuming in urban, denser markets.
STANDARDIZED PATIENTS	Provider practice (knowledge + performance)	Standardization of case and patient mix, allowing valid comparisons of different doctors and clinics Can assess practice with rare but serious illnesses Free from observation bias Diagnoses known, so care can be benchmarked against medical guidelines	Limited set of illness conditions that can be portrayed to maintain accuracy and to protect enumerator safety Often measuring quality of care of first time patients Providers often do not articulate diagnoses	*High* Need to identify set of questions and examination necessary for a proper differential diagnosis. Requires standardized responses for every possible question and examination posed to patient, including medically irrelevant questions. Often requires	*Medium* Training (2 weeks) is key for ensuring accuracy and consistency of case presentation and for limited detection of standardized patients. Ordinary people (not actors) perform very well.	*Low* Depends on length of typical clinical encounter, often very short in both middle- and low-income settings.

Method	What is measured	Advantages	Limitations	Development requirements	Personnel Requirements	Time requirements for fieldwork
				clearance from a research ethics board.		
PROVIDER OBSERVATIONS (OR DIRECT CLINICAL OBSERVATIONS)	Provider practice (knowledge + performance)	Can document provider behavior with real patients Can characterize patient mix	Case and patient mixes differ across doctors Actual diagnoses unknown, so often difficult to assess adherence to medical guidelines Low frequency of more serious illnesses Providers rarely articulate diagnoses Potential for observation bias	*Medium* All captured variation should fit on a single page.	*High* Need high ability enumerators. They must be able to record all information of quick clinical encounters without interruptions or requests for clarification. More than a week of training might be required.	*High* Ideally enumerators sit with provider all day or for part of a day and record all patient visits.
PATIENT EXIT INTERVIEWS	Provider practice (knowledge + performance), patient background, and patient satisfaction	Can measure patient perceptions of quality and compare them to measures obtained during provider observations Can characterize patient mix	High potential for recall bias Patients unable to recall or evaluate providers' technical capacity Actual diagnoses unknown, so often difficult to assess adherence to medical guidelines Providers rarely articulate diagnoses	*Low*	*Low*	Should not be more than 10 minutes to ensure that all patients, rather than a chatty non-random sample, are interviewed.

Method	What is measured	Advantages	Limitations	Development requirements	Personnel Requirements	Time requirements for fieldwork
			Patients might feel inhibited to speak about provider outside of the clinic			
MEDICAL VIGNETTES	Provider knowledge	Standardization of case and patient mix, allowing valid comparisons of different doctors and clinics Unrestricted set of illness conditions that can be used Can obtain providers' diagnoses.	Does not capture what providers would do when faced with case presented by real patients Could underestimate quality if providers nervous about evaluation	*High* Need to identify set of questions and examination necessary for a proper differential diagnosis. Requires standardized responses for every possible question and examination posed to patient, including medically irrelevant questions.	*High* Need high ability enumerators. They must be able to record all information of quick clinical encounters without interruptions or requests for clarification. More than a week of training might be required.	*Approximately 15 minutes per vignette.*
CHART ABSTRACTION (OR RECORD REVIEW)	Provider practice	Information about real patients, including providers diagnoses, which they often do not articulate.	Charts, when they exist, are often incomplete, even in high income settings.	*Low*	*High* Requires local person with medical background who	*Medium* Need to scan all records.

Method	What is measured	Advantages	Limitations	Development requirements	Personnel Requirements	Time requirements for fieldwork
					can read and evaluate charts.	

BIBLIOGRAPHY

Basinga, P., P.J. Gertler, A. Binagwaho, A.L. Soucat, J. Sturdy, and C.M. Vermeersch. "Effect on maternal and child health services in Rwanda of payment to primary health-care providers for performance: an impact evaluation." *Lancet* 377, no. 9775 (2011): 1421-8.

Chan, Brian, Veena Das, Manoj Mohanan, Diana Tabak, Alaka Holla, and Jishnu Das. "Standardized Patients and the Measurement of Healthcare Quality: Field Guide, Manual, and Sample Instruments." Medical Advice, Quality, and Availability in Rural India (MAQARI), 2012.

Das, Jishnu and Jeffrey Hammer. "Money for Nothing: The Dire Straits of Medical Practice in India." *Journal of Development Economics*, 2007: 1-36.

Das, Jishnu and Jeffrey Hammer. "Which Doctor: Combining Vignettes and Item Response to Measure Doctor Quality." *Journal of Development Economics* 78 (2005): 348-83.

Das, Jishnu and Paul Gertler. "Variations in Practice Quality in Five Low-Income Countries: A Conceptual Overview." *Health Affairs* 26, no. 3 (2007): w296-w309.

Das, Jishnu and Thomas Pave Sohnesen. "Patient satisfaction, doctor effort, and interview location: evidence from Paraguay." *Policy Research Working Paper* (World Bank), 2006.

Das, Jishnu, Alaka Holla, Veena Das, Manoj Mohanan, Diana Tabak, and Brian Chan. "Standardized Patient Study Reveals Severe Quality Deficits in Rural and Urban India." *Health Affairs*, Forthcoming.

Das, Jishnu, Jeffrey Hammer, and Carolina Sanchez. "The Impact of Recall Periods on Reported Morbidity and Health Seeking Behavior." *Journal of Development Economics* 98, no. 1 (2012): 76-89.

Grad, R. Tamblyn, R., McLeod, P.J., et al. "Does knowledge of drug prescribing predict drug managment of standardized patients in office practice?" *Medical Education* 31 (1997): 132-137.

Hutchinson, B., Woodward, C.A., Norman, G.R., Abelson, J., Brown, J.A. "Provision of preventive care to standardized patients." *CMAJ* 158, no. 2 (1998): 185-193.

Igun, U.A. "Reported and actual prescription of oral rehydration therapy for childhood diarrhoeas by retail pharmacists in Nigeria." *Social Science and Medicine* 39, no. 6 (1994): 797-806.

Leonard, Kenneth and Melkiory Masatu. "The use of direct clinical observation and vignettes for health services quality evaluation in developing countries." *Social Science and Medicine*, 2005: 1944-1951.

Lundberg, Mattias. "Client Satisfaction and the Perceived Quality of Primary Health Care in Uganda." In *Are You Being Served: New Tools for Measuring Service Delivery*, edited by Jishnu Das, and Markus Goldstein Samia Amin. World Bank, 2008.

Madden, J.M., J.D. Quick, D. Ross-Degnan, K.K. Kafle. "Undercover Careseekers: Simulated Clients in the Study of Health Provider Behavior in Developing Countries." *Social Science and Medicine* 45, no. 10 (1997): 1465-1482.

McGlynn, Elizabeth A., Steven M. Asch, John Adams, Joan Keesey, Jennifer Hicks, Alison DeCristofaro, and Eve A. Kerr. "The Quality of Health Care Delivered to Adults in the United States." *New England Journa of Medicine*, no. 238 (2003): 2635-45.

Ozuah, P.O. and M. Reznik. "Residents' asthma communication skills in announced versus unannounced standardized patient exercises." *Ambul.Pediatr* 7, no. 6 (2007): 445-448.

Peabody, John W., Jeff Luck, Peter Glassman, Timothy Dresselhaus, and Martin Lee. "Comparison of Vignettes, Standardized Patients, and Chart Abstraction: A Prospective Validation Study of 3 Methods for Measuring Quality." *Journal of American Medical Association* 283, no. 13 (200): 1715-1722.

Peabody, John W., Jeff Luck, Sharad Jain, Dan Bertenthal, and Peter Glassman. "Assessing the Accuracy of Administrative Data in Health Information Systems." *Medical Care* 42, no. 11 (2004): 1066-1072.

Rethans, Jan-Joos, Ferd Sturmans, Riet Drop, Cees van der Vleuten, and Pie Hobus. "Does competence of general practitioners predict their performance? Comparison between examination setting and actual practice." *British Medical Journal*, 1991: 1377-80.

Russel, N.K. "Using Unannounced Standardized Patients to Evaluate Sexual Risk Assessment And Risk Reducation Skills of Practicing Physicians." *Acad. Med.* 66 (1991): 37-39.

Schulman, K.A., J.A. Berlin, W. Harless, J.F. Kerner, S. Sistrunk, B.J. Gersh, R. Dube, C.K. Taleghani, J.E. Burke, S. Williams, J.M. Eisenberg, J.J. Escarce. "The effect of race and sex on physicians' recommendations for cardiac catheterization." *The New England Journal of Medicine* 340, no. 8 (1999): 618-626.

Thamlikitkul, V. "A correlation of clinical performance on a written test and a standardized patient." *Journal of the Medical Association of Thailand*, November 1991: 513-517.

World Bank. *Healthy Partnerships: How Governments Can Engage the Private Sector to Improve Health in Africa.* Washington, DC: The World Bank, 2011.