

Quality of Primary Care in Low-Income Countries: Facts and Economics

Jishnu Das^{1,2} and Jeffrey Hammer³

¹World Bank, Washington, DC 20433; email: jdas1@worldbank.org

²Centre for Policy Research, New Delhi 110021, India

³Woodrow Wilson School of Public and International Affairs, Princeton University, Princeton, New Jersey 08544; email: jhammer@princeton.edu

Annu. Rev. Econ. 2014. 6:525–53

The *Annual Review of Economics* is online at economics.annualreviews.org

This article's doi:
10.1146/annurev-economics-080213-041350

Copyright © 2014 by Annual Reviews.
All rights reserved

JEL codes: I10, I11, I15, L15

Keywords

health markets, credence goods, audit studies, public and private care

Abstract

New research on the quality of care in public and private primary care facilities has significantly enriched our understanding of how health care is delivered in low- and middle-income countries. First, this article summarizes recent advances in the measurement of quality, distinguishing between measurements of provider knowledge and provider effort. Second, it looks at the determinants of practice quality variation in low-income settings, highlighting the limited role of structural constraints such as infrastructure, the supply of materials including drugs, and provider training—the mainstay of much of global health policy today. In contrast, practice quality variation is clearly linked to provider effort, an aspect of provider behavior that can be altered through a variety of means. Third, it provides a broad economic framework to interpret the findings. We look for evidence of specific market failures in the provision of primary care and emphasize that the key difficulty is (and always was) the transaction-specific nature of medical advice. Providers can do too much or too little (or both), and the extent of either depends on the specific patient and the specific disease. We document specific ways in which it is difficult for both consumers and governments to monitor every transaction to detect potentially errant behavior.

1. INTRODUCTION

Recent years have seen an upsurge in interest in global health. This is to be applauded, as the health and well-being of poor people in poor countries clearly have a solid claim as a moral imperative. However, with progress on some of the challenges facing improved health worldwide, harder second-generation problems linked to the quality, rather than the availability, of care have emerged. These form the focus of our review. We summarize new research from primary care settings in low-income countries that helps us understand how quality is produced, how it is valued in the market, and how it can be improved.¹

But first, we provide some history. The Alma-Ata Declaration of 1978 emphasized two broad categories of policies necessary for adequate health coverage of the people in poor countries. One was basic preventive and promotive health services, including safe water and improved sanitation. The second was the extension of primary curative medical care to achieve universal access. As to the first, it can be argued that budgets for basic prevention measures have often been given short shrift relative to those for medical services. It is always worth noting that real public goods, nonexcludable and nonrival, and traditional public health interventions, many of which were handled by currently rich countries by the early twentieth century with significant mortality impacts, have yet to be extended to most citizens of the currently poor world (see Cutler & Miller 2005, Cutler et al. 2006).

As to primary curative care, initial efforts were to make sure that curative care was accessible to all citizens, particularly in rural areas. Increased access to medical care has meant more medical providers; fewer barriers from, for example, fees charged at clinics; and, as a nod to economics, more cost-effective interventions. We argue in this article that, perhaps as a consequence of this focus, access is not the main problem for many poor people anymore. However, it is increasingly evident that access to quality remains a serious issue, with severe deficiencies in both the public and the private sectors.² The nature of these deficiencies and the determinants of quality variation are two major themes of this article.

This transition from access to access with quality in the health sector mirrors that in education. Just as an emphasis on ensuring that school enrollment in poor countries is increased has given way to measuring and concentrating attention on achieving more learning by children, so has the emphasis in health shifted: from measuring and ensuring access alone to asking if medical encounters are of high-enough quality to be effective in improving health. Similarly, as we document below, the focus has started to shift from purely technical solutions (more construction of schools or better curriculum, for example, in education; specific cost-effective medical interventions in health) to the behavior of the providers needed to make sure the technical components have any traction at all.

A third theme of this article looks at the rewards to quality and therefore the incentives to invest in improving quality. In most sectors of the economy, discussions of appropriate policy start by defining the specific market failures and, from there, try to find the appropriate interventions. Unfortunately, although the private sector is larger than the public sector in many of the settings we study, the nature of equilibrium—the determination of prices, quantity, and quality—is not

¹ A comprehensive look at health policy would include much more than medical care and certainly more than primary medical care. Attacking the main market failures in global health would require much more attention to basic, preventive, public health, on the one hand, and protection from major catastrophic financial loss due to endemic insurance market failure, on the other. These should be considered the opportunity cost of all policies mentioned in this article but are not the subject here.

² This discussion on quality comes early to low-income countries. In the United States, for instance, the Institute of Medicine's report on patient safety spurred significant institutional responses and new research, but this was only in 2000 (Kohn et al. 2000).

well understood.³ Therefore, it is incumbent on us to ask more basic questions concerning these characteristics of equilibrium, the welfare losses of distortions in private markets, and the ability of governments to improve on the entire market's functioning given its own informational constraints. In brief, how bad is medical care, significant fractions of which are in the private sector; how bad is the welfare loss of these market failures; and what policy levers are available to fix them?

The question is relevant not only for the low-income countries that we study, but also for our broader understanding of how medical markets work. Specifically, these environments are examples of medical markets with little *de facto* regulation, insurance, or administrative price setting. Understanding pricing, quality, and provider behavior in these relatively free markets thus helps isolate failures (or the lack thereof), which is harder to do in the controlled and insurance-based settings commonly found in high-income countries.

The article proceeds as follows: We first present some basic facts about access to medical treatment in low-income countries, finding it to be widespread for many poor populations around the world. Second, we discuss problems of defining and measuring the quality of care and cull some empirical generalizations from the recent literature. These relate both to the level of quality in public and private primary clinics (as a preview, it is not good) and to the main constraints to improving quality. Third, we discuss hypotheses that might explain some of these generalizations and anomalies. Little is known about markets for health care in poor countries and the market failures that characterize them such that the quality of medical care emerges as a consequence of the incentives in both private markets and public facilities. We suggest elements of theory that might contribute to an explanation, trying to stay as close to relevant empirical work as possible. We conclude with a brief discussion of the policy interventions that may be feasible given both the behavior of providers (and patients) and the information constraints under which a public regulator or delivery system operates.

2. HEALTH MARKETS IN LOW-INCOME COUNTRIES

One concern in the health literature is that there is too little access to medical care in low-income countries. For instance, Chen et al. (2004) report on human resources for health, sounding the alarm for countries with (supposedly) less than 2.5 health workers per 1,000 population. Besides the arbitrariness of the particular number chosen, as economists, we worry about whether such ratio policies can be used to determine optimal policy in a meaningful manner; after all, there is no market in the economy in which we prejudge what the right ratio of sellers to buyers should be. But by simply looking at some data, we can see that the point that there are too few providers is a little hard to maintain.⁴

That there is considerable access to providers can be examined from both sides of the market: How often patients visit providers and how many providers there are can be independently observed in areas where poor people live. Poor people seek medical care often. The Demographic and

³Although there is variation across regions, much of care seeking in low- and middle-income countries is in the private sector. Montagu (2010) provides readily accessible regional and country-level analysis of care seeking (defined as health seeking for children below age 5 in the three months prior to the survey) using DHS from multiple countries. These data show that the private sector accounts for close to 80% of care in South Asia, 66.3% in Southeast Asia, 60.1% in North Africa and Europe, 50.8% in sub-Saharan Africa, and 33.7% in Latin America and the Caribbean. The income elasticity of private sector use differs across regions. For instance, in South Asia, 79.8% of the poorest and 85.4% of the richest quintiles seek care in the private sector relative to 22.5% of the poorest and 61.3% of the richest quintiles in Latin America and the Caribbean.

⁴The term "some" in the phrase "some data" should be taken literally: Similar to the quality of care, there are few systematic data over time on the use of the private sector in most low- and middle-income countries.

Health Surveys (DHS) from around the world provide a snapshot of visits to health care providers for children under the age of 5 for two tracer conditions—acute respiratory infections and diarrhea. These data suggest that, in 41 out of 70 countries, more than 50% of children with acute respiratory infections or diarrhea sought care from a health provider and, furthermore, that the likelihood of seeking care was not very sensitive to asset ownership. One way to assess whether this is “often” would be to compare it to high-income countries. For instance, in the United States, data from the 1988 National Medical Expenditure Survey show that 52% of children seek care at a health facility when they are sick with pharyngitis (throat infection); among the uninsured, the rate is 32%. The rough comparability of the numbers from low-income countries to US data, in conjunction with the fact that the rural health facility usage rates in these countries are within 80% of the urban rates, suggests that access to health care may be more widespread than usually imagined (data compiled from MEASURE DHS 2013).

Detailed surveys on health care utilization paint a similar picture. For instance, people in rural Rajasthan, a low-income and low-density state in India, visit a doctor about six times a year (Banerjee et al. 2004). In urban India, individuals visit doctors about five times a year (Das & Sánchez 2003). This usage of health facilities in India—both in a relatively rich urban sample and in a relatively poor rural sample—is higher than the US average of 3.44 visits per person per year (data from Natl. Cent. Health Stat. 2009). Results are similar in other low-income countries. From Burkina Faso [purchasing power parity adjusted gross national income (GNI) or GNI per capita of \$1,510 in 2012] to Thailand (purchasing power parity adjusted GNI per capita of \$9,430 in 2012), health care utilization is high among the population in general, and even among the poorest quintiles (Makinen et al. 2000). Although it is well known that expenditures on medical care are highly elastic with respect to income (the elasticity of expenditure on health care with respect to income tends to cluster around 1.5 in surprisingly disparate studies), the number of visits is not.⁵

These numbers themselves may be lower bounds on the use of medical care in poor populations owing to the significant effects of recall periods on health seeking. Decreasing the recall period from one month to one week in an experimental setting in Delhi, India, sharply increased the number of doctor visits that households reported. This recall bias was correlated with income, with larger increases among the poor. In fact, with weekly recall periods, the poor sought health care more than the rich both unconditional and conditional on reporting an illness, whereas in monthly recall surveys, the use of health care providers increased with income (Das et al. 2012a).

From the other side of the market, a few studies map the number of providers in a village or urban neighborhood, highlighting the complexity of health markets in low-income countries in terms of the choices that households face.⁶ For example, our previous work in Delhi (urban India) shows that in seven neighborhoods studied, a 15-min walking radius surrounding the neighborhood yields an average of 75 providers of all stripes available to patients (Das & Hammer 2005).

Even in India’s rural areas, there are plenty of people willing to offer services. The Medical Advice Quality and Availability in Rural India (MAQARI) project counted the numbers and types of providers in representative village samples for 19 Indian states, covering more than 90% of its

⁵There are regions where access remains a significant problem. In countries such as Chad, Mali, and Togo (all in francophone West Africa), less than 25% of children with acute respiratory infections or diarrhea are taken to a health facility for treatment (data compiled from MEASURE DHS 2013). This could be because these countries have highly dispersed, low-density populations that are difficult for any provider to reach. That people in these countries do not visit anyone, even local traditional healers, is an interesting phenomenon but not one that is common worldwide.

⁶These studies are different from more standard surveys in which households are asked how far they live from the health facility. In these more standard surveys, the number of health facilities available to communities cannot be ascertained.

rural population. In addition, in the state of Madhya Pradesh—one of India’s poorest states, with low levels of education and health outcomes—the study mapped providers in the village and its surrounding health market. There were several noteworthy results.⁷

In Madhya Pradesh, an average village has 0.5 public doctors but also a wide variety of other providers, a fact that becomes apparent once we include the relevant market area (often quite clear in context) in the mapping. For instance, one village had two public providers and one private one, but villagers also sought care from the larger settlement located on the state highway two miles away. In this larger settlement, there were more than 70 providers, ranging from fully qualified MBBS doctors (the standard medical degree, equivalent to an MD in the United States) to people trained in various traditional medical systems such as Ayurvedic medicine and several people with no formal training at all. However, all providers were found to dispense allopathic medicine (including prescription-only antibiotics and steroids) and carried many common drugs.

Statewide, the study identified 1,190 providers across the health markets of 100 villages. Of these, 34% (347) were chemists or government functionaries in nursing roles (auxiliary nurse midwives, Accredited Social Health Activists, or multipurpose health workers). Among the 653 primary care providers who self-identified as doctors, a small fraction (just over 10%) reported an MBBS degree, and one-third reported a number of alternate degrees, including Ayurveda, Unani, and homeopathy degrees. The single largest category was those without any medical qualifications, with 5.4 such providers in the health market of the average village.

Neither was Madhya Pradesh an exception: Provisional numbers from the all-India study that counted providers within villages (excluding those in wider markets possibly just outside the village) showed an average of 4.4 providers for every village, 3.4 of whom had no medical training, 0.8 had some degree, and 0.18 were MBBS doctors (Cent. Policy Res. 2011, based on data from the MAQARI project).

Few studies of this nature have been done outside India, but those few find similar numbers and structures. Makinen et al. (2011) map formal health facilities in seven districts of Ghana. Across the seven districts, they located 765 facilities, 47% of which were “chemical sellers” and 22% were retail pharmacies. Clinics, health centers, hospitals, and maternity homes together accounted for another 29% of health facilities, with the majority in urban areas.

Sudhinaret et al. (2013) look at the role of informal health care providers in low-income countries, reviewing 334 references between 2000 and 2011. First, their review suggests that the percentage of health care provided by informal providers is consistently high across many countries, accounting, for instance, for 65–77% of care seeking in Bangladesh, 36–49% in Nigeria, 33% in Kenya, and 55–77% in Thailand. Second, they point to the very few studies that attempt to determine the size of the informal sector; in Bangladesh, estimates range from 88% to 96%, and in Uganda, the estimate is 77%. Our data from India suggest that close to 80% of all health care providers fall into this category.⁸

In all these exercises, the working definition of a provider is someone who receives payments (either through a salary or via fee-for-service from the patient or a third party) for providing

⁷The MAQARI project collected data from 19 states in India on summary measures of availability and quality between 2008 and 2011. In one state, Madhya Pradesh, a detailed mapping was completed of 100 villages in five districts. The sampling was representative of rural Madhya Pradesh beyond a 10-km radius of large towns and cities. The project team included Jishnu Das and Alaka Holla (World Bank), Karthik Muralidharan (University of California, San Diego), and Michael Kremer (Harvard University).

⁸Expanding the set of countries in these studies could reveal significant diversity in the number and types of providers. In recent work in Kenya, for instance, we find private clinics run by nurses, clinical officers, and medical officers, but typically they are both registered and licensed.

medical advice beyond a product, such as medicine. Consequently, we are counting all manner of providers the same in these exercises—fully qualified doctors as well as people with no medical training that can legitimately be called quacks (and often are called such).

The preponderance of providers with no medical training raises the obvious question of whether we have got this all wrong: Perhaps there is a dearth of genuine medical professionals, and access remains the key issue, as access should be counted as the availability of real doctors, rather than quacks. Whether this is a valid concern depends on both whether consumers view public/private providers as substitutes and the difference in quality between public and private providers.

With regard to consumer demand, people switch back and forth between public and private providers with some frequency, so from the consumers' perspective, these providers appear to be close substitutes. Determining the elasticity of substitution—with respect to price; location (distance); perceived severity of illness; or, the point of this article, perceived differences in quality—is a difficult task that is unlikely to yield general results.⁹ Summarizing early studies on this issue, Filmer et al. (2000) find all possible numbers, ranging from nearly zero to 100% crowding out of the private sector by expansion of the public sector, either by new construction closer to villages or by lower fees at existing facilities.¹⁰ However, we observe both types of providers being used by the same people at different times and frequently for the same episode of illness. This makes the position that they are not at all substitutable hard to maintain.

The degree of substitutability between public facilities and private providers suggests that quality differentials between these types of providers may not be as large as usually assumed. Therefore, we turn specifically to the issue of quality. As a preview, when directly measured, the difference in the quality of clinical advice between the typical public doctor and even untrained practitioners is small, and sometimes the so-called quack is better than the doctor. Replacing the term access with the phrase access to real doctors does not alter the picture of high availability described above.

3. QUALITY: MEASUREMENT AND PRACTICE QUALITY VARIATION

First-generation studies used the presence or absence of drugs combined with the availability of medical equipment as a proxy for quality. Das et al. (2008) critique such studies on two grounds. First, the availability of drugs is clearly problematic because stockouts (facilities running out of drugs more often) are potentially correlated with good, not bad, quality. Second, sick people go to a doctor because they do not know what is wrong with them. They rely on a doctor's expertise to give an accurate diagnosis and recommendation for treatment that they would not be able to provide themselves. The quality of the clinical encounter has to do with the accuracy of the advice, and it is this accuracy that represents the true value added of the provider. The presence of drugs may measure the degree of subsidy the visit involves, but if the wrong drugs are given, the value of the subsidy (as opposed to its cost) is nil. Quality involves information.

⁹Leonard & Zivin (2005) point to complex patterns of care seeking in Cameroon. If patients suspect malaria, they visit a public clinic because they know they will just get an antimalarial drug, probably at a subsidized price. If they suspect something that is more complicated or expect that treatment will be time intensive, such as chronic pain, they will go to any number of alternatives, including traditional healers, as they know that providers in public facilities will not spend the time to figure out what to do or, given previous experience, they know that there is little to be expected from modern doctors.

¹⁰If the presence of a public doctor displaces an actually harmful quack, this may be an added benefit. If the difference is not very great, or even perverse, then the substitution clearly undermines the net value of public provision, that is, net of what would happen in its absence.

Measuring the accuracy of the clinical encounter is harder than counting drugs on the shelf. Three methods have been used for such measurements: medical vignettes, direct observation, and standardized patients, each of which we describe briefly below.¹¹

3.1. Measuring Quality: Vignettes, Observations, and Standardized Patients

A vignette is a hypothetical case in which the interviewer acts as an (unblinded) patient and provides a very brief description of symptoms. The specific process described below has been implemented and validated in low-income settings by Das & Hammer (2005) and Leonard & Masatu (2005), building on earlier applications in Jamaica and Indonesia.¹² The doctor, who knows that the interviewer is not a real patient, is then invited to proceed exactly as he or she would under normal circumstances, asking questions about the history of the illness and performing necessary examinations. The “patient” provides standardized predetermined answers to the questions and examination procedures appropriate for the underlying condition. Usually, a second interviewer is present to provide answers to questions that the patients may not know, such as the results of a blood test should the provider say he or she would ask for one. The second interviewer also notes the treatment prescribed.

Several vignettes are asked of each provider and are intended to capture the provider’s behavior both for cases that should be treated at primary care clinics and for those that should be triaged to higher levels. For instance, diarrhea in an infant or child is usually included, and separate vignettes have been developed that lead to a (correct) conclusion of viral diarrhea without dehydration, diarrhea with severe dehydration, or dysentery. Along with these more common tracer conditions, the set of vignettes typically includes conditions that are relatively uncommon on a day-to-day basis but are important for a provider to detect and triage appropriately, such as a heart attack, pre-eclampsia (a serious complication of pregnancy), tuberculosis, or pneumonia.

Interview results—the number of relevant questions asked, including those necessary to rule out more complicated or serious conditions; examinations conducted; tests requested; diagnoses offered (if any); or treatment suggested—are evaluated against a protocol or the judgment of a team of medical experts. These are used to extract a composite score for each provider, using standard psychometric methods.¹³ Because providers realize that they are being tested, vignettes test what a doctor knows, which we call competence.

Direct observation measures what a doctor does when faced with a real patient. Interviewers sit for a while, usually a day, in the doctors’ clinics. The interviewer records details of every interaction, including the time spent, the questions asked, the examinations performed, the treatment dispensed, and the price charged. Sometimes these observations are supplemented by exit surveys of patients to see, for example, if treatment varies by patient characteristics, such as education or apparent income.

Time spent with each patient, the number of (relevant) questions asked, and examinations performed can then be aggregated into an index of effort. When combined with vignettes on

¹¹These methods of measuring quality—which require direct contact with the provider—contrast with more common methods of quality measurement in OECD countries with greater reliance on outcome data in hospital settings and on data abstraction from patient charts in the primary care settings. These measures were developed and validated to differing degrees in OECD countries (see, e.g., Peabody et al. 2000, Beullens 1997) and then extended to low-income countries.

¹²Used both as a training technique and in medical exams, medical vignettes were first extended to the survey setting in Jamaica and Indonesia.

¹³Details of these techniques can be found in Das & Hammer (2005) for India and Das et al. (2008) for a broader set of countries.

common ailments done with the same doctor some time before, direct observation can be used to make a direct comparison between what health care providers knew (and what researchers knew they knew) and what they did.

Standardized patients provide the most ambitious of the measurement techniques and give the most accurate picture of what providers really do, albeit for a more limited set of cases. Standardized patients are people from the local community who are extensively trained as actors to present the same case to multiple providers. After the interaction, they are debriefed with a structured questionnaire to recall all history questions asked, examinations done, and diagnoses given by the provider. The quality of medical advice is assessed by the time spent with patients, by providers' adherence to case-specific checklists of recommended care, the likelihood of correct diagnosis, and the appropriateness of treatment. For a number of reasons, discussed below, standardized patients are widely regarded as the gold standard in assessing the quality of medical care delivered in outpatient settings (see Rethans et al. 2007 for a review of the medical literature on standardized patients). Standardized patients will be easily recognized as the medical counterpart of audit studies, which have become an important tool in labor economics for measuring discrimination or the behavior of agents under different institutional frameworks (recent examples include Hanna & Linden 2012; Bertrand & Mullainathan 2004; and, in the context of credence goods, Balafoutas et al. 2014).

In low-income countries, the first large-scale population-based study in a representative sample of providers was completed in India in 2008–2009 (Das et al. 2012b). This study included coaching by a professional standardized patient trainer, doctors, and an anthropologist to consistently portray the emotional, physical, and psychosocial aspects of the case. Standardized patients were also thoroughly trained to make plausible excuses to avoid thermometers, needles, and pelvic exams and to hide medicine that doctors requested them to ingest in the clinic. In this study, the standardized patients presented with unstable angina, asthma, and dysentery in a child who was sleeping at home (i.e., the standardized patient plays the parent who went to the doctor).

3.2. Comparisons of Different Quality Measurements

There are pros and cons for each of the methods (Table 1). Vignettes are the most artificial of the techniques as the provider knows he or she is being interviewed and tested. Vignettes really do just measure the provider's knowledge and are an upper bound on the quality of care of which the provider is capable. It is best to use them in conjunction with one of the other techniques if a measure of what happens in the real world is the goal.

Direct observations of provider-patient interactions are one way to observe practice in the real world. Clearly, there are shortcomings. First, observers never know the real sickness of patients, and therefore, the accuracy of the diagnosis and treatment cannot be evaluated.¹⁴ Second, most people go to a doctor with a minor ailment, either to receive palliative care or to receive assurance that their condition is not something serious. Therefore, in one day of observation, observers almost never see a patient with a heart attack or another potentially life-threatening condition.¹⁵ Third, comparing the practice of different providers is fraught with interpretational issues owing to patient selection. Different providers may see vastly different mixes of cases, making comparison across them difficult. Finally, respondents may change their behavior because they know they are

¹⁴Even if the observer suspects a condition that the provider misses, unsolicited advice can alter the doctor-patient relationship, harming the research subject. For this reason, observers are typically medically untrained.

¹⁵This fact also makes it hard to know whether a statement such as "20% of patients who felt they were sick did not seek medical care" represents a problem. Maybe the decision to not seek care was correct.

Table 1 Comparison of different quality measurement tools

Measure of quality	Description	Measures knowledge?	Measures practice?	Accounts for case mix?	Accounts for patient mix?	Hawthorne effects?	Illnesses covered
Medical vignettes	Standardized presentations of hypothetical patients in which providers know they are being tested	Yes	No	Yes	Yes	Yes, by design (vignettes measure the maximum a provider can do)	All. Typically providers are given between three and five vignettes, and these include both cases that the provider would be expected to treat in the primary care setting and cases that he or she should be able to identify as serious and triage into higher levels of care. Example tracer conditions include diarrhea or dysentery in a child, pulmonary tuberculosis, unstable angina, viral pharyngitis, and malaria with anemia.
Clinical observation	Observations in which enumerators sit with providers for some time (typically a day) and record various attributes of every interaction	No	Yes	No	No	Yes, given the evidence that when an observation begins, providers increase effort, and this is followed by a rapid return to preobservation levels	Limited in two ways. First, serious illnesses such as unstable angina will show up on a sporadic basis. Second, the observer never knows what the patient actually has—and doctors frequently make incorrect diagnoses.

(Continued)

Table 1 (Continued)

Measure of quality	Description	Measures knowledge?	Measures practice?	Accounts for case mix?	Accounts for patient mix?	Hawthorne effects?	Illnesses covered
Chart abstraction	Patient charts maintained by providers and assessed and graded by expert teams, providing an important way to measure quality in OECD countries	No	Yes, to the extent that the relevant information is maintained in the chart	No	No	No, unless providers know that the charts will be assessed at a later stage	Severely limited for all illnesses. There has been no success with chart abstraction in low-income countries. Providers rarely maintain patient records, particularly in the private sector. Even when they exist, charts tend to be incomplete and do not accurately reflect patient-provider interactions. For instance, in a study in Madhya Pradesh, less than 20% of standardized patients sent to clinics were found in patient records with accurately recorded symptoms. There was no information on what the provider did.
Standardized patients	People recruited from the local community and extensively trained to present the same case to multiple providers	No	Yes	Yes	Yes	No	Typically limited to (a) adults only, (b) diseases that do not have any obvious physiological symptoms (which cannot be mimicked), and (c) conditions that do not require invasive exams—particularly in low-income countries. The use of a thermometer could be a dangerous invasive exam in some contexts.

being tested (the Hawthorne effect).¹⁶ Although it is always a danger with research in which subjects are under overt observation, in one study that carefully measured the Hawthorne effect, it was not as bad as one might think. In Tanzania, doctors put in 20% more effort in terms of items on a checklist they are supposed to follow when observers first entered the clinic, but this initial bump was short-lived and returned to previsit levels within a short time (Leonard & Masatu 2006, Leonard 2008).

Standardized patients solve most of the problems of direct observation studies. Researchers know what the underlying problem is because they designed it. The choice of ailments can be made to fit the circumstances relevant for study rather than relying on chance for a heart attack patient to show up, for example. Using standardized patients also directly avoids the case-mix problem, whether it was induced by chance or by conscious selection by patients. Furthermore, in comparison to methods such as patient exit interviews, there is no recall bias or heterogeneity, and relative to chart abstraction, reports are standard and complete.

However, similar to audit studies in labor economics, there are limitations to standardized patient studies and the interpretation of the results that follow from them. One limitation is the exclusion of illnesses, usually infectious, that have clear physical manifestations such as a high fever. Sometimes these can be avoided, such as when the actor plays the role of the parent of an infant left at home, or when tests that would indicate a disease can be done in a separate facility. Sometimes the constraint binds, and standardized patients for those diseases cannot be studied. A second limitation is that standardized patients (at least to date) are unknown to the provider and do not typically present for follow-up. Therefore, providers may bias their care toward emergency medicine rather than continuing care, and in cases in which the illness may justifiably require two to three visits to resolve, we observe only part of the full interaction. This partial observability is known to cause problems, for instance, as discussed in the context of car sales by Ayres & Siegelman (1995) and Goldberg (1996). Finally, the extent to which statistical discrimination may account for differences in observed outcomes has yet to be studied.

In the absence of administrative data on quality, these methods are now becoming part of the standard toolkit for health researchers in low-income countries. One recompense for the challenges of primary data collection involves the flexibility and possibility of innovation that such measurements imply.¹⁷ Examples include new research on the nonprice determinants of provider behavior through lab-in-the-field measurements of attributes such as altruism, research on the relative performance of the same providers in their public and private clinics using standardized patients, and research that alters the characteristics of standardized patients to examine the sensitivity of antibiotic use to patient characteristics. In many cases, these measurements are further combined with experimental manipulations. These studies shed light on the deep determinants of practice quality variation across doctors and provide valuable insights not only on the functioning of health care markets in low-income countries, but also on the behavior of health professionals in a market setting more broadly. We turn next to a finer description of these studies and what we learn from them.

3.3. Patterns of Practice Quality Variation

We first document the nature and extent of practice quality variation in low-income countries. We then turn to the determinants of practice quality variation, grouping them into two broad

¹⁶For vignettes, this is not a problem because the whole point is to find out as much as the respondents know. Similarly, it is not at all a problem for standardized patients because the doctor does not know that the patient is a fake.

¹⁷Attempts to study quality face serious hurdles in rich countries. In the United States, a proposal to use mystery shoppers to assess the time taken to obtain an appointment could not be implemented owing to resistance from doctors (see Sounart 2011).

categories. The first are structural determinants, mostly equipment and case load, which are weakly (or not at all) correlated with the quality of medical advice.¹⁸ This is of interest primarily because structural factors are widely viewed as the usual culprits leading to low quality, and alleviating such constraints has been the mainstay of global health interventions. The second are behavioral determinants, most importantly, provider effort. We show that provider effort is closely linked to accurate diagnosis and treatment. Several instruments improve provider effort in low-income countries with implications for health outcomes.

3.3.1. Competence. Medical vignettes have been used in several countries, and although cases have differed, several basic patterns are emerging. First, overall quality can be poor, with low completion of checklist items, poor diagnosis, and treatment skills. In India, where close to 500,000 children die every year of diarrhea, only 25% of providers in the richest state, Delhi, asked about blood/mucous in the stool, 49% asked whether the child had a fever, and 7% checked for a depression in the skull fontanel. These essential questions and examinations allow the provider to differentiate viral from bacterial causes and to assess the degree of dehydration—thus, whether the child needs immediate hospitalization. In Tanzania, these numbers are only slightly better, and even in Indonesia, except for asking about the frequency of stools, there is still less than a 50% chance that any of the other essential procedures would be requested (Leonard et al. 2007, Das et al. 2008).

In terms of knowledge regarding the correct treatment, Das & Hammer (2005) use an independent panel of providers to grade all treatments by the providers in their sample from urban Delhi, the richest state in India. They report that a provider had to have above average competence to have a 50% chance of not harming their patients. Even among the top 20% of providers, the likelihood of harming the patient was more than 50% for viral diarrhea, 25% for preeclampsia (a potentially life-threatening condition of hypertension in pregnancy), and 7% for tuberculosis. Leonard et al. (2007) document similarly low levels of competence in Tanzania, and pilot studies from Senegal and Ghana through the World Bank's Service Delivery Indicator project (see below) reveal similar deficiencies (World Bank 2011).

Second, there is tremendous geographical variation in competence. Over the past five years, two initiatives have measured the same set of conditions in a large number of geographical settings. The MAQARI project covered 19 Indian states and measured the availability and competence of over 6,000 providers sampled in these states. Preliminary results from this assessment of the availability and quality of health care providers suggest a difference of close to two standard deviations on an aggregate measure of provider competence across Indian states (the southern states are better). In terms of diagnostic accuracy, in the worst states (Uttar Pradesh, Jharkhand, and Bihar), rates of correct diagnosis were very similar at 51%, 16%, and 12% for tuberculosis, dysentery in a child, and preeclampsia, respectively. In the best states (Gujarat and Tamil Nadu), diagnostic accuracy for the same tracer conditions was 93%, 91%, and 94%, respectively. The differences across these states are so large that providers with no medical qualifications in the better states are 1–1.5 standard deviations more competent than the fully trained MBBS providers in the worst-performing states.¹⁹

¹⁸The use of the term structural is based on a widely used framework of quality comprising structure, process, and outcomes (Donabedian 2005). Here, structure refers to the physical aspects of a clinic.

¹⁹These data are based on a preliminary analysis as part of the MAQARI project by Monisha Ashok (Harvard University), Jishnu Das and Alaka Holla (World Bank), Karthik Muralidharan (University of California, San Diego), and Michael Kremer (Harvard University).

Similarly, the World Bank started measuring the competence of providers in several countries in sub-Saharan Africa through its Service Delivery Indicators initiative (<http://www.sdindicators.org>). Results so far show that competence among providers is much higher in Kenya relative to Tanzania or Senegal (World Bank 2011, 2013b). For five tracer conditions that the initiative covers, diagnostic accuracy was 34% in Senegal (with 22% adherence to clinical guidelines) compared to 57% for Tanzania (35% adherence to clinical guidelines) and 72.2% for Kenya (43.7% adherence to clinical guidelines).²⁰

These large differences in competence are arguably related to the quality of medical training: All the studies consistently find a positive correlation between medical training and competence (it would be worrying if they did not); they also find little or no correlation between competence and experience. The latter could reflect the joint (positive) effects of experience and cohort effects, whereby younger cohorts are better trained.

3.3.2. Effort. Low competence is compounded with worryingly low effort in doctor-patient interactions. In urban and rural India, the average consultation time is three minutes, during which time the provider asks three questions, completes one examination, and gives three different types of medicine. One-third of interactions lasted less than one minute, with one question (“What’s wrong with you?”) and no examinations. Das et al. (2008) present basic characteristics of doctor-patient interactions across several countries and find similar practices across several low-income countries, compared to consultation times that are three to four times higher in OECD (Organization for Economic Co-operation and Development) contexts.

Of particular interest is that in the two countries for which there are data (Paraguay and India), roughly half the variation in effort is across providers, and half is within providers. The variation in effort across providers in India is clearly tied to whether the provider is in the public sector; for instance, among the interactions that lasted less than one minute, most were in the public sector, and time spent in the public sector is 30–50% lower than in the private sector (Das & Hammer 2007). Variation in effort across providers is also linked to their competence. More competent providers exert higher levels of effort, suggesting that effort and knowledge are complements in the production of quality. Notably, however, the correlation is qualitatively small (Das & Hammer 2007, Leonard et al. 2007).

Within-provider variation in effort is harder to explain. In Paraguay, researchers conducted exit surveys with patients leaving the providers’ clinics, asking them about both their socioeconomic backgrounds and their physical health. Das & Sohnesen (2007) find no correlation between provider effort and various measures of physical health, including self-reported health status, activities of daily living, the number of days sick, and the presenting symptoms. Neither is there a correlation between effort and the wealth or education of the patient; a fair amount of the variation within providers remains a mystery. The one variable that does affect effort within providers is patient order—in Paraguay and Tanzania, patients seen later in the day receive less time, with fewer questions and fewer examinations (Das & Sohnesen 2005, Leonard 2008, Brock et al. 2013a).

3.3.3. Under- and overtreatment. Evidence on the extent of under- and overtreatment when providers face real patients requires both that researchers know what the conditions were that the patients presented with and what the providers did with these patients. In the absence of

²⁰These tracer conditions were malaria with anemia, diarrhea, pneumonia, pelvic inflammatory disease, and pulmonary tuberculosis.

administrative data and/or chart review, standardized patients are the only measurement tool that meets both these requirements. The first study based on a representative sample of providers sent standardized patients with three tracer conditions—unstable angina, asthma, and dysentery in a child (sleeping at home)—to a large sample of providers in rural Madhya Pradesh and urban Delhi (Das et al. 2012b).

Das et al. (2012b) document a series of deficiencies in medical practice. First, a large percentage of health care providers are not medically trained—not surprising given our previous discussion. In the rural Indian setting, 67% of health care providers reported no medical qualifications at all. More surprisingly, when standardized patients visited public clinics, they were seen by whoever was providing care at that time. In 63% of interactions in public clinics in rural Madhya Pradesh, this was also a provider without medical training, pointing to the combined problems of finding qualified doctors for rural postings and widespread absences among doctors in rural public clinics (Banerjee et al. 2004, 2008; Chaudhury & Hammer 2004; Chaudhury et al. 2006).

Second, overall quality was poor in both urban and rural settings and across all types of providers. Visits lasted 3.6 minutes on average in the rural setting, providers completed 33.7% of essential questions and exams recommended by medical guidelines, and 32.6% gave any diagnosis at all. Of those who gave a diagnosis, 12.2% were correct, and 41.2% were partially correct. The correct treatment was given 30.4% of the time (treating unknown treatments as missing), but unnecessary or harmful treatment was given 41.7% of the time. The results were somewhat better in the urban setting, in which providers spent more time (5.3 minutes) and were more likely to give the correct treatment (47.8%).

This study was then replicated in a sample of 36 village clinics and 12 township health centers in southern Shaanxi province in China (Sylvia et al. 2013). Despite the much higher incomes of the area surveyed (GNI per capita was \$583 in Madhya Pradesh in 2011, compared to \$3,179 in Shaanxi), the study found very similar results. On average, village clinicians spent 1.6 minutes consulting with patients and asked 18% of the recommended questions. For unstable angina, only 15% of the recommended exams were completed. Fully correct diagnoses were given in 26% of interactions, and the diagnoses provided were completely incorrect in 41% of interactions, with correct or partially correct treatments in 53% of interactions. Results were somewhat better in the larger township health centers, with a 52% rate of correct diagnosis and 6% higher correct treatment rates for unstable angina, although correct treatment rates were lower for dysentery, which is surprising because this is the easier problem to treat. These differences were entirely attributable to differences in medical qualifications and education.

3.4. Determinants of Practice Quality Variation

Despite the low average quality observed in different countries, there is substantial variation in the overall population and within subgroups of health care providers with differing qualifications. A persistent question is whether these variations reflect differences in the inputs that providers have access to, such as better equipment or physical infrastructure, or facets of provider behavior that are independent of the inputs that they can use. We turn to this next.

3.4.1. Structural quality and case load. It is often believed that these low levels of quality reflect poor structural inputs (medical equipment) and high case loads. In fact, across all studies, there is no correlation between structural inputs and practice quality; one reason may be that the cases used thus far do not require much equipment, and there are very few clinics without stethoscopes and/or sphygmomanometers to measure blood pressure (see Das & Gertler 2007 for a broader discussion).

There is also no link between case load and practice quality. For instance, Maestad et al. (2010) use the size of the catchment area to instrument for case load in Tanzania and show no link between patient load and a number of variables measured at the level of the doctor-patient interaction. This is surprising if we believe that access to health care is a serious problem in low-income countries, but is less so given the actual data on patient load.

In fact, direct observations in rural areas reveal enormous excess capacity in public clinics. In Tanzania, Senegal, Kenya, and India, the regular patient load rarely exceeds 15 patients, and usually averages between 8 and 10 patients a day. On average, providers spend no more than a half-hour per day actively seeing patients. In India, the average amount of time doctors spent with patients was about 40 minutes per day, with the 5th percentile at 10 minutes per day and the 95th around 2 hours. In Kenya and India (where we also have data on private providers), excess capacity is not restricted to the public sector; it is rare to find a rural provider who spends more than 1 hour a day actively seeing patients, although these providers are in their clinics for far longer (World Bank 2011, 2013b; authors' analysis based on MAQARI data). Although not the main purpose of these studies, these new data provide mounting evidence against mere access as the defining problem of health care in resource-poor settings.

3.4.2. Provider effort. In stark contrast to structural constraints, provider effort is a consistent determinant of quality. The clearest manifestation of this is an empirical regularity, discovered when vignettes are combined with measures of observed effort from either direct observation or standardized patients. This is the phenomenon increasingly known as the know-do gap. What providers say they would do, determined from vignettes, is often vastly different from what they are observed to do in practice for patients with the very same set of symptoms. Using data from Delhi, India, **Figure 1** plots on the horizontal axis what doctors said they would do faced with a particular patient; for instance, they may tell surveyors in vignettes that for a patient with diarrhea, they would check the frequency of urination, the nature of the stool, and whether the patient has experienced vomiting. On the vertical axis, the figure plots what doctors actually did faced with a similar patient.

If providers did everything that they knew, we would expect a plot of what providers know versus what they do to lie on a 45° line. In fact, it does not. What we find is that at low levels of medical knowledge, providers pretty much do whatever they know how to do—at this end of the distribution, knowledge is indeed the constraining factor.²¹ However, as knowledge increases, practice does not keep up, leading to a gap between knowledge and practice that widens with knowledge. In areas where we have data on the public and the private sectors (India), the gap is larger for the public sector but is also large and qualitatively significant in the fee-for-service private sector.

Given excess capacity in most clinics, this gap does not reflect optimal rationing on the part of the provider. Neither does it reflect structural constraints arising from a lack of equipment—there is no correlation between effort and various indices of infrastructure and the availability of equipment. However, the gap does respond to a variety of financial and social incentives, as well as to attempts to improve intrinsic motivation. In turn, closing the gap improves patient satisfaction and health outcomes. Four studies demonstrate these results.

Das et al. (2013) focus on the treatment of standardized patients across the same providers in public and private clinics in rural Madhya Pradesh, India. In their setting, public sector providers

²¹The precise calculation has been made for India in Das & Hammer (2007), for Tanzania in Leonard et al. (2007), and for Rwanda in Gertler & Vermeersch (2012). There is also evidence for a know-do gap for the Netherlands in Rethans et al. (1991) and for the United States in Schwartz et al. (2013).

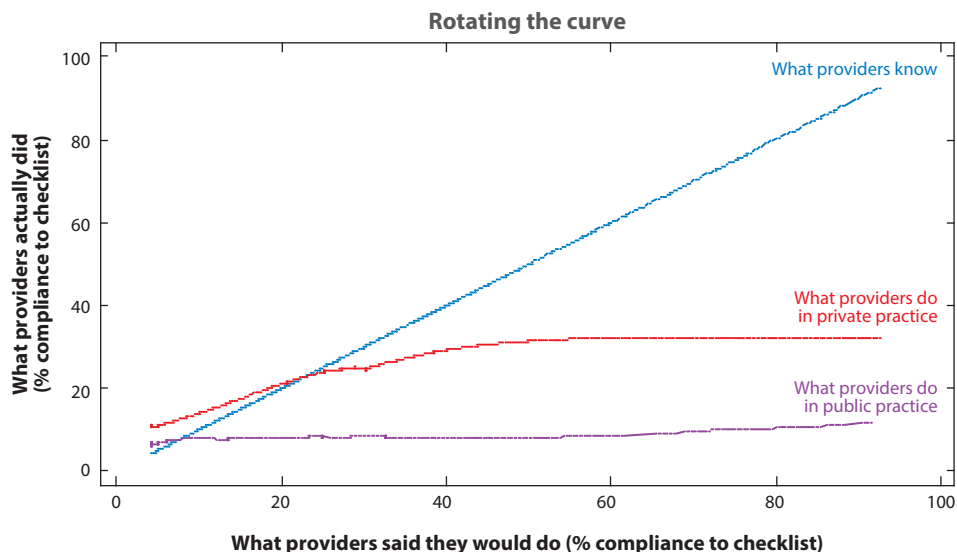


Figure 1

The know-do gap in medical care. The horizontal axis plots what a provider knows, as measured by medical vignettes, using percentage compliance with a medically necessary case-specific checklist of history questions and examinations. The vertical axis plots what the provider actually did with a similar patient, observed in practice. Every history question and examination can be compared in a pairwise comparison. If providers did everything they told us they would do, we should observe them on the 45° line. At very low levels of knowledge, practice is constrained by knowledge; at higher levels of knowledge, there is a significant gap between knowledge and practice; and the know-do gap is larger in the public sector (in which there is no correlation between practice and knowledge), but even in the private sector, there is a significant gap at higher levels of knowledge.

are paid a salary and are accountable to administrative norms and criteria. In their sample, 83% of public providers also have a private clinic. There is no formal health insurance, and in private clinics, providers are paid on a fee-for-service basis and are fully accountable to their customers. The same standardized patients, presenting with asthma, unstable angina, and dysentery, visited both the public and private clinics of the providers.²²

Das et al. (2013) report three results. First, providers spend more time with the patient, ask more questions, perform more exams, and cover more items in the checklist when the standardized patient visits their private rather than public practice. Second, correct treatment rates are higher in the private clinic—by 28.4 percentage points for unstable angina and 11.8 percentage points for asthma. Third, incorrect treatment rates (most patients receive some incorrect treatments) are no different between public and private clinics. Das et al. (2013) examine several structural explanations for the difference and find little evidence—excess capacity in both public and private clinics rules out optimizing behavior; providers have more equipment in the public clinic; public providers with and without a dual practice behave similarly and on process outcomes; and the differences are identical among the real patient population. The authors argue that the incentives generated through customer accountability in the fee-for-service private market lead to higher

²²For the dysentery case, a parent comes to seek advice for a child who is sleeping at home. Providers were visited by the asthma and dysentery standardized patient in both public and private clinics. The unstable angina case was randomly allocated to either the public or the private clinic.

quality care than those generated through administrative accountability in the public sector—at least in this example, in which administrative accountability is poor. Furthermore, customer accountability does not increase the likelihood of incorrect treatment relative to the public setting, although we note that this does not imply that incorrect treatment rates are low or anywhere close to efficient in the private market. In fact, interviews with providers (with one example further below) suggest that there are significant incentives to overprovide in the private sector; the statement here compares the dispensation of medicine in the private sector at market prices with the dispensation of medicine in the public sector at zero prices. That unnecessary medication in both sectors is similar suggests that the combination of price and incentive effects balances out. We do not know what results would look like if medicine were priced positively in the public sector.

More evidence linking financial incentives to performance and health outcomes comes from an experiment in Rwanda (Gertler & Vermeersch 2012). In 2004, Rwanda introduced performance-based pay in public sector clinics, and clinics were paid a performance bonus on the basis of complex measurements related to task completion and patient load. A control group did not receive the performance bonus but did receive equivalent revenue so that comparisons would elicit the difference due to incentives rather than an overall relaxation of budgetary constraints. Gertler & Vermeersch (2012) report three key findings. First, they confirm the know-do gap among Rwandan doctors and again find that the gap increases with training. Second, they show that the know-do gap reduced owing to the introduction of performance pay—although the reductions were qualitatively small. Third, they show that children in the catchment of treated clinics were 0.2 standard deviations taller in the end-line survey. The implied elasticity of improvements in anthropometric outcomes to provider effort is large, and the trial confirms a link among higher provider effort, better process quality measures, and health outcomes.

Incentives need not be monetary. For instance, Leonard & Masatu (2006) and Leonard (2008) interviewed patients as they left the clinic without the attending provider's knowledge. They asked patients what the doctor did, as well as their satisfaction with the clinical encounter. They then sent a separate team of surveyors to the provider's clinic and document an immediate jump in task completion by the providers, attributable to a Hawthorne effect. Increased task completion drops off quite rapidly, but importantly, patient satisfaction tracks the task completion. Although waiting time presumably increased as providers spent more time with each patient, patients were able to perceive better care and evaluated the increase in effort positively against the increase in waiting time. Because there was no other intervention apart from the Hawthorne effect induced through observation, clearly providers knew that their effort levels were less than optimal and increased the effort levels with observation. In a sense, this study exploits the Hawthorne effect to hint at the incentive effects of being monitored. Similar results have been noted in the United States: Colonoscopies that are videotaped have lower error rates than those for which the doctor is not observed, and announced standardized patients receive better care than unannounced standardized patients (Rex et al. 2010, Schwartz et al. 2013).²³

²³In our Delhi study (Das & Hammer 2007), we note that the best performers overall, taking into account both effort exerted and the avoidance of overprescribing medicine, particularly antibiotics, were public doctors in major public hospitals. This was most notable among the most knowledgeable doctors as measured by their vignette scores. Doing very much worse were public doctors in primary care clinics. A plausible explanation is that the best young doctors (this was all in the outpatient clinic where few senior doctors work) in the best hospitals may be motivated by career concerns or are sensitive to peer pressure in environments in which there are multiple providers [although Doshi (2010) reports rates of antibiotic use in public hospitals as high as in private clinics in Mumbai, India]. Primary care clinicians have an entirely different career path than the elite of the profession, and that career path is much less likely to depend on the quality of their practice. There are many issues and mysteries surrounding the deployment of doctors in the public service in India, which are well beyond the scope of this article (see World Bank 2013a).

Such monitoring of behavior can be done by any number of actors, who need not be medical professionals. Björkman & Svensson (2009) evaluate the impact of a community empowerment intervention in Uganda. The intervention bundled community meetings and informal agreements similar to contracts with local public providers with information on these providers. There was no attempt to increase either the training of providers or the availability of equipment for health clinics. As Björkman & Svensson (2009, p. 757) note, “in the experiment we consider, on the contrary, no new health interventions were introduced and the supply of health inputs was unchanged. Instead we focus on incentivizing health workers to carry out their tasks through strengthened local accountability.” They show that under-five mortality rate in the experimentally treated villages declined 33%, bringing the rate down from 144 in the control group to 97 in the treatment group. Although the causal chain is complex, the study demonstrates a clear link between greater provider effort and health outcomes.

Finally, provider behavior can be improved by appealing to intrinsic motivation, rather than financial incentives or social sanctions. For instance, Brock et al. (2013a,b) combine lab-in-the-field methods with measurements of provider effort in Tanzania. To begin, they implement a standard dictator game in which providers were given (approximately) \$12 and asked to donate an amount of their choosing to an anonymous person they had been paired with. Clinicians could observe the pool of receivers, but not communicate, and did not know the specific person they had been paired with. Because the receivers had been recruited from the local market, clinicians would realize that they were likely to be poorer. The authors argue that the amount donated can be construed as a measure of altruism, and in their data, there is a strong fairness norm: 36.8% of the participants gave at least half of their money to the stranger. The authors then look at the link between altruism thus measured in the experiment and the performance (checklist completion) with actual patients.

Brock et al. (2013a,b) demonstrate a striking result: More altruistic providers exert greater effort in their interactions with patients. However, in ongoing work, the authors show that this effect is only for providers in the public sector. In the private sector, there is no evidence that altruism and effort are linked (K. Leonard, personal communication). A much lower effect of altruism in the private sector is precisely what we would expect if market pressures equated the marginal returns to effort in the private sector, but these disciplining mechanisms were absent in the public sector.

Brock et al. (2013a,b) then take on two key questions linked to performance in the public sector. That variation in practice quality in the public sector links to variation in measures of innate altruism implies that public sector performance may be improved by better selection of providers. As Brock et al. (2013a) point out, this is a tough task for governments: Although there are a fair number of altruistic providers in their sample, it is unclear how such a test could be administered on a large scale without strategic behavior or gaming on the part of the providers. Furthermore, any selection on altruism would leave out half the available pool of providers. This leaves open another question: Is motivation malleable, and thus a relevant policy variable?

Surprisingly, Brock et al. (2013a,b) show that it is—and that changing levels of motivation is not a Herculean task. They report on three sets of interventions in a randomized trial. In one intervention, providers are visited by peer monitors. In a second intervention, providers are given a motivational book that talks about the nobility of medical practice and the saving of human lives. In a third intervention, providers were encouraged to follow a set of clinical protocols by a peer. All experiments led to significant improvements in provider effort measured as adherence to clinical protocols elicited through exit surveys, and in fact, large gains were observed among clinicians by just participating in the experiment, which involved an encouragement visit followed by the presence of the research team at several points to collect data. A follow-up study currently in process shows that these gains appear to have lasted up to (at least) two years after the initial experiment.

These studies demonstrate the extent to which the quality of care can be improved by increasing provider effort and demonstrate a range of possibilities that can be used to do so. However, except for noting the distinct differences in the behavior of public sector doctors in their private practices, the studies discussed so far are public sector oriented. Part of the reason for this is that teasing out incentive effects in the private sector requires unraveling the whole nature of equilibrium in health markets. Public providers can be studied within their very constrained work environments. Being mostly salaried workers, their incentives can be varied on the margin. In the private sector, all relevant variables (prices, effort exerted, patient choice, provider location) are endogenous and simultaneously determined. To get any further on the determinants of quality in the private sector, as well as the even harder question of the welfare implications of higher or lower quality, empirical work that is not guided by theory is of limited use. So we have to venture into uncharted waters.

4. WHY IS THE QUALITY OF MEDICAL CARE LOW?

One advantage of studying health care in low-income rather than in OECD countries is that these markets are much closer to laissez-faire markets. It is hard to study the markets for medical care in isolation, particularly in rich countries. This is because they are bound inextricably to insurance systems that are heavily regulated or publicly provided. In poor countries, we have more of an opportunity to see how such markets work without the additional complications of third-party payers, tightly controlled public systems, or heavily regulated private providers. Without insurance or an effective public presence, medical care is just another service between a buyer and seller, albeit with very special features that we discuss below.²⁴

There are three classes of market models that are candidates for explaining the poor quality of care in these (free) markets. These could potentially be nested but have not been thus far, suggesting an open area for future research. One is that medical markets are credence goods and subject to problems of asymmetric information that may or may not be corrected by market forces. The second is that medical markets are similar to the market for “lemons” as in Akerlof (1970), also dependent on asymmetric information (and patients’ knowledge that such asymmetric information exists) but in this case resulting in a complete unraveling of the market such that only low-quality providers operate. The third is markets work in textbook fashion and are efficient—apparent problems are all a consequence of poverty. We discuss these in turn, focusing on the credence good approach.

4.1. Medical Care as a Credence Good

It is widely believed that private health care markets are prone to manipulation by the provider because of asymmetric information. Clearly, doctors know more than patients; otherwise, why would anyone seek their advice? The question is, Does this asymmetry reveal itself in the formation of prices, quality, and number of visits we can observe? Where is the market failure?

A synthesis of the nature of, and the potential problems with, credence goods is due to Dulleck & Kerschbamer (2006). In these markets, consumers observe the treatment they received but cannot tell whether they needed it. For instance, a mother will observe whether she received a cesarean section but not whether it was required. Therefore, the consumer needs to trust the

²⁴Although this description is apt for India, we do not wish to claim that all low-income countries are similar. Our field visits in Kenya suggest higher levels of regulation among private providers, but preliminary results from Cambodia suggest a similar laissez-faire situation as in India.

knowledge of the provider who knows what treatment the patient optimally needs but may have incentives to manipulate information in his or her favor by overcharging or by providing unnecessary services, in which unnecessary is defined as those that would not be demanded at their marginal cost should the consumer be fully informed. Dulleck & Kerschbamer present a characterization of such markets under a variety of assumptions and ask when they result in efficiency loss and when market forces have self-correcting properties.

Crucial to the definition of credence goods and the source of the problem is the characteristic of what Dulleck & Kerschbamer (2006) call commitment or the existence of economies of scope. This holds when the provider who makes the diagnosis has a strong advantage in providing it as well. Emergency situations are examples, as is a complicated case in which the patient would have to duplicate a long and expensive set of tests to get a second opinion. Without any economies of scope, the possibility of shopping around undermines the market power of the provider and the need for the patient to trust (put credence in) his or her opinion. With economies of scope, the diagnostician is most likely to be the treatment giver, conferring a barrier to entry into the specific transaction, and it is the “stuck in this particular transaction” part that leads to real problems. If it were only asymmetric information, then one could obtain more information. It is the inability or the high cost of obtaining this information that makes it impossible to break that asymmetry.

Dulleck & Kerschbamer (2006) then show how markets for credence goods can, under certain circumstances, still yield efficient outcomes. It is possible that other features of the product or the institutional setting can undo the damage of the asymmetry of information exacerbated by being stuck. For example, if a patient can get restitution for not getting enough treatment (an assumption they call liability) or if a patient can tell whether treatment seems commensurate with the price (an assumption called verifiability), then the market can be self-correcting. The surprising part of the theory is that this can happen even without actually shopping around (that is the market part).²⁵ However, neither liability nor verifiability can be assumed in poor countries, so the market mechanism for corrections is weak. Liability would most often be ensured by malpractice or well-functioning professional associations. Verifiability depends on the nature of the service.

We should then first ask, What is the evidence that economies of scope lead to inefficiencies in real life? Most of the examples we have are either from public facilities or under a regime of administered prices and so are not obviously related to markets at all. However, the payment structure in many places is such that, even if the facility is public or regulated, the individual providers within the facility have substantial discretion as to how to practice, and their response to incentives gives some hints to the behavior of private markets. Payment schemes exacerbate the problem of economies of scope.

In China, drugs sold by hospital pharmacies account for 40–50% of their revenue, and hospital pharmacies can charge a markup of 15% over the wholesale price (Currie et al. 2012). Although they are putatively civil servants, doctors in hospitals earn performance bonuses that are linked to the revenues generated from their own hospitals and therefore are able to share in the profits from drug sales within the hospital (but not those outside).

It is often claimed in the literature on health care that physician-induced demand leads to overtreatment in primary care, a possible outcome in credence goods models. In two audit studies, Currie et al. (2011, 2012) examine the reasons for the overuse of antibiotics in China. They send standardized patients with symptoms of a sore throat, low-grade fever, and poor appetite to health

²⁵Consider a provider choosing between a cesarean section and a normal birth. The provider knows what procedure is required, and if certain assumptions are met (the patient knows which procedure she ultimately received), equal markups to the provider independent of the procedure will lead to the efficient outcome.

care providers. Although the symptoms are *ex ante* consistent with both bacterial and viral etiologies, appropriate history taking and examinations would reveal that sore throat is viral in nature. The optimal treatment in this case does not involve antibiotics.

Indiscriminate antibiotic use was rampant: 63.3% of all patients were given an antibiotic, and 15.3% were given second-generation antibiotics—typically reserved for cases resistant to more common types. Currie et al. (2011, 2012) then randomize the standardized patients into one of four types: the “baseline” patients, who present the case; the “gift” patients, who offer a small pen to the provider; the “display” patients, who signal to the doctor that they have read on the Internet that antibiotics may not be appropriate for their symptoms; and the “no purchase” patients, who ask the doctor for a prescription because they have a relative who works in a pharmacy and can purchase the medicine from their relative’s store.

The differences in antibiotics prescribed are stunning. From the baseline of 63.3%, those providers who received the offer of a gift reduced their prescription rates to 50%. The patients who signaled their knowledge received antibiotics 43.3% of the time. Finally, the patients who indicated that they would not buy from the hospital pharmacy received a prescription for antibiotics 11.7% of the time, and the complementary treatment of “no purchase” and “display” led to antibiotic prescription rates of 8.3%. Further of note, there is no discernible impact on the likelihood of taking the patient’s temperature, asking about sputum, or using a stethoscope. These results led the Currie et al. (2012) to conclude that

at least in Beijing hospitals, physicians do not prescribe antibiotics primarily because patients demand them, because doctors believe that patients want antibiotics, or because physicians erroneously believe that such prescriptions are in the best interests of the patient. Indeed, physicians who are offered a small gift improve service quality and reduce prescriptions of antibiotics. Hence, doctors who are making a greater effort to please patients reduce rather than increase antibiotic prescriptions.

A third standardized patient study in Chinese hospitals interacted consumer signals of where they will buy the drug from with whether the patient has health insurance (Lu 2014). In this case, there is a further twist: Providers significantly increased the numbers of drugs prescribed when the patient was insured and when they thought that the patient would buy the medicine in-house. Consequently, prescriptions were 43% more expensive when patients were insured and doctors believed that they would purchase the medications from the hospital pharmacy. In addition, the author argues that 80% of this increased drug expenditure is motivated by doctors’ financial interests, representing a welfare loss for patients.

System-wide legislation that restricts such economies of scope reduces expenditures on drugs, as in Chen & Gertler’s (2013) study of the Taiwanese health system. On March 1, 1997, Taiwan decoupled diagnosis and treatment, prohibiting the sale of drugs at all outpatient clinics. Following protests from physicians, the eventual compromise was that “clinics with an onsite pharmacist would continue to dispense drugs. . . . Clinics without an onsite pharmacist would have to ‘release’ their prescriptions to the patients to fill at an outside pharmacy” (Chen & Gertler 2013).

As a consequence of the legislation, clinics without an onsite pharmacist reduced their drug expenditures by 8.6% owing to a simultaneous reduction in the likelihood of prescriptions, number of drugs prescribed, and the duration of the treatment. However, they also increased laboratory expenditures by 3.9%, suggesting that overtreatment was displaced to those services for which economies of scope remained unrestricted. There was no change in the clinics that always had an onsite pharmacist. Finally, among physicians who consequently hired a pharmacist

(but did not have one to begin with), most indicators of prescription behavior rebounded to the prereform patterns.

The studies highlighted thus far emphasize the critical link between mistreatment and economies of scope. However, economies of scope in these studies are either fully or partly administratively determined; the pricing of the drugs and the revenue sharing are governed by administrative arrangements. In pure markets, patients will (rationally) realize that providers who both diagnose and treat will have an incentive to overtreat, and the market should compensate for this behavior by, for example, reducing the price of consultation. This prediction has yet to be tested, but to begin, it is useful to ask whether economies of scope arise generically in free markets for medical care.

In Delhi, India, 84.6% of all patients did not receive any prescriptions but were dispensed medicine directly by the provider (authors' calculations from data presented in Das & Hammer 2007). This medicine is typically taken from wholesale medicine bottles one tablet at a time, crushed into powder in a mortar, and put into paper packets for the patient. As the provider's competence increases, so does the likelihood that he or she uses prescriptions. Private providers in the bottom third of the competence distribution dispense medicine in the clinic 91% of the time, and this decreases to 74% for the top third; simultaneously, the likelihood of giving a prescription increases from 31% to 68%. We find similar patterns in the rural data, with providers with an MBBS degree almost never dispensing medicine in the clinic. Therefore, economies of scope arise endogenously in pure markets for medical care. Providers with low levels of knowledge tend to dispense medicine, whereas those at the higher end of the distribution are more likely to prescribe—thus consciously breaking the vertical integration between diagnosis and treatment.

So providers certainly do behave such that we can suspect market failures due to asymmetric information, as predicted in credence good models, without corrective effects of the markets through liability. Fixing this market would require breaking connections between diagnosis and prescription, on the one hand, and treatment, on the other.

4.2. Medical Care as a Market for Lemons

The credence good model has, as a special case, a Wild East market that can degenerate into a market for lemons (Akerlof 1970). In this special case, patients neither have the ability to verify what treatment they received nor hold providers accountable for what they have done. In this case, there is little or no incentive for providers to exert even minimal effort (quality); this is expected by patients, and there is an equilibrium with too low quality—people would be willing to pay more for extra effort and ability charged at their true costs. The reason is that patients know that providers are not likely to be very good—either because they have little training or because they put in so little effort that they are of limited use to the patient. Therefore, patients do not pay more for quality, and this reinforces the lack of supply of quality. This market is clearly inefficient and could explain the observed phenomenon of the know-do gap, at least among well-trained providers. Nevertheless, we document below that one of the main implications of this model—that patients do not recognize or are willing to pay for higher quality care—does not seem to hold. This raises one last possible model for medical care.

4.3. Markets as Poor but Efficient

In ordinary markets with differentiated products, prices serve two functions: They equilibrate supply and demand, and they act as a signal of quality. If markets are operating properly, then the problem of observed low quality may not be a problem at all. Although the quality of care is

indisputably low, is it too low? In one sense, of course it is—who wants low-quality anything? If the metric is whether the quality of medical advice in poor countries is up to the standards of medical professionals visiting from rich countries, then of course it is a problem. However, by that standard, the quality of all commodities that the poor consume in poor countries (food, housing, clothing, drinking water, education) is too low. What is it about low-quality medical care available in the market that has a larger claim on public resources than, for example, food and water?

In a sense, this line of questioning is similar to Schultz's (1964, 1979) hypothesis on agriculture in developing countries, maintaining that it was "poor but efficient," with the problem being squarely that of poverty and not agriculture at all. In the area of medical care, maybe the markets are working OK (recalling that there is little real intervention in them at all) but we do not happen to like their results because we just do not like to see quality of care that low.

What would the elements of a narrative of medical care such that it is poor but efficient look like? In discussing the other two models above, we make the point that there may not be asymmetric information that characterizes the market, but rather no information at all. Ignorance is a problem of sorts but not necessarily one that is inefficient. An additional piece of information is that there is a lot of shopping around for medical care. So the economies of scope argument may be weaker in these markets than in the Chinese health system as there does not seem to be a lot of dependency of patients on a particular doctor and, in primary care settings, distinct limits to how much providers can extract from poor patients. These are *a priori* types of arguments. There are also results from empirical work that raise some questions about asymmetric information characterizing the private market in all respects.

To support a poor but efficient interpretation in the market for medical care, we need to have a link between quality and prices. This appears to be true in various sorts of empirical work. In Delhi, India, prices are well correlated with scores that providers get on the vignettes, controlling for degrees earned and overall experience. Patients must be able to judge the knowledge of the provider for this correlation to emerge. This connection carries over into measures of effort.

When standardized patients are used to judge quality, a method that guarantees that the same service is being demanded, there is a strong association between quality received among standardized patients and the prices they are charged. Das et al. (2013) find that consumers are charged more when providers spend more time with them. Furthermore, conditional (and unconditional) on the time spent, providers who complete more of the recommended checklist for each case charge higher prices. Finally, correct treatments are rewarded with higher prices—but this higher price results from the direct correlation between greater checklist completion and correct treatment. Thus, the market rewards compliance with the checklist (a combination of effort and knowledge) but cannot discern the difference between correct and incorrect treatments conditional on checklist completion. These hedonic price relations must imply a significant role for consumer choice in price determination. Providers cannot be rewarded for higher quality if consumers cannot recognize it and are not willing to pay for it.

Notably, in the rural Indian study from Madhya Pradesh, the association between prices and checklist completion (and prices and consultation time) is the same across and within providers. This implies that when providers choose their optimal effort levels, they are moving along the price-quality curve, and that for a given provider, increasing the level of effort brings in the same price benefits as moving to a provider with higher quality. In the urban Indian data from Delhi, the across and within provider price-consultation time gradient is identical for low-quality providers but is small (and statistically insignificant) for high-quality providers. This is largely because at the top end of the market, providers charge (virtually) a fixed consultation fee per patient.

Another piece of the puzzle is provided by results on the total earnings of providers as functions of their characteristics. Preliminary calculations from rural Madhya Pradesh suggest that private

providers earn their opportunity cost when they try to sell their services in the medical marketplace. Hence, college graduate providers (not with a degree in medicine) earn the same as other college graduates, secondary school-educated providers earn the same as other secondary school graduates (and less of course than college graduates), primary school graduates the same as others, and so on. Differences in earnings comprise both differential prices and different levels of business across providers. Although we know little of the determinants of the supply of medical care, this result indicates that the elasticity of supply at different quality levels is quite high, and there are few barriers to entry.

This evidence is consistent with the hypothesis that information about quality is sufficiently widespread among potential patients that there is no market failure. With perfect information, the appropriate model is simply one in which hedonic prices move to equilibrate supply and demand at different levels of quality. However, that prices and quality move generally together is not sufficient to prove optimality, although it is sufficient to rule out cluelessness on the part of consumers. Currently, there is no way to tell whether the price/quality relationship is optimal: They have to be related, but there is little guidance from the theoretical literature as to how well they have to be related. The market is certainly competitive in the sense that entry is easy and suppliers earn about as much in medicine as they can earn elsewhere. However, the empirical evidence holds open the possibility that the markets are efficient but still very low quality.

Based on our field work, we also raise a couple of possibilities for specific failures in these markets that could be investigated but have yet to be integrated into a theoretical framework. First, it could be that the market is a bit too competitive, as providers may be prone to give patients what they ask for, even if it is not medically warranted. Currie et al. (2012) show that if patients signal their knowledge that they may not need antibiotics, doctors are less likely to prescribe them. Alternatively, it could be that if patients signal that they want antibiotics, private providers may follow the patients' lead. In our work in Delhi, a frequent refrain from the private providers followed that of Dr. S.:

Dr. S.: Yes, there is a lot of diarrhea and dysentery in this locality—what can they do as well? The water is dirty and people do not know to boil it—that's why their children are always falling sick.

J.D.: So, what do you do for children with diarrhea?

Dr. S.: What can we do? The usual things—we tell the mother to give water with salt and sugar to the baby and then also give some medicines.

J.D.: Such as?

Dr. S.: The usual—Metrogyl (metronidazole), loperamide (an anticholinergic), Furoxone (furazolidone).

J.D.: But isn't ORS enough?

Dr. S.: Of course the WHO and others keep saying that we should only give ORS. But if I tell the mother that she should go home and only give the child water with salt and sugar, she will never come back to me; she will only go to the next doctor who will give her all the medicines and then she will think that he is better than me.

In some cases, public doctors refrain from giving such unnecessary treatments, as they are less sensitive to patient demand (Das & Hammer 2007), but this result does not hold across all settings, even in India (Das et al. 2013, Doshi 2010). Once again, it is important to understand both in what sense this is a problem and how a government would practically be able to solve it. In terms of the

first issue, it is unclear that we can make progress without prespecifying a welfare function for the patient. For instance, whether a steroid that is medically unnecessary but makes the patient feel good is overtreatment depends on whether we view the steroid purely through the prism of medical science or broaden our definition to thinking of the steroid like any other drug of choice, such as alcohol or nicotine. In terms of the second issue, what would it take to correct these errors? Does it require that every clinical contact be monitored and subject to second opinions?²⁶

Another possibility based on results of our research that challenges the poor but efficient perspective is the know-do gap. In many cases, real doctors do much less than they know—many fewer questions are asked and procedures undertaken. However, we do observe a payoff for the doctor to ask those questions as, except for the very top end of the market, the price rewards to effort are identical across and within providers, and most providers operate in a setting with excess capacity. If the doctor puts in more effort, he or she will earn more in a setting in which the marginal cost is, for all appearances, very close to zero. In public clinics, we might say there is no incentive for the doctor to ask questions and conduct examinations, but the same gap appears in both public and private practices. Why is it that private doctors do not work harder and charge more? What is the nature of the short-run labor supply elasticity, and is there a possibility of a market failure?

This broad review highlights areas in which we have made considerable progress over the past decade. These include measuring quality; documenting practice quality variation; ruling out structural factors as a major constraint; ruling in provider effort as a critical variable affecting the quality of care and health outcomes that can be altered through a variety of interventions; demonstrating the link between economies of scope and overtreatment; providing evidence on the functioning of the public sector; and documenting the price-quality relationship in unregulated, private markets. The review also demonstrates areas in which both theoretical and empirical work is required. Unregulated and unsubsidized private markets are a key setting for understanding the behavior of health markets and the market failures that arise in them. Four theoretical and empirical priorities are to understand (a) endogenous economies of scope; (b) the role for patient preferences over different treatments; (c) the ways in which a variety of quality offerings can affect provision; and (d) the market structure that leads to the know-do gap, together with excess capacity in medical care.

5. WHAT CAN BE DONE TO IMPROVE THE QUALITY OF CARE?

Although clearly speculative, there is now evidence on the key ingredients necessary for improving the quality of care in poor countries. First, access is not the problem for much of the world's poor, but access with quality is, and simply assigning a public doctor to a village cannot be assumed to meet reasonable thresholds of quality. Second, improving quality has little to do with improving supply chain management or decreasing patient loads, both mainstays of health interventions in many countries. Although training—another often-used strategy—can help at low levels of knowledge, such programs are unlikely to increase quality at knowledge levels typical of trained providers (for whom most training programs are targeted). In contrast, large improvements in the rate of accurate diagnosis and treatment can result from changing the level of effort that providers exert in their interactions with patients. Third, health care is difficult to fix because

²⁶Cohen et al. (2013) study a similar problem in the context of malaria treatment in Kenya. They show that subsidizing malaria treatment leads to greater use of malaria drugs for both patients with and without malaria. In contrast, lowering the cost of diagnosis increases the likelihood of matching the drug to the disease.

providers can either do too much or do too little, which is both patient and disease specific. In general, any market in which efficiency is transaction specific is extremely hard to address through both the public or the private sector. When we see things being done wrong in practice, there is a tendency to say “just do it right.” But supervising each patient/doctor transaction is patently impossible.

Given the notable lack of evidence that the private sector is causing egregious harm relative to the public, a first generation of reforms could focus entirely on the public sector. Salaried personnel, especially with weak supervision, have little reason to perform. Public doctors in their private clinics in Madhya Pradesh are 23 percentage points more likely to give the right treatment for unstable angina than in their public clinics, which is consistent with the lack of incentives to perform in the public sector.

In such a situation, the clearest way to improve their effort is to increase their accountability, either to administrators or to patients. Reforms such as paying for performance, introducing capitation regimes, or making provider welfare in any way responsive to patient demand could lead to improvements in quality, although whether they will actually do so depends on the precise design of the program and its monitoring (see, e.g., Miller & Babiarz 2013). But there is a broader point that any such incentive-based reform has to eventually take into account the underlying rationale and the limitations of the reform.

The fundamental question here is, Precisely how sensitive do we want public sector providers to be to patient demand? The argument that the way to improve public provision is by basing payments on consumer demand (or measures of user satisfaction) is equivalent to claiming that the private sector can provide the service more efficiently. But of course this is reforming the public sector by making sure that it behaves more like the private sector—but not really because key features that drive performance in the private markets are missing. This is in part because price signals are ruled out in the public sector by the use of administrative pricing so that consumers lose out on valuable market signals regarding the relative quality of different providers. In addition, the decision space is dramatically reduced in the public sector so that there is little discussion of the overall structure of provision (China had to legislate economies of scope in the provision of care). And finally, the use of statistics derived from average rather than marginal benefits in health reforms has little theoretical backing. User satisfaction is one such example; why measures of average satisfaction (even if it were to perfectly measure welfare) with a consciously selected provider should be maximized is difficult to reconcile with standard efficiency criteria in economics.

The alternative to consumer demand–based methods of accountability in the public sector is greater administrative accountability. But this requires building the legitimacy both among public providers to introduce sanctions for poor performance and encourage good performance, perhaps through peer monitoring and motivation-building exercises, and among the political institutions that allow such sanctions (and rewards) to be implemented. Brock et al. (2013a,b) take a step in this direction and represent one example of how such systems of greater administrative accountability, peer monitoring, and supported supervision may be operationalized.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

We thank Pratap Bhanu Mehta, Veena Das, Ranen Das, and Kenneth Leonard for many conversations on the issues discussed here. Brian Chen, Gayle Martin, and Dominic Montagu shared

important material with us for this piece. Some results in this article are drawn from ongoing work on the Medical Advice Quality and Availability in Rural India (MAQARI) study, funded through grant 50728 from the Global Health Program of the Bill & Melinda Gates Foundation. Funding for this article was also obtained from the Knowledge for Change program at the World Bank through a grant for research on the quality of primary care (KCPH TF010746). The findings, interpretations, and conclusions expressed in this article are those of the authors and do not necessarily represent the views of the World Bank, its Executive Directors, or the governments they represent.

LITERATURE CITED

- Akerlof GA. 1970. The market for “lemons”: quality uncertainty and the market mechanism. *Q. J. Econ.* 84:488–500
- Ayres I, Siegelman P. 1995. Race and gender discrimination in bargaining for a new car. *Am. Econ. Rev.* 85:304–21
- Balafoutas L, Beck A, Kerschbamer R, Sutter M. 2014. What drives taxi drivers? A field experiment on fraud in a market for credence goods. *Rev. Econ. Stud.* 80:876–91
- Banerjee A, Deaton A, Duflo E. 2004. Health care delivery in rural Rajasthan. *Econ. Polit. Wkly.* 39:944–49
- Banerjee A, Duflo E, Glennerster R. 2008. Putting a Band-Aid on a corpse: incentives for nurses in the Indian public health care system. *J. Eur. Econ. Assoc.* 6:487–500
- Bertrand M, Mullainathan S. 2004. Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination. *Am. Econ. Rev.* 94:991–1013
- Beullens J, Rethans JJ, Goedhuys J, Buntinx F. 1997. The use of standardized patients in research in general practice. *Fam. Pract.* 14:58–62
- Björkman M, Svensson J. 2009. Power to the people: evidence from a randomized field experiment on community-based monitoring in Uganda. *Q. J. Econ.* 124:735–69
- Brock MJ, Lange A, Leonard KL. 2013a. *Generosity norms and intrinsic motivation in health care provision: evidence from the laboratory and field*. Work. Pap. 147, Eur. Bank Reconstr. Dev., London
- Brock MJ, Lange A, Leonard KL. 2013b. *Giving and promising gifts: experimental evidence on reciprocity from the field*. Unpublished manuscript, Univ. Maryland, College Park
- Cent. Policy Res. 2011. *Mapping medical providers in rural India: four key trends*. Policy Brief, Cent. Policy Res., New Delhi. http://www.cprindia.org/sites/default/files/policy%20brief_1.pdf
- Chaudhury N, Hammer J. 2004. Ghost doctors: absenteeism in rural Bangladeshi health facilities. *World Bank Econ. Rev.* 18:423–41
- Chaudhury N, Hammer J, Kremer M, Muralidharan K, Rogers FH. 2006. Missing in action: teacher and health worker absence in developing countries. *J. Econ. Perspect.* 20(1):91–116
- Chen B, Gertler P. 2013. *Stark contrasts: the impact of prohibiting physician self-referrals on the prevalence of overtreatment in health care*. Unpublished manuscript, Univ. Calif., Berkeley
- Chen L, Evans T, Anand S, Boufford J, Brown H, et al. 2004. Human resources for health: overcoming the crisis. *Lancet* 364:1984–90
- Cohen J, Dupas P, Schaner S. 2013. *Price subsidies, diagnostic tests, and the targeting of malaria treatment: evidence from a randomized trial*. NBER Work. Pap. 17943
- Currie J, Lin W, Meng J. 2012. *Using audit studies to test for physician induced demand: the case of antibiotic abuse in China*. NBER Work. Pap. 18153
- Currie J, Lin W, Zhang W. 2011. Patient knowledge and antibiotic abuse: evidence from an audit study in China. *J. Health Econ.* 30:933–49
- Cutler DM, Deaton A, Lleras-Muney A. 2006. The determinants of mortality. *J. Econ. Perspect.* 20(3):97–120
- Cutler DM, Miller G. 2005. The role of public health improvements in health advances: the twentieth-century United States. *Demography* 42:1–22

- Das J, Gertler P. 2007. Variations in practice quality in five low-income countries: a conceptual overview. *Health Aff.* 26:w296–309
- Das J, Hammer J. 2005. Which doctor? Combining vignettes and item response to measure doctor quality. *J. Dev. Econ.* 78:348–83
- Das J, Hammer J. 2007. Money for nothing: the dire straits of medical practice in Delhi, India. *J. Dev. Econ.* 83:1–36
- Das J, Hammer J, Leonard K. 2008. The quality of medical advice in low-income countries. *J. Econ. Perspect.* 22(2):93–114
- Das J, Hammer J, Sánchez-Paramo C. 2012a. The impact of recall periods on reported morbidity and health seeking behavior. *J. Dev. Econ.* 98:76–89
- Das J, Holla A, Das V, Mohanan M, Tabak D, Chan B. 2012b. In urban and rural India, a standardized patient study showed low levels of provider training and huge quality gaps. *Health Aff.* 31:2774–84
- Das J, Holla A, Kremer M, Mohpal A, Muralidharan K. 2013. *Quality and accountability in healthcare delivery: evidence from an audit study of healthcare providers in India*. Unpublished manuscript, World Bank, Washington, DC
- Das J, Sánchez C. 2003. *Short but not sweet: new evidence on short duration morbidities from India*. Work. Pap. 2971, World Bank, Washington, DC
- Das J, Sohnesen T. 2005. *Patient satisfaction, doctor effort and interview location: evidence from Paraguay*. Policy Res. Work. Pap. 4086, World Bank, Washington, DC
- Das J, Sohnesen T. 2007. Variations in doctor effort: evidence from Paraguay. *Health Aff.* 26:w324–37
- Donabedian A. 2005. Evaluating the quality of medical care. *Milbank Q.* 83:691–729
- Doshi S. 2010. *Trust me, I'm a doctor*. Master's Thesis, Woodrow Wilson Sch., Princeton Univ., Princeton, NJ
- Dulleck U, Kerschbamer R. 2006. On doctors, mechanics and computer specialists: the economics of credence goods. *J. Econ. Lit.* 44:5–42
- Filmer D, Hammer JS, Pritchett LH. 2000. Weak links in the chain: a diagnosis of health policy in poor countries. *World Bank Res. Obs.* 15:199–224
- Gertler PJ, Vermeersch C. 2012. *Using performance incentives to improve health outcomes*. Policy Res. Work. Pap. 6100, World Bank, Washington, DC
- Goldberg PK. 1996. Dealer price discrimination in new car purchases: evidence from the Consumer Expenditure Survey. *J. Polit. Econ.* 104:622–54
- Hanna RN, Linden LL. 2012. Discrimination in grading. *Am. Econ. J. Econ. Policy* 4:146–68
- Kohn LT, Corrigan JM, Donaldson MS, eds. 2000. *To Err Is Human: Building a Safer Health System*. Washington, DC: Natl. Acad.
- Leonard KL. 2008. Is patient satisfaction sensitive to changes in the quality of care? An exploitation of the Hawthorne effect. *J. Health Econ.* 27:444–59
- Leonard KL, Masatu MC. 2005. The use of direct clinician observation and vignettes for health services quality evaluation in developing countries. *Soc. Sci. Med.* 61:1944–51
- Leonard KL, Masatu MC. 2006. Outpatient process quality evaluation and the Hawthorne effect. *Soc. Sci. Med.* 63:2330–40
- Leonard KL, Masatu MC, Vialou A. 2007. Getting doctors to do their best: the roles of ability and motivation in health care. *J. Hum. Resour.* 42:682–700
- Leonard K, Zivin JG. 2005. Outcome versus service based payments in health care: lessons from African traditional healers. *Health Econ.* 14:575–93
- Lu F. 2014. Insurance coverage and agency problems in doctor prescriptions: evidence from a field experiment in China. *J. Dev. Econ.* 106:156–67
- Maestad O, Torsvik G, Aakvik A. 2010. Overworked? On the relationship between workload and health worker performance. *J. Health Econ.* 29:686–98
- Makinen M, Sealy S, Bitran RA, Adjei S, Munoz R. 2011. *Private health sector assessment in Ghana*. Work. Pap. 210, World Bank, Washington, DC
- Makinen M, Waters H, Rauch M, Almagambetova N, Bitran R, et al. 2000. Inequalities in health care use and expenditures: empirical data from eight developing countries and countries in transition. *Bull. World Health Organ.* 78:55–65

- MEASURE DHS. 2013. *STATcompiler*. MEASURE DHS, Calverton, MD, accessed on September 14. <http://www.statcompiler.com/>
- Miller G, Babiarz KS. 2013. *Pay-for-performance incentives in low- and middle-income country health programs*. NBER Work. Pap. 18932
- Montagu D. 2010. *Private healthcare in developing countries: data by region*. <http://www.ps4h.org/globalhealthdata.html>
- Natl. Cent. Health Stat. 2009. *National Ambulatory Medical Care Survey: 2009 summary tables*. Ambul. Hosp. Care Stat. Branch, Natl. Cent. Health Stat., Hyattsville, MD. http://www.cdc.gov/nchs/data/ahcd/namcs_summary/2009_namcs_web_tables.pdf
- Peabody JW, Luck J, Glassman P, Dresselhaus TR, Lee M. 2000. Comparison of vignettes, standardized patients and chart abstraction: a prospective validation study of three methods for measuring quality. *JAMA* 283:1715–22
- Rethans JJ, Gorter S, Bokken L, Morrison L. 2007. Unannounced standardized patients in real practice: a systematic review of the literature. *Med. Educ.* 41:537–49
- Rethans JJ, Sturmans F, Drop R, van der Vleuten C, Hobus P. 1991. Does competence of general practitioners predict their performance? Comparison between examination setting and actual practice. *BMJ* 303:1377–80
- Rex DK, Hewett DG, Raghavendra M, Chalasani N. 2010. The impact of video recording on the quality of colonoscopy performance: a pilot study. *Am. J. Gastroenterol.* 105:2312–17
- Schultz TW. 1964. *Transforming Traditional Agriculture*. New Haven, CT: Yale Univ. Press
- Schultz TW. 1979. Nobel lecture: the economics of being poor. *J. Polit. Econ.* 88:639–51
- Schwartz A, Weiner SJ, Binns-Calvey A. 2013. Comparing announced with unannounced standardized patients in performance assessment. *Joint Comm. J. Qual. Patient Saf.* 39(2):83–88
- Sounart A. 2011. *AMA postpones decision on secret shoppers*. Rep., AMN Healthcare, San Diego, CA. <http://www.amnhealthcare.com/latest-healthcare-news/ama-postpones-decision-secret-shoppers>
- Sudhinaraset M, Ingram M, Lofthouse HK, Montagu D. 2013. What is the role of informal healthcare providers in developing countries? A systematic review. *PLoS ONE* 8:e54978
- Sylvia S, Shi Y, Xue H, Tian X, Wang H, et al. 2013. *Survey using incognito standardized patients shows poor quality care in China's rural clinics*. Unpublished manuscript, Stanford Univ., Stanford, CA
- World Bank. 2011. *Service delivery indicators: pilot in education and health care in Africa*. Rep., World Bank, Washington, DC
- World Bank. 2013a. *Can public service systems deliver effective social services? Human resource management practices in India's public health sector*. Rep., World Bank, Washington, DC
- World Bank. 2013b. *Education and health services in Kenya: data for results and accountability*. Rep., World Bank, Washington, DC



Contents

Probabilistic Expectations in Developing Countries <i>Adeline Delavande</i>	1
Ill-Posed Inverse Problems in Economics <i>Joel L. Horowitz</i>	21
Financing Old Age Dependency <i>Shinichi Nishiyama and Kent Smetters</i>	53
Recent Developments in Empirical Likelihood and Related Methods <i>Paulo M.D.C. Parente and Richard J. Smith</i>	77
Belief Elicitation in the Laboratory <i>Andrew Schotter and Isabel Trevino</i>	103
Models of Caring, or Acting as if One Cared, About the Welfare of Others <i>Julio J. Rotemberg</i>	129
Exchange Rate Stabilization and Welfare <i>Charles Engel</i>	155
Copulas in Econometrics <i>Yanqin Fan and Andrew J. Patton</i>	179
Firm Performance in a Global Market <i>Jan De Loecker and Pinelopi Koujianou Goldberg</i>	201
Applications of Random Set Theory in Econometrics <i>Ilya Molchanov and Francesca Molinari</i>	229
Experimental and Quasi-Experimental Analysis of Peer Effects: Two Steps Forward? <i>Bruce Sacerdote</i>	253
Coordination of Expectations: The Eductive Stability Viewpoint <i>Gabriel Desgranges</i>	273

From Sudden Stops to Fisherian Deflation: Quantitative Theory and Policy <i>Anton Korinek and Enrique G. Mendoza</i>	299
China's Great Convergence and Beyond <i>Kjetil Storesletten and Fabrizio Zilibotti</i>	333
Precocious Albion: A New Interpretation of the British Industrial Revolution <i>Morgan Kelly, Joel Mokyr, and Cormac Ó Gráda</i>	363
Disclosure: Psychology Changes Everything <i>George Loewenstein, Cass R. Sunstein, and Russell Golman</i>	391
Expectations in Experiments <i>Florian Wagener</i>	421
Optimal Design of Funded Pension Schemes <i>Lans Bovenberg and Roel Mehlkopf</i>	445
The Measurement of Household Consumption Expenditures <i>Martin Browning, Thomas F. Crossley, and Joachim Winter</i>	475
Empirical Revealed Preference <i>Ian Crawford and Bram De Rock</i>	503
Quality of Primary Care in Low-Income Countries: Facts and Economics <i>Jishnu Das and Jeffrey Hammer</i>	525
The Endowment Effect <i>Keith M. Marzilli Ericson and Andreas Fuster</i>	555
Decentralization in Developing Economies <i>Lucie Gadenne and Monica Singhal</i>	581
Local Labor Markets and the Evolution of Inequality <i>Dan A. Black, Natalia Kolesnikova, and Lowell J. Taylor</i>	605
People, Places, and Public Policy: Some Simple Welfare Economics of Local Economic Development Programs <i>Patrick Kline and Enrico Moretti</i>	629
Applying Insights from Behavioral Economics to Policy Design <i>Brigitte C. Madrian</i>	663
The Economics of Human Development and Social Mobility <i>James J. Heckman and Stefano Mosso</i>	689
Firms, Misallocation, and Aggregate Productivity: A Review <i>Hugo A. Hopenhayn</i>	735
Endogenous Collateral Constraints and the Leverage Cycle <i>Ana Fostel and John Geanakoplos</i>	771

Teacher Effects and Teacher-Related Policies	
<i>C. Kirabo Jackson, Jonah E. Rockoff, and Douglas O. Staiger</i>	801
Social Learning in Economics	
<i>Markus Mobius and Tanya Rosenblat</i>	827
Rethinking Reciprocity	
<i>Ulrike Malmendier, Vera L. te Velde, and Roberto A. Weber</i>	849
Symposium: The Institutional Underpinnings of Long-Run Income Differences	
Institutions, Human Capital, and Development	
<i>Daron Acemoglu, Francisco A. Gallego, and James A. Robinson</i>	875
Growth and the Smart State	
<i>Philippe Aghion and Alexandra Roulet</i>	913
The Causes and Consequences of Development Clusters: State Capacity, Peace, and Income	
<i>Timothy Besley and Torsten Persson</i>	927
Under the Thumb of History? Political Institutions and the Scope for Action	
<i>Abhijit V. Banerjee and Esther Duflo</i>	951
Indexes	
Cumulative Index of Contributing Authors, Volumes 2–6	973
Cumulative Index of Article Titles, Volumes 2–6	976
Errata	
An online log of corrections to <i>Annual Review of Economics</i> articles may be found at http://www.annualreviews.org/errata/economics	



ANNUAL REVIEWS

It's about time. Your time. It's time well spent.

New From Annual Reviews:

Annual Review of Statistics and Its Application

Volume 1 • Online January 2014 • <http://statistics.annualreviews.org>

Editor: **Stephen E. Fienberg**, *Carnegie Mellon University*

Associate Editors: **Nancy Reid**, *University of Toronto*

Stephen M. Stigler, *University of Chicago*

The *Annual Review of Statistics and Its Application* aims to inform statisticians and quantitative methodologists, as well as all scientists and users of statistics about major methodological advances and the computational tools that allow for their implementation. It will include developments in the field of statistics, including theoretical statistical underpinnings of new methodology, as well as developments in specific application domains such as biostatistics and bioinformatics, economics, machine learning, psychology, sociology, and aspects of the physical sciences.

Complimentary online access to the first volume will be available until January 2015.

TABLE OF CONTENTS:

- *What Is Statistics?* Stephen E. Fienberg
- *A Systematic Statistical Approach to Evaluating Evidence from Observational Studies*, David Madigan, Paul E. Stang, Jesse A. Berlin, Martijn Schuemie, J. Marc Overhage, Marc A. Suchard, Bill Dumouchel, Abraham G. Hartzema, Patrick B. Ryan
- *The Role of Statistics in the Discovery of a Higgs Boson*, David A. van Dyk
- *Brain Imaging Analysis*, F. DuBois Bowman
- *Statistics and Climate*, Peter Guttorp
- *Climate Simulators and Climate Projections*, Jonathan Rougier, Michael Goldstein
- *Probabilistic Forecasting*, Tilmann Gneiting, Matthias Katzfuss
- *Bayesian Computational Tools*, Christian P. Robert
- *Bayesian Computation Via Markov Chain Monte Carlo*, Radu V. Craiu, Jeffrey S. Rosenthal
- *Build, Compute, Critique, Repeat: Data Analysis with Latent Variable Models*, David M. Blei
- *Structured Regularizers for High-Dimensional Problems: Statistical and Computational Issues*, Martin J. Wainwright
- *High-Dimensional Statistics with a View Toward Applications in Biology*, Peter Bühlmann, Markus Kalisch, Lukas Meier
- *Next-Generation Statistical Genetics: Modeling, Penalization, and Optimization in High-Dimensional Data*, Kenneth Lange, Jeanette C. Papp, Janet S. Sinsheimer, Eric M. Sobel
- *Breaking Bad: Two Decades of Life-Course Data Analysis in Criminology, Developmental Psychology, and Beyond*, Elena A. Erosheva, Ross L. Matsueda, Donatello Telesca
- *Event History Analysis*, Niels Keiding
- *Statistical Evaluation of Forensic DNA Profile Evidence*, Christopher D. Steele, David J. Balding
- *Using League Table Rankings in Public Policy Formation: Statistical Issues*, Harvey Goldstein
- *Statistical Ecology*, Ruth King
- *Estimating the Number of Species in Microbial Diversity Studies*, John Bunge, Amy Willis, Fiona Walsh
- *Dynamic Treatment Regimes*, Bibhas Chakraborty, Susan A. Murphy
- *Statistics and Related Topics in Single-Molecule Biophysics*, Hong Qian, S.C. Kou
- *Statistics and Quantitative Risk Management for Banking and Insurance*, Paul Embrechts, Marius Hofert

Access this and all other Annual Reviews journals via your institution at www.annualreviews.org.

ANNUAL REVIEWS | Connect With Our Experts

Tel: 800.523.8635 (US/CAN) | Tel: 650.493.4400 | Fax: 650.424.0910 | Email: service@annualreviews.org

