



The World's Largest Open Access Agricultural & Applied Economics Digital Library

This document is discoverable and free to researchers across the globe due to the work of AgEcon Search.

Help ensure our sustainability.

Give to AgEcon Search

AgEcon Search

<http://ageconsearch.umn.edu>

aesearch@umn.edu

*Papers downloaded from **AgEcon Search** may be used for non-commercial purposes and personal study only. No other use, including posting to another Internet site, is permitted without permission from the copyright owner (not AgEcon Search), or as allowed under the provisions of Fair Use, U.S. Copyright Act, Title 17 U.S.C.*

No endorsement of AgEcon Search or its fundraising activities by the author(s) of the following work or their employer(s) is intended or implied.

**Defining Access to Health Care:
Evidence on the Importance of Quality
and Distance in Rural Tanzania**

by

Heather Klemick, Kenneth L. Leonard, and Melkiory C. Masatu

WP 08-03

Department of Agricultural and Resource Economics

The University of Maryland, College Park

Defining Access to Health Care: Evidence on the Importance of Quality and Distance in Rural Tanzania^{*}

April 24, 2008

Heather Klemick^a

Kenneth L. Leonard^b

Melkiory C. Masatu, M.D.^c

Abstract

We examine the implications of health seeking behavior on access to quality health care using a unique dataset that combines a household survey from rural Tanzania with the location and quality of all health facilities available to households. Patients do not always visit the nearest facility, but choose from among multiple facilities, improving the quality of care they receive by bypassing low quality facilities. Recognizing this behavior alters the projected benefits to health interventions, reducing the value of focusing on the staff qualifications and increasing the value of focusing on travel time and the motivation of current staff.

Copyright © 2008 by Heather Klemick, Kenneth L. Leonard, and Melkiory C. Masatu.

All rights reserved. Readers may make verbatim copies of this document for non-commercial purposes by any means, provided that this copyright notice appears on all such copies.

^{*} This work was funded by NSF Grant 00-95235 and The World Bank, and was completed with the assistance of R. Darabe, M. Kyande, S. Masanja, H. M. Mvungi and J. Msolla. We extend our appreciation to the Commission for Science and Technology (COSTECH, Tanzania) for granting permission to perform this research. This paper has benefited from extensive conversations with Jishnu Das, as well as seminar participants at the London School of Hygiene and Tropical Medicine.

^a US Environmental Protection Agency, National Center for Environmental Economics, 1200 Pennsylvania Ave. NW, Washington, DC 20460

^b Department of Agricultural and Resource Economics, University of Maryland, 2200 Symons Hall, College Park MD 20742

^c Centre for Educational Development in Health, Arusha Tanzania

Rural households in developing countries face considerably greater obstacles to obtain health care than urban households because they live further from health facilities and because rural facilities are of lower quality. For these households, access to health care is a matter of both distance and quality. Since the Alma-Ata declaration (World Health Organization 1978), developing countries have focused on expanding the coverage of curative health services. During this expansion of health care services, much of the literature focused on measuring the distance to the nearest facility, and showed that usage of facilities highly correlated with distance (see Stock 1983 and Kloos 1990 for examples). Despite the fact that some parts of Africa are still remote from health facilities, much progress has been made in improving physical access to health care.¹ However, gains in health care outcomes have not followed health infrastructure investments (Filmer, Hammer and Pritchett 2000).

Research has increasingly turned from measuring the distance to a health care provider to measuring the quality of care offered at health facilities. Das, Hammer and Leonard (2008) point out that the quality of care provided at the average health facility in developing countries is low, and much research points to the fact that the quality of care provided to poor and rural populations is lower still.² In this article, we point out that, in the rural areas of developing countries, distance and quality both matter because households do not seek care at the average facility nor at the closest facility, but rather they frequently choose to bypass low quality health facilities in search of higher quality care. Thus, access is not simply a function of the distance to the nearest facility, or of the

¹ By 1992, 93% of the population of Tanzania lived within 10kms of a health facility (Abel-Smith and Rawal 1992).

² See for example, Banerjee, Deaton and Duflo (2004a,b); Chaudhury and Hammer (2004); Das and Hammer (2005, 2007); Das and Sohnesen (2007); Leonard and Masatu (2005, 2007); Wagstaff (2002); Wagstaff et al. (2004); and Fabricant, Kamara and Mills (1999).

quality of care at the nearest or average facility, but of the distances and qualities of **all** facilities within a household's health facility portfolio.

We demonstrate the importance of household behavior on measures of access by comparing the benefits of five policy interventions in terms of the quality of medical providers at the closest facility and at the facility that households choose to visit. We use data on health seeking behavior from a household survey from rural Arusha, Tanzania paired with data on the quality of clinicians at all the health facilities available to these households. Our examination of the data highlights the low levels of competence and lower levels of performance of the health care practitioners posted to these facilities. We show that, while residents live reasonably close to facilities, they have limited access to acceptable care because competent providers are overwhelmingly concentrated in urban and peri-urban areas. The health system fails to provide adequate services to rural communities on multiple counts: almost a quarter of health personnel are absent from their posts, almost no facilities are staffed up to nationally mandated standards, the personnel on duty in these facilities often exhibit poor diagnostic skills, and they also frequently fail to perform up to their skill level. Indeed, while rural Tanzanians do seek modern medical care regularly, the failure of clinicians to appropriately diagnose and treat easily curable illnesses leads to avoidable deaths.³

Against this backdrop, we consider the value of eliminating absenteeism, eliminating vacancies, upgrading the level of staff training, improving the road network and reducing the gap between competence (ability to properly care for patients) and performance (actual quality provided to the average patient) found in government-run

³ A survey in rural Tanzania found that 79% of children who die of malaria do seek care at modern health facilities rather than home care or traditional healers (de Savigny et al. 2004).

facilities. For each of these policies, we examine the gains by four measures of access: (1) the probability of getting the correct diagnosis at the closest facility, (2) the travel time required to reach the closest facility, (3) the probability of getting the correct diagnosis at the facilities patients are most likely to visit, and (4) the average travel time required to reach the facilities patients are most likely to visit. We also define the probability of correct diagnosis two alternative ways: first, using data on clinician competence, and second, using information on clinician performance. We propose that a clinician's competence represents the best quality a patient could expect for any given clinician, and performance represents the most likely level of quality.

The data on health seeking behavior clearly demonstrate that households do not always visit the nearest health facility, but rather frequently bypass that facility and travel to facilities with higher quality. We show that this behavior changes the expected benefits to the health interventions examined. In particular, interventions that focus on improving quality in all health facilities (such as eliminating vacancies or upgrading staffing levels) are less beneficial when patients bypass because their behavior had already improved the quality of care received. On the other hand, policies that reduce travel times (such as improving the road network) lead to significant reductions in costs—even if there is no improvement in quality at any particular facility—because patients spend less to access acceptable care. In addition, our results suggest that policies improving the quality of care provided at a few facilities have greater benefits when we take into account the fact that patients will respond by seeking out these improved facilities.

Figure 1 illustrates these points using the example of competence in diagnosing the causes of infant diarrhea. This map of our research area shows the location and

population of sampled sub-villages, all health facilities, all roads (and road conditions) and clinician competence. A 5 kilometer radius, a commonly used measure of accessibility, is shown around each facility. By this traditional measure, if a household lives within such a circle, it has adequate access. Thus, most rural households in this area have reasonable access to facilities. However, our data demonstrate that not all clinicians can properly diagnose illnesses that they are trained to diagnose. Facilities with at least one staff member competent to diagnose the causes of infant diarrhea are indicated by shaded circles. Clearly, the average rural household has poor access to adequate care once competence is considered.

Figure 1 also illustrates the contrast between two measures of health care access. If patients always visit the nearest facility, then households in locations A and B face a similar (unacceptable) level of access. In this case, the only way to improve access to health care for households in locations A and B is to make sure that the health facilities nearest them have competent providers. On the other hand, if patients know where to find acceptable quality and travel to reach acceptable facilities, then households in location B have better access than households in location A. In addition, improvements in some facilities or reduction in travel costs are both effective means of improving access. Note that households in location A see a significant improvement in access to acceptable care if any of the facilities or the road network in their local area is improved.

The policy interventions that we consider in our simulations reflect important debates in the health literature, which has called attention to travel cost, equipment and supplies, and staffing patterns as important elements of patient access. In particular, there is considerable evidence that patients' ability to travel to medical facilities is governed by

the road infrastructure linking them to urban and peri-urban areas where the best-performing hospitals are often located. Many studies have shown that visits to health centers decline significantly with travel costs.⁴ Villages in Indonesia, the Philippines, Sri Lanka, Vietnam and Morocco participating in rural roads projects reported shorter travel times and higher usage of modern facilities compared to non-project villages (Hettige 2006, van de Walle and Cratty 2002, World Bank 1996).

Absenteeism and vacancies have also received attention as factors contributing to poor health care access in developing countries. A six-country study revealed average absentee rates of 35% among medical staff, considerably higher than the 19% absentee rate among teachers (Chaudhury, Hammer and Kremer 2006). Absentee rates were also higher among doctors compared to other medical staff in all countries in the study. In Bangladesh, high absentee and vacancy rates among medical staff (35% and 26% nationwide, respectively) are widespread (Chaudhury and Hammer 2004). While absenteeism in Bangladesh did not vary significantly with the income level of the region, the vacancy rate does, leaving poorer areas with lower access to medical personnel.

In addition to travel cost and staff attendance, we examine clinician quality. Even if a facility is accessible and the doctor is present, it matters whether the doctor is competent to diagnose a patient's illness. Some studies have considered facility quality as indicated by the presence of staff, equipment and material inputs (e.g., Collier, Dercon and MacKinnon 2002 and Lindelow 2004). In contrast, we focus on aspects of care that are non-contractible and non-tradable. When a facility provides the correct diagnosis but

⁴ The relationship between demand and distance has been a standard feature of health seeking models since at least Acton (1975), and some studies in developing countries have gone so far as to estimate the willingness to pay for health care solely based on variation in travel costs (Dor, Gertler and Van der Gaag 1987; Gertler and Van der Gaag 1990).

does not provide the appropriate medicines to treat the illness, the patient can buy medicine on the market or bribe the doctor to get access to private stores. However, good diagnoses, since they are subject to asymmetric information, cannot be purchased in the market place or secured with bribes (Leonard 2002, 2003). Instead of infrastructure and pharmaceutical stocks, we focus on the quality of diagnosis, following Das and Hammer (2005), Das, Hammer and Leonard (2008) and Leonard, Masatu and Vialou (2007).

Finally, patients' behavior is a critical piece of the puzzle that has been under-examined. Bypassing, in which households pass closer facilities in order to seek care at facilities that are further away, has been documented in a number of developing country settings (Akin and Hutchinson 1999; Leonard, Mliga and Mariam 2002; Gauthier and Wane 2008; Hanson, Yip and Hsiao 2004). There is empirical evidence that households know the quality of both visited and bypassed facilities; households bypass low quality facilities in search of high quality facilities when they suffer from illnesses that are responsive to high levels of quality (Leonard, Mliga and Mariam 2002; Leonard 2007). However, to date, no studies have explicitly incorporated bypassing behavior in measuring access.

In the next section, we examine the data, discuss our measures of competence and performance, and report summary statistics on the quality of clinicians in our sample. In the third section, we estimate a model of patient choice of health facilities. We then turn to the simulation of several hypothetical policy scenarios to examine how access to health care changes under these interventions in the following section. We discuss the results and conclude in the final two sections.

Empirical Background and Data Description

Tanzania spends very little on health care and has too few fully trained medical officers.

At five medical officers per 100,000 people, Tanzania is one of the most poorly served countries in the world (United Republic of Tanzania, 2002). The public health care budget in Tanzania allocated \$1.50 per person to recurrent expenditures and \$2.50 per person to capital development in 2002. Combining these figures with cost sharing revenues and donor funding, the Ministry of Health in Tanzania spends almost \$5 per person annually. Total private and public expenditure is \$12 per person, less than half of the average for sub-Saharan Africa.

Most health care in rural Tanzania is delivered by practitioners with much less training than a full MD. Clinical Assistants (CAs) have an elementary school education and three years of medical training. Clinical Officers (COs) have four years of secondary schooling and two years of medical training. Assistant Medical Officers (AMOs) are clinical officers with two additional years of training. Medical Officers (MOs)—fully trained MDs—have six years of secondary schooling and five years of university-level medical training. These clinicians are trained to treat all of the illnesses that we consider in our competence and performance measures. In addition, in the rural areas, nurses occasionally diagnose patients despite their lack of training for this job. We refer to all types of medical personnel as clinicians.

Four types of organizations provide health services: public, private, nongovernmental organizations (NGOs) and parastatal organizations (publicly owned, independently managed).⁵ Health care is delivered in three basic types of facilities: dispensaries, health centers and district hospitals. Dispensaries deal with ambulatory care;

⁵ In this article, we consider parastatal facilities as private facilities.

maternal and child care and the most common outpatient conditions. They do not admit patients. Few dispensaries have any laboratory facilities. Dispensaries should be staffed with at least one CO and one CA. Health centers provide ambulatory care similar to dispensaries, but they also have a ward for inpatient care. Health centers conduct normal deliveries for pregnant women. They are required to have at least one AMO. There is a laboratory for diagnosis of parasitic diseases such as malaria and intestinal worms, as well as bacterial diseases such as tuberculosis. There are no NGO-operated health centers in our data, though they exist in other areas. Hospitals provide general ambulatory and inpatient care, as well as specialized care such as surgery. Most outpatient care is done by AMOs and COs, but there should always be an MO present. The functions of the three levels of health care facilities (dispensaries, health centers, hospitals) are similar irrespective of ownership.

On the other hand, the organizational structure of these facilities is different. Most NGO dispensaries are supervised by Medical Officers headquartered in Arusha, and these supervisors have full authority over both the facilities and their staff. Public facilities, on the other hand, are supervised from the district headquarters (Monduli, Arumeru or Arusha), but these supervisors have less authority over the staff they supervise, and final authority rests with the Ministry of Health and Social Welfare in Dar es Salaam. Leonard, Masatu and Vialou (2007) point out that this decentralization of decision-making authority has important implications for the gap between competence and performance; clinicians who work in the decentralized NGO facilities examined in this data are much more likely to perform at levels near their competence.

Transportation infrastructure in Tanzania is poor. In 2003, 6,808 of 78,891 kilometers of roadways were paved (CIA 2007). A review of sub-Saharan African transportation infrastructure noted that in 1990, only 24% of Tanzania's main roads were in good condition, falling to 8% for regional roads—considerably worse than the neighboring countries of Malawi, Zambia and Kenya (Platteau 1996). In Arusha, many villages are not connected to any road network at all. Improving transportation networks is a high priority for the central government but is never explicitly considered as a health policy.

The data used in this article come from a 2002 survey in Monduli and Arumeru districts of Arusha that collected data on 4,102 individuals in 529 households from 44 sub-villages in 22 villages. The households in the survey are weighted according to the sampling process and represent a population of approximately 211,157.⁶ Data on illnesses suffered within these households (and the locations where the patients sought care) come from two one-year recall surveys collected in 2002 and 2003. From the households interviewed over this two-year period, 1,345 illnesses resulted in visits to one of the health providers examined in this study.

Data on clinician attendance and competence were gathered from all health facilities that could be accessed by the population in the study region (including urban facilities) in 2002-2003. All health facilities in the study area were visited at least twice, but not all clinicians were seen for each of these visits. We assessed the competence and performance of 106 clinicians at 44 facilities. In this article, competence refers to a clinician's ability to examine and diagnose a case study patient and performance refers to

⁶ This estimate is close to the population of 206,764 reported by the official 2002 census for the study area.

the effort and activities that clinicians undertake with their normal patients, as observed by our research team.

We recorded the locations of all sub-villages, health facilities and roads. The data on roads include measures of road quality based on the research teams' observation in the field (not official statistics or maps). Households in the survey were assigned a travel time to each facility, assuming the fastest available means of transportation for each trip segment.⁷ Travel time does not account for time spent waiting for transport, but since this cost is both large and inevitable, travel time is more accurately thought of as the effort or cost required to reach a facility rather than the actual length of time spent in transit. Travel times and distances to the closest health facilities range from under a minute (0.03 km) to four and a half hours (30 km). Approximately 82.5% of the population lives within 5 km of a health facility.

Absenteeism and Understaffing

Almost all surveyed facilities were understaffed due to absent clinicians, vacant positions, and posts filled with poorly qualified personnel. According to Ministry of Health standards, a dispensary is supposed to be staffed with at least one CO and one CA. We found that 42% of dispensaries had only a CO, and the remainder were only assigned a less well-trained CA. Health centers (used only in the public health system), were supposed to be assigned at least one AMO and four COs, but one of the five health centers was only staffed with a CO, and all of the remaining health facilities had fewer COs than mandated. Vacancy rates were higher among MOs than other clinicians,

⁷ We calculate travel costs from all subvillages to all health facilities using information on road locations and average road speeds using ArcView's Shortest Network Path extension. We assumed travels speeds of 100 km/hr for paved and all-weather roads, 40 km/hr for passable roads, and 5 km/hour for footpaths and tracks, based on the authors' observations.

reflecting the dearth of MDs in Tanzania. All hospitals in our sample had an MO on duty, but most of them had fewer than required and also fell short on other types of clinicians.

Table 1 shows the patterns of absenteeism in our sample. Overall, 73% of assigned and scheduled clinicians were present or ready to work on the day of the unannounced visit. Sixty three percent of clinicians at rural posts were present compared to 74% and 77% in semi-rural and urban facilities. Absenteeism is slightly higher in the public service (29%), but clinicians in both the private and NGO sector were also absent (20 and 27%, respectively). These rates are slightly better than those reported in Bangladesh by Chaudhury and Hammer (2004).

Medical Competence and Performance

Each clinician in the sample was tested for competence using case study patients with an actor (vignettes), and most clinicians were assessed on performance by being observed consulting their regular patients in a similar outpatient setting. Vignettes are purposefully-designed case studies presented to clinicians and paired with a measurement tool that examines the inputs (history taking and physical examination) required by national protocol for that illness. The clinician knows that the patient is an actor, and there is a clinician from the research team present in the room to record his activities. Clinicians were tested for six case studies: malaria with complications, pelvic inflammatory disease (PID), infant diarrhea, pneumonia, the flu and worms.⁸ We also created a seventh case study of standard malaria, using the results of the malaria with complications case. If a clinician diagnosed such a patient with malaria or malaria with

⁸ For 14 of the 106 doctors we only have information on four of these vignettes (the flu and worms cases are excluded).

complications, we determine that he would have correctly diagnosed standard malaria. Each illness is common, and diagnosis is possible without laboratory tests. Specifically, every illness is part of the standard training in protocol for personnel at all facilities in our sample. Since we designed the case study, we tailored the quality measurement tool to fit the protocol required for the illness. For outpatient consultation with actual patients, we used three quality measurement tools designed to fit patients who presented with a fever, cough, diarrhea and a more general measurement tool for other cases.

We can judge whether clinicians gave the correct diagnosis for each vignette because we know the correct diagnosis for each case study. For clinicians' performance with real patients, we cannot assess whether they correctly diagnosed the illness, but since we know what inputs they provided we can predict the probability that they would get the correct diagnosis.

Although vignettes measure competence rather than actual performance and represent an upper bound on competence (Leonard and Masatu, 2005; Das and Hammer, 2005), they are a valuable tool for comparing quality across providers because every clinician in the sample faces exactly the same patients. Observing clinicians in practice is a more accurate representation of the effort that clinicians exert with their regular patients, but it is more difficult to control for the different case mix across clinicians in the sample. We use both measures in our analysis to enhance the robustness of our results and compare their implications.

Creating Competence and Performance Scores

For each illness, we had a list of required protocol items that clinicians did or did not implement for both the case study and with real patients. For a patient presenting with a

fever, for example, the checklist asked, “Did the clinician ask about the pattern of the fever?” or “Did he check the patient’s temperature?” There were 61 such items across the six vignettes, and 41 across the four categories of symptoms observed with actual patients. From these items, we derive a competence and a performance score for each clinician in the sample, assuming that competence and performance are constant across illnesses within each type of score. We derive these two scores independently using the Item Response Theory (IRT) analysis. This methodology was developed for the analysis of education exams (Birnbaum 1967, Bock and Leiberman 1970) and was pioneered in the analysis of medical vignettes by Das and Hammer (2005). We follow the implementation used in Leonard, Masatu and Vialou (2007).

For vignettes, the probability that clinician i implements protocol item j is modeled as a function of that clinician’s competence (A_i), an item-specific discrimination factor (α_j) and an item-specific difficulty parameter (β_j). Discrimination is the degree to which more competent clinicians are likely to implement an item properly, and difficulty is the degree to which all clinicians are likely to implement an item correctly. To solve for all three sets of parameters, we use the logit model as follows:

$$\text{prob}(a_{ij} = 1) = \frac{\exp(\hat{\alpha}_j \hat{A}_i - \hat{\beta}_j)}{1 + \exp(\hat{\alpha}_j \hat{A}_i - \hat{\beta}_j)}.$$

For performance, we implement the same basic procedure, but add information about the characteristics of the patient (Z_k) and the length of time the clinician had been under observation (t). Patient characteristics include whether the patient was an infant or child and whether he or she presented with multiple symptoms. We control for the number of previous consultations observed by the research team because clinicians in this

sample exhibited high but declining quality as they continued to consult in the presence of the research team—the Hawthorne effect.⁹ Thus, we estimate parameters for patient characteristics (ω) and for the Hawthorne effect (δ), as well as for performance (A_i), item-specific discrimination (α_j) and item-specific difficulty (β_j).¹⁰

$$\text{prob}(a_{ijk} = 1) = \frac{\exp(\hat{\alpha}_j(\hat{A}_i + \hat{\delta}t + Z_k\hat{\omega}) - \hat{\beta}_j)}{1 + \exp(\hat{\alpha}_j(\hat{A}_i + \hat{\delta}t + Z_k\hat{\omega}) - \hat{\beta}_j)}$$

This methodology generates one competence and one performance score for each clinician that controls for the fact that some items reveal more information about competence and performance than others (they have a high discrimination score), and for the observable variation in case mix across patients.¹¹

The IRT process does not identify the level or scale for either score, so we use the following process to put these scores on a similar scale. Some of the patients seen in actual practice are similar to three of the vignette case studies (malaria, pneumonia and diarrhea), and many of the protocol items are the same for these cases. On average, clinicians in our sample completed 52.0% of these 26 comparable items on the vignette (with a standard deviation of 17.2%) but only 39.6% of these same items in actual practice (with a standard deviation of 10.0%). We convert the distribution of IRT competence scores and IRT performance scores to match these two distributions,

⁹ Leonard and Masatu (2006) examined this phenomenon and collected additional data demonstrating that doctors increased their effort when the research team first arrived and allowed it to fall as they became accustomed to the research team's presence. By examining the quality of care provided by doctors who were never directly observed by the research team, the authors showed that declining quality was not a feature of the normal practice of doctors but was caused by the presence of the research team.

¹⁰ We use the same notation for quality, discrimination and difficulty in both the competence and performance regressions, but the parameters are estimated independently.

¹¹ The IRT scores are highly correlated with the raw percent of items completed. For performance, the IRT index and raw performance score have a correlation coefficient of 0.87 (p-value<0.001) over 95 observations. For competence, the IRT index and raw score have a correlation coefficient of 0.94 (p-value<0.001) over 106 observations. The list of items and estimated discrimination and difficulty coefficients for the vignette and patient observation checklists are available from the authors upon request.

respectively. These converted scores can be interpreted as the percentage of items completed. Figure 2 shows the distribution of these converted scores over the sample, plotting the competence and performance score pair for each clinician. In addition to the competence-performance score pairs, we show the frontier that represents performance equal to competence. If competence is the maximum level of performance and we have scaled our measures correctly, then performance should always be less than or equal to competence. With the exception of five clinicians, this frontier is a reasonable representation of this concept: most clinicians are either close to the frontier, or well below it.

Figure 2 also indicates each clinician's facility type, either public or non-public (which combines private, parastatal and NGO). The dashed and dotted lines show the overall trends for the relationship between competence and performance for public (dotted) and non-public (dashed) facilities. The difference between the levels of these two lines is significant ($p = 0.011$). The slope of performance with respect to competence is significantly positive ($p=0.010$)—indicating that increased competence is associated with increased performance—but the difference in the slopes is not significant ($p=0.36$). This evidence shows that the average non-public sector clinician implements approximately seven percentage points more items in practice than does a public sector clinician with the same level of competence. Indeed, Leonard, Masatu and Vialou (2007), working with the same data as the present study, showed that autonomy (decentralization of decision-making authority) at non-public facilities encouraged clinicians to out-perform their public-facility peers, even though clinicians at both types of facilities were equally competent.

Clinician Quality Summary Statistics

Clinicians in our sample performed poorly on a variety of measures even when they were assumed to practice at their competence level, as shown in table 2. Clinicians correctly implemented only 48% of items on the checklist for vignettes, and 41% of items when examining an actual patient. Competence varied considerably across the most common illnesses. While clinicians correctly diagnosed malaria, worms, and flu vignettes 82% of the time and pneumonia vignettes 81% of the time, competence fell to a 61% chance of correct diagnosis for pelvic inflammatory disease and 58% for infant diarrhea. Clinicians were least able to diagnose malaria with complications; only 10% correctly diagnosed the case study.

Competence and the Probability of Correctly Diagnosing an Illness

Using the vignette data described above, we estimate a model to determine the effect of clinician and facility characteristics—particularly input provision, as measured by IRT scores—on the probability of correct diagnosis. We posit that clinicians differed in the probability that they would give the correct diagnosis because (1) they differed in the use diagnostic inputs designed to differentiate illnesses with common presenting symptoms, (2) they differed in the degree to which they were trained to recognize the distinctive characteristics of illnesses, (3) they differed in the types of patients they normally see and therefore their experience with particular types of illness, (4) they differed in their overall experience or ability and (5) good (or bad) luck. Thus, we model the probability of giving the correct diagnosis as a function of input provision (given by the competence scores), years of training, workplace environment (location [rural, urban], type of facility [hospital, dispensary] and facility ownership [public, NGO, private]), a clinician-specific effect (clinician random effect) and error (or luck).

Table 3 shows the determinants of correct diagnosis in our sample. Column 1 shows the results of a random effects logit regressing a dummy variable indicating whether the vignette diagnosis was correct on the IRT competence score, years of training, and dummy variables for each illness and for facility location, type and ownership. Each clinician was observed for at least five vignettes, and most were observed for seven, allowing us to include random effects to control for omitted variables that could be correlated with the regressors and the probability of correct diagnosis.

We find that the IRT competence score is highly significant, meaning that clinicians who ask questions and examine the patient are more likely to get the correct diagnosis. Training is not significant after we control for input provision; clinicians with more training diagnose more accurately because they do more. Clinicians in urban facilities are less likely to correctly diagnose an illness (after controlling for input provision). The clinician random effect is significant, suggesting that even after controlling for input provision, training, location, facility type and ownership, some clinicians are simply better than others, for reasons we cannot observe.

Column 2 estimates the same equation as column 1 using a linear probability model (with clinician random effects) rather than a logit. This model allows us to recover each clinician's mean random effect. Because the clinician random effect remains significant even after controlling for competence, training and work environment, we include it in our analysis of the illness-specific diagnoses. Columns 3 through 7 examine the impact of competence, training, work environment and the clinician random effect on the diagnosis of malaria with complications, pelvic inflammatory disease, infant diarrhea, pneumonia and the combined cases of uncomplicated malaria, worms and the flu. We

combine these three illnesses into one category because most clinicians correctly diagnose these illnesses, and in the individual regressions, only the clinician random effect is significant.¹² The insignificant coefficient on competence indicates that these three illnesses require little skill to correctly diagnose. For pelvic inflammatory disease, only the clinician random effect is a significant determinant of correct diagnosis, but we do not pool PID with malaria, worms and the flu because diagnosis is not easy (clinicians have a lower rate of correct diagnosis), and it does matter that the patient receive appropriate care, not just the default medicine. Malaria with complications, infant diarrhea and pneumonia show a different pattern: both competence and the clinician random effect are important determinants of the correct diagnosis, though not training or the work environment.

From the regressions in shown in columns 3 through 7, we can derive the predicted probability of correct diagnosis for five illness categories. Assuming clinicians work at their competence level, we can use the IRT competence score. If instead we assume clinicians practice at the level observed with actual patients, we can replace the competence score with the performance score, holding the coefficient constant, to derive the predicted probability of correct diagnosis with real patients.

Model of Health Facility Choice

In this section, we examine how clinician quality and other facility characteristics affect patients' decisions about which health facility to visit. In order to assess whether changes to the health care system are likely to improve access, it is important to consider patients' behavior: Do they bypass, and why?

¹² We reject the hypothesis that the coefficients for each of the three illnesses individually are significantly different from each other ($p=0.279$).

To answer this question, we use data collected from households about illness episodes, the symptoms experienced and the first facility visited for that illness. Unlike the case study patients, we do not know what illness the patient suffered from.¹³ Of the full set of 2,220 recorded illnesses, 1,345 resulted in visits to one of the health facilities examined in our study. All visits are first visits; referrals and follow-ups are not included. The remaining illnesses resulted in not seeking care or using traditional remedies (28%), visiting a pharmacist (6%), visiting a traditional healer (1%) or visiting a facility outside the study area (2%). In this analysis we examine the choice of facility conditional on the choice to visit a modern facility in the research area. We observed patients visiting 35 of the 44 facilities in our sample.

Our survey asked patients to describe the symptoms of the illness, the length of time they had been sick, the number of days bedridden (if any), and whether or not they could perform a series of activities before and during the illness. Clinicians from the region evaluated this information and graded each illness by the following criteria (on a scale of 1 to 10):

Responsiveness to effort: The degree to which more effort in examination (by clinicians) can improve the chances of a successful outcome.

Responsiveness to skill available at a dispensary: The degree to which low levels of training and access to equipment are adequate to properly diagnose and treat the illness.

Responsiveness to skill available at a hospital: The degree to which training and access to better laboratories or other equipment can improve the chances of a successful outcome.

¹³ Patients often know the diagnosis that they were given by the doctor, but if doctors routinely misdiagnose patients, this information is of little use to the econometrician.

Chance of a successful outcome with the best possible care: The degree to which a patient is likely to fully recover if a clinician provides all necessary effort and has all necessary skill.

Chance of a successful outcome with poor quality care: The degree to which a patient is likely to recover if a clinician provides no effort or has no skill.

Thirty-seven clinicians examined the full set of illnesses, and each illness was coded by at least three different clinicians. For the purposes of this analysis, we examine three scores derived from the scores above: (1) the responsiveness to effort, (2) net value of skill, defined as the net gain from skill available at a hospital over the skill available at a dispensary, and (3) net value of quality, defined as the net gain in the chances of a cure from good quality care over poor quality care. Since the illnesses were randomly assigned to clinicians for coding, we aggregate the scores by standardizing these three scores for each coder and then averaging over all coders. Thus, rather than examining the seven diseases presented in the case studies, we use three continuous variables describing the characteristics of each illness. Importantly, these characteristics are assigned by clinicians working on the research project, not by patients. By using these characteristics to model the choice of health facility, we imply that patients understand the importance of these characteristics.

We expect patients to prefer facilities with higher quality care and to differentially prefer quality when they suffer from illnesses that are more responsive to quality. The quality of care and the level of training are represented as averages across all clinicians at the facility. We expect patients to avoid traveling long distances, and therefore seek out closer facilities. In addition, since costs vary across public, private and NGO facilities,

patients are likely to prefer less expensive facilities if they have the same level of quality and are the same distance as other facilities. In our specification, we include a dummy variable for each type of facility (owner and type) to capture the average difference in expected costs, as well as any systematic differences between these types of facilities. In addition, patients should prefer facilities where clinicians are less likely to be absent. Absenteeism is represented by a variable indicating the percentage of clinicians present during each of the site visits. If patients know where clinicians are more likely to be absent, they could use this information in their choices. A positive coefficient indicates that patients avoid absenteeism.

As quality measures, we include the average competence score, the average performance score directly, and the average clinician random effect (from the linear probability model shown in column 2 of table 3). In addition, we include the average performance score interacted with responsiveness to effort, as well as the average competence score interacted with responsiveness to effort, net value of skill and net value of quality for each of 44 dummy variables representing the 44 sub-villages. This last variable is meant to capture the variance across sub-villages in the willingness to pay for quality, conditional on the return to skill, reflecting the fact that average income and wealth (or labor market opportunities) could vary significantly across sub-villages in the sample. We also include the average level of training interacted with the net return to skill, and a dummy variable for whether the facility is a hospital interacted with the net return to skill. Travel costs are represented by travel time (in minutes), squared and cubed terms and a dummy variable for whether a particular facility was the closest facility to the sub-village.

Conditional Logit Model Results

Table 4 shows the coefficients for latent utility in a conditional logit model of facility choice over the 35 sampled health facilities visited by patients in the household survey. Overall, households behaved as if they preferred facilities that are closer and house more competent clinicians. In addition, they preferred hospitals when they suffered from illnesses with a high net gain to skill. (Note that a direct preference for hospitals would enter through the dummy variables for each type of facility). Performance was less important on the margin, though a test that the coefficients for both performance and performance interacted with the return to effort are jointly equal to zero is rejected (p-value 0.06). The clinician random effect had a negative effect on patient choice, suggesting that some feature of clinicians unassociated with training, location or effort leads to above average diagnostic ability and below average demand for services.

The dummy variables for each type of facility reflect the differences in costs as well as any differences in average quality. In this sample, public facilities had lower performance than all other facilities, but did not have lower levels of competence. They were also much less expensive than other facilities. Thus, these dummy variables reflect a combination of the value of higher performance and lower fees. The negative (though not significant) coefficient on average performance does not mean that patients did not value performance, only that after controlling for competence and facility type (which explains most of the variance in performance), the marginal impact of performance is not significant. The coefficient for less absenteeism is positive but not significant. Unsurprisingly, patients preferred to travel less, as indicated by the negative coefficient on travel time. In addition, patients tended to visit the closest facility, even when controlling for travel time.

This model of health seeking behavior attempts to predict household behavior and to allow us to measure the impact of quality changes at any health facility on access to health care. The many interaction terms included in the model to capture the impact of quality on facility choice reduce the t-statistics for several of the variables due to collinearity, but the model captures behavior reasonably well, as seen in the high correspondence between the characteristics of the facilities that patients actually chose and the facilities the model predicts they would choose (discussed below).

Health Seeking Behavior and Bypassing

In only four of the 44 sub-villages did all of the households in our sample visit the same facility. In one sub-village, households visited 9 different facilities, and on average households from each sub-village visited 4.24 facilities. Thus, it is important to model the portfolio of facilities visited by households, not just to describe the one closest or most likely to be visited by a sub-village. The overall significance of the interactions with illness characteristics included in table 4 demonstrates the value of multiple health facilities; patients did not always choose one facility, but varied their choice according to the illness from which they suffered.

The first three columns of table 5 compare the travel time and clinician characteristics at the facilities closest to patients, the facilities patients actually visited, and the characteristics of all facilities weighted by the predicted probability that the patient would visit that facility. Predicted facility characteristics represent a view of all health care available to the average patient weighted by the health seeking behavior described in the conditional logit model.

Overall, the predicted facility characteristics are similar to those at the facility chosen. Patients received slightly lower competence and performance and travel slightly farther than predicted by the model, but the difference is small in magnitude (though it is statistically significant for travel time, years of training and performance score). However, characteristics at the nearest facility were on average quite different than those at predicted facilities. We expect patients on average to travel 6 minutes (22%) more and to see clinicians with an additional half year of training who were significantly more competent and better performing than those at the nearest facility. Thus, the baseline results suggest that once patient behavior is taken into account, patients routinely accessed better quality care than was available nearby. In the following section, we investigate the implications of patient behavior on health care access under a variety of hypothetical policy improvements.

Policy Scenarios and Resulting Changes in Access

In this section, we compare the travel cost and clinician quality patients would encounter at the closest facility and the facility they are most likely to visit under current staffing conditions and five hypothetical policies.

Baseline: includes clinicians present at the time of the survey in 2003, including clinicians observed during other site visits but not scheduled to be on duty during the survey.

Eliminating absenteeism: The first policy scenario assumes that absent clinicians return to work. Some of the clinicians who were absent during the 2003 survey were seen during other site visits, but other absent clinicians were never observed by the research

team, and we impute their quality based on clinicians of the same cadre and facility location, type and ownership.

Eliminating vacancies: The second scenario assumes that absenteeism is eliminated and that all vacant posts are filled. Therefore, in this scenario all posts are staffed according to current government guidelines.

Upgrading staff qualifications: For this scenario, there is no absenteeism, and all dispensaries are staffed with an AMO and a clinical officer, and all health centers and hospitals are staffed with a MO and an AMO.

Road upgrade: The fourth scenario uses baseline staffing patterns but reduces travel times between patients and facilities by assuming that passable roads are upgraded to all-weather, and that tracks and footpaths (including foot traffic not represented on the maps) become passable roads (see figure 1).

Public Sector Reform: This scenario combines a reduction in absenteeism with an increase in the performance of public sector clinicians so that they practice at levels similar to those of non-public sector clinicians. We achieve a partial elimination of absenteeism by supposing that every clinician ever observed at their post is present; we assume that it is easier to get these clinicians to show up at their posts than it is for clinicians whom we never observed. We simulate an increase in performance by assuming that every public sector clinician performs seven percentage points more items, per the observed difference in public and non-public sector clinician performance (holding competence constant) indicated in figure 2.

Simulating Facility Characteristics under Hypothetical Scenarios

In order to predict the average quality of care that patients would encounter at any facility under these scenarios, we need to simulate the qualities of new clinicians (for eliminating absenteeism, vacancies or upgrading facilities), improve the qualities of existing clinicians (for public sector reform), or change travel costs (for upgrading roads).

To examine the impact of policies targeting absenteeism, vacancies and staff qualifications, we simulate the quality of new clinicians who would replace or augment the clinicians already practicing. We assign each simulated clinician a competence, performance, level of training, work environment and clinician random effect in the following manner. We predict competence and performance from a regression of these scores on the observable characteristics of studied clinicians: years of training and work environment.¹⁴ The policies we examine determine the years of training and work environment variables. The clinician random effect is, by design, independent of any observable characteristics of clinicians; therefore we assign every simulated clinician a random draw from the sample of observed clinician random effects estimated in the random effects linear probability model (column 2 of table 3). Using work environment, clinician random effect and predicted competence and performance, we predict the expected probability of a correct diagnosis for all simulated clinicians under the assumption that they worked at their competence level and, separately, that they worked at their observed performance level using the coefficients estimated from columns 3-7 in table 3.

To obtain reasonable confidence intervals for the five scenarios, we bootstrap the whole process for determining quality 5,000 times. First, we rerun the random effects

¹⁴ These regression results are available from the authors upon request.

linear probability model (column 2 of table 3), sampling with replacement from clinicians observed in each type of work environment (urban/rural, hospital/dispensary, public/NGO/private). Second, using this bootstrapped sample and the clinician random effects, we rerun each of the five logit regressions on the probability of correct diagnosis for malaria with complications, pelvic inflammatory disease, diarrhea, pneumonia and malaria/worms/flu (columns 3-7 of table 3)¹⁵ Third, we re-estimate the relationship of observed competence and performance on training and work environment for clinicians in the bootstrapped sample. Fourth, we draw randomly from the sample of clinician random effects for all simulated clinicians and for observed clinicians who are not in the bootstrapped sample. Fifth, we determine the predicted probability of correct diagnosis for all clinicians (in and out of the bootstrapped sample). Thus, the sample of clinician (observed and simulated) never changes, but the qualities assigned to these clinicians does change with the sample used in the bootstrap.

To simulate the results of reducing absenteeism and vacancies and upgrading qualifications, we add new clinicians to the staff and recalculate the average clinician quality at each facility.¹⁶ To simulate the road building scenario, we assume that road speeds increase for all tracks, footpaths and passable roads (from 5 km/hour to 40

¹⁵ For malaria with complications, some of the bootstraps have no correct diagnoses among doctors at certain types of facilities. This occurs alternately at rural facilities, private facilities, NGO facilities or dispensaries, 294 times out of 5,000 bootstraps, and in these cases we set all diagnoses to wrong in the relevant type of facility and solve the logit model only for remaining facilities. For pneumonia, some of the bootstraps have all correct diagnoses among (alternately) NGO or private facilities. This occurs 1,278 times out of 5,000 bootstraps, and in these cases we set all diagnoses to correct for these types of facilities and solve the logit model for only the remaining facilities. In 11 draws, doctor random effect perfectly predicts whether a doctor gets the correct diagnoses, and it is not possible to implement a logit model for any observations. In these cases we drop the draw and replace it with another random draw.

¹⁶ As vacancies are filled or staff are upgraded, we keep the number of staff assigned to each facility at the government-mandated level by dropping the least qualified doctor when a new doctor is posted. In most cases, this results in dropping the less qualified staff, but, in order to prevent facility quality from falling under any scenario, if the new doctor is of lower quality than an existing doctor, the new doctor is dropped; we assume that no highly-skilled doctors are ever removed from facilities.

km/hour for tracks and paths and from 40 km/hour to 100 km/hr for passable roads). We then recalculate the shortest distance to all facilities and report average clinician quality at these facilities under current staffing conditions. For the public sector reform scenario, we eliminate absenteeism at public sector facilities, assume that clinician performance increases by 7 percentage points at public facilities and then recalculate average clinician performance.

In addition to calculating clinician quality and travel costs assuming that patients visit the closest facility, we also calculate these measures of access taking into account patient behavior for each policy scenario. Under each of the five interventions, a patients' probability of visiting a particular facility changes based on the changes in all facility characteristics attributable to the policy. We use the estimated coefficients from table 4 to predict the probability of visiting every facility in the sample and calculate the weighted average of facility characteristics.

Note that we do not allow for any changes in the value of the dummy variables or the designation of a facility as a hospital, NGO, private, etc. Thus, to the extent that certain features of clinicians are consistent across types of facilities, these will not change in our model. For example, even when we assign an MO to a health center, it does not become a hospital, and patient preferences for hospitals do not change. In addition, since the conditional logit regression models the decision of where to seek health care conditional on seeking health care at one of the providers in the sample, we are not looking at the extensive margin; no new patients now choose to seek care.

Effect of Hypothetical Policies on Facility Characteristics

Here we turn to the effect of the policy scenarios on facility characteristics. Each of the five policy scenarios discussed above is represented in table 5. We consider how these policies affect clinician quality and travel times faced by patients. In particular, we test whether each of these policies changes access relative to the baseline. For each policy considered, we compare the characteristics of the closest facility with the intervention to the characteristics at the closest facility in the baseline, and we compare the characteristics of the predicted facility with the intervention to the characteristics of the predicted facility in the baseline.¹⁷

Eliminating absenteeism leads to a small but statistically significant boost in clinician training, competence and performance at predicted and chosen facilities. It also slightly decreases travel times to predicted facilities. It does not affect travel time to the closest facility (nor do any of the other policies considered other than rural road upgrades). Eliminating vacancies has a larger impact on average clinician training and competence at both the closest and the predicted facilities. Notably, patients now travel farther but reach facilities with more qualified providers, as indicated by the increase in travel time above the baseline. Clinician training, competence and performance improve even more substantially when staff qualifications are upgraded. Patients encounter clinicians with more than two years of additional training relative to the baseline, whether they visit the closest or the predicted facility, and these clinicians have a higher skill level and perform better in practice. Similar to the eliminating vacancies scenario, travel time

¹⁷ The significance of these comparisons is indicated in the table with asterisks, and the use of parentheses indicates a fall from the baseline. Thus, for example, travel time to the predicted facility rises significantly when we eliminate vacancies, indicated with ***.

to the predicted facility increases, reflecting the fact that patients have greater reason to travel past their nearest facility.

Upgrading rural roads reduces average travel times to the closest and predicted facilities by a substantial amount—by over 11 minutes in both cases. It also slightly but significantly improves clinician quality (by training, competence and performance measures) at predicted facilities, indicating that patients make use of reduced travel times to seek out superior facilities. The magnitude of the quality changes is lower than in the facility upgrade scenario but similar to the policies addressing absenteeism and vacancies. Surprisingly, quality falls by all measures at the closest facility. This result arises because the cases in which road upgrades change the identity of the closest facility are facilities at the end of poor roads—generally lower quality facilities.

Public sector reform (reducing absenteeism and improving performance in public facilities) also leads to clinician quality improvements. Average performance rises at both closest and predicted facilities, but the gain is much greater at predicted facilities. In fact, this scenario leads to the largest changes in performance of all the scenarios examined. Similar to the vacancy and staff upgrade scenarios, travel times increase slightly as patients bypass nearby facilities to seek out better performing clinicians.

Effect of Views of Patient Behavior on the gains to policies

Table 5 also allows us to compare whether the view of patient behavior affects the gains to difference policy scenarios and in particular whether the gains are larger when we take into account actual patient behavior. For example, when facilities are upgraded, the gain in years of training from the baseline for predicted behavior (6.58-4.55), though

positive, is significantly smaller than the gain when we assume patients visit the closest facility (6.54-4.00).¹⁸

Comparing the results for facility upgrades to those for road improvements illustrates the importance of understanding how patients choose health facilities. For upgrading facilities and reducing vacancies, assuming that patients visit the closest facility results in overestimates of the gain in travel time, training, competence and performance if patients bypass. On other hand, for road improvements, assuming that patients visit the closest facility leads to underestimates of the gains in all of these measures. The pattern is not as clear for eliminating absenteeism, though the magnitude of the changes is very small, so the significance of these results has less meaning. Under public sector reform, the increase in travel time is greater when patients bypass (it is zero when patients do not bypass), but bypassing leads to larger gains in training, competence and performance relative to visiting the closest facility.

The Impact of Policies by Type of Illness

Table 6 shows the probability of correct diagnosis for the specific illnesses included in the vignettes: malaria with complications, pelvic inflammatory disease, infant diarrhea, child pneumonia, and the combined cases of malaria, worms, and flu. This table addresses the question, “Do any of these policies lead to real changes in outcomes?” Since we do not model how the behavior of patients **with these illnesses** would differ from that of the average patient, this table does not predict what would happen to a patient with these illnesses—their choices might be considerably different from those of the average patient. The characteristics of facilities are used to simulate what would

¹⁸ If the gains to any given policy are larger when patients bypass than when they visit the closest facility, we indicate significance with the plus symbol, +. If the gains are smaller when patients bypass than when they visit the closest facility, we indicate significance with the plus symbol in parentheses (+).

happen if a patient with one of these illnesses visited the average facility visited by all types of patients observed in our data.¹⁹ Similar to table 5, we compare diagnostic outcomes when patients visit the closest facility with those when they visit the facility predicted by the conditional logit model. In addition, we report diagnostic outcomes based on clinician competence and performance for each illness under each policy scenario.

The standard errors in this table are much larger, reflecting the fact that we cannot perfectly predict whether a clinician would correctly diagnose any given illness; the clinician random effect has a large impact, and because it is randomly distributed, it generates noise. In addition, in columns 3 and 6 of table 3, the coefficients on training for malaria with complications and pneumonia have large variance which is reflected in the variance in this table. Nonetheless, there are some important patterns in the data.

The gains to eliminating vacancies and upgrading facilities are reflected in significantly higher probability of correct diagnoses for most illnesses, although the gains for illnesses like malaria, worms or the flu are relatively small. On the other hand, upgrading roads has few benefits in terms of increased probability of a correct cure. Recall however, that it has a large impact on travel time; patients are traveling much shorter distances to reach the same quality of care. Public sector reform has a smaller impact on illnesses, and except for pelvic inflammatory disease, the gains only occur if we assume that medical providers work at their level of performance. Conversely, for all other policies, the gains are smaller if clinicians work at their level of performance than if they work at their level of competence.

¹⁹ The conditional logit model controls for characteristics of patient illnesses, but we do not map these characteristics to the illnesses examined with vignettes, and therefore the number shown correspond to the average illness reported in the household data.

Similar to table 5, when we compare the gains at the closest and the predicted facility, eliminating vacancies and upgrading facilities look better if patients visit the closest facility (particularly for pelvic inflammatory disease and child pneumonia), and public sector reform looks better when patients bypass (especially for infant diarrhea and child pneumonia).

In general, the results in tables 5 and 6 show that the gains realized from policy interventions differ according to our assumptions about health seeking behavior, but not in the same manner for all policy interventions. In particular, public sector reform looks much better when we take into account bypassing, whereas eliminating vacancies and upgrading facilities look worse. While public sector reform is not better than upgrading facilities or eliminating vacancies for any of the specific illnesses considered, the relative gains from policy are larger.

Discussion

These simulations suggest certain policies that have the highest potential for improving access to health care for rural residents of Arusha, Tanzania. They do not offer a general prescription for how to improve rural health care in developing countries; rather, they highlight the importance of accounting for variations in staff quality, travel costs, and performance incentives and, in particular, for how patients respond to these factors when choosing which facility to visit.

Care should be taken in the interpretation of these results because of the nature of absenteeism in this area and the fact that we do not take into account any changes in caseload that may result from these policy changes. In our data, clinicians who are occasionally absent are not better than clinicians who are present at work, so we assume

that chronically absent clinicians are also not better than those who are present. Thus, bringing back these clinicians has few benefits. This result may not be applicable to all developing country settings for at least two reasons. First, there are limited private sector opportunities for underqualified clinicians in Tanzania, in contrast to settings where absent staff are actually the best clinicians who neglect their posts to work in private practice.

Second, where reducing absenteeism increases the number of clinicians present but does not affect whether the facility is open for business (say the number goes from 1 to 2, or from 2 to 3), our data do not show an increase in quality. This is reasonable, as patient loads in Arusha are relatively low; in almost all facilities, clinicians are able to see every patient who visits on a given day (though wait times may be longer). However, if only one clinician were present and patients were turned away, eliminating absenteeism could improve quality even if the returned clinician was worse than the one who was always present. In our study area, high caseloads do not decrease the quality of clinicians who are present. If we regress performance on the number of patients seen on the day of the visit, the coefficient on caseload is negative, but not significant ($p\text{-value} = 0.833$). If we include average caseload in the conditional logit model for facility choice, the coefficient is positive (and significant), suggesting that patients prefer longer waiting times, all else equal. Naturally, this does not mean that patients prefer to wait; this result reflects the fact that certain facilities are more popular and therefore more crowded. We do not model changes in caseload with our policy scenarios, so we do not include this variable in the model reported in table 4. Bypassing clearly increases the potential for overcrowding, and in many cases, it may be undesirable because patients who need

primary care bypass the appropriate facilities causing inefficient use of higher order facilities. However, bypassing is an empirical reality, not a policy intervention. Indeed, Darkaoui et al. (1999) and London and Bachmann (1997) point out that many patients who bypass “appropriate” facilities to seek care at higher levels do so because of inadequate care at these same “appropriate” facilities. Since we see no reasonable evidence of avoiding overcrowding in our data, we do not model how patient choices would change with overcrowding. This assumption, however, may not apply in all settings. We note that public sector reform increases visits to the appropriate facilities, and this result is more likely to apply generally.

Although we compare the overall outcomes predicted in the various scenarios, the scenarios do not reveal an “optimal” policy based on a cost-effectiveness analysis because we do not estimate costs. We intend to show that the benefits of a policy depend on our understanding of patient behavior, and we have shown that this behavior significantly alters the gains to different policies, but it does not alter the costs. Estimating precise costs for these policy scenarios is difficult for a variety of reasons. In particular, eliminating absenteeism and reforming public sector facilities carry no clear financial costs. Rather, reforming public sector institutions is primarily a matter of political will. Reducing absenteeism has been shown to be difficult in a variety of context, but targeted and relatively low-cost interventions such as photographing teachers (Duflo, Hanna and Ryan 2007) hold some promise if they can be scaled up and adapted to the health sector. Increasing the performance of public sector clinicians by seven percentage points corresponds, roughly, to getting providers to observe the breathing pattern of children with a cough, take the pulse of a patient with a fever, or check the eyes

of an infant with diarrhea. These are not radical changes in behavior. They do not require that clinicians receive more training, but rather that supervisors be given the authority to act in cases where clinicians fail to do these simple things.

We do estimate rough costs for three of the simulated policies: filling vacant posts, hiring more qualified staff, and upgrading roads. Using information on medical staff salaries and educational costs collected by Dr. Masatu (and spreading educational costs over twenty years), we calculate that eliminating vacancies and upgrading staff would require US\$20,929 and US\$50,386 per year respectively. These expenditures total US\$0.10 and US\$0.24 per capita, given the 2002 Arusha population (10 to 20% of the annual per capita recurrent budget). These costs are likely to be underestimates because they do not account for the cost of equipment or other complementary inputs that might be required for clinicians to do their jobs. In addition, we do not consider whether current salaries are sufficient to induce trained clinicians to serve in the vacant or newly created rural posts we simulate. Chomitz et al. (1998) investigated options to induce doctors to serve in rural areas of Indonesia, finding training and cash incentives to be somewhat effective. While we do not account for these potential additional expenditures, our estimates give a rough approximation of the costs of implementing these policies.

We estimate that upgrading and maintaining Arusha's rural road network would cost at minimum US\$2,743,452 annually (or US\$13.27 per capita), far outstripping the cost of staff upgrades. We calculate this figure using road building and maintenance cost estimates given by Lebo and Schelling (2001) and divide road building investment costs over twenty years. While upgrading the road system carries a cost two orders of magnitude greater than training and hiring large numbers of medical staff and could not

be justified solely as a health care policy on a benefit-cost basis, we do not infer that building roads is a poor investment choice. Unfortunately, our analysis of road-building is rough at best because improved health care access is but one of the many benefits of improving rural transportation infrastructure.

As has been discussed extensively in the development literature, improving rural roads raises incomes and access to a variety of public services, making it one of the most effective investments developing country governments can make. For instance, road investments in rural India have had the largest impact on poverty reduction and productivity growth of any government expenditure (including health), with each 100 rupee investment decreasing the number of poor by 124 (Fan, Hazell, and Thorat 2000). Building roads, particularly of lower quality, has had large net benefits in China (Fan and Chan-Kang 2005). Road investments also yielded the greatest marginal returns in the less-favored rural areas of India and China (Fan and Hazell 2001). Platteau (1996) comprehensively reviewed transportation infrastructure constraints in sub-Saharan Africa, where sparse and low quality rural roads in most countries have hindered long-term prospects for economic growth, particularly in the agricultural sector. Improving existing transportation networks through maintenance of roads and vehicles, along with targeted road construction, can raise agricultural productivity considerably. Our simulation results underscore the consensus about the importance of improving rural transportation infrastructure by demonstrating the incidental benefits of road building for improved health care access. Additionally, they suggest that some of the development gains attributed to rural road construction could be caused by an interaction with health sector investment. Better roads improve the quality of care received by patients because

they improve access to existing high quality facilities, thereby increasing the return to past investments in health care quality.

Conclusions

Using unique data that matches health-seeking behavior with measures of clinician quality in all available health facilities, we show that patients choose from among a large set of available health facilities, even when these choices imply significant additional travel costs. In bypassing lower quality facilities, households manage to improve the quality of care that they receive. This behavior has important implications for policy in Tanzania and similar settings, and points out the importance of patient behavior in all developing country settings.

While existing health facilities ostensibly serve the rural residents of Arusha, access to competent care is poor because national staffing standards are too low, these low standards are not met, and many clinicians fail to report to work. Even when clinicians are present, many of them are unqualified to do their jobs and do not perform up to their skill level. In this setting, we investigate the discrepancy between patients' proximity to health facilities and their access to quality diagnostic care. We also examine the possible benefits from interventions designed to increase the number and the average qualifications and performance of personnel and to reduce the time to travel to better facilities. We simulate the impact of eliminating absenteeism, filling staff vacancies, upgrading staff qualifications, improving roads and improving clinician incentives at public sector facilities.

Our simulations indicate that aggressive policies of staff training and hiring—such as filling vacant posts and upgrading the average qualifications at rural facilities—

are required to achieve significant improvements in access to competent health care. While reducing absenteeism alone has a minor impact, coupling absenteeism reductions with stronger incentives for clinicians at public facilities to perform up to the level of their non-public sector counterparts does improve the quality of care. Improving rural roads offers an alternative approach to increasing patient access without new investments in clinician quality, leading to considerable reductions in patient travel times. This result highlights a benefit of improved transportation infrastructure that has received less attention in the literature on road-building and development. The results of our simulation analysis might also have broader relevance for the rural areas of other sub-Saharan African countries, many of which have conditions similar to those in Arusha: few and poorly trained clinicians, public sector facilities with few incentives for good performance, sparse rural road networks, and low population densities. Such regions might improve health care access by increasing clinician staffing, performance incentives, and road networks

The data presented in this article allow us to consider the combination of health-seeking behavior in a rural setting and the distribution and determinants of two important measures of health care quality in all facilities available to patients. Even without any interventions, patient behavior has important impacts on the quality of care. In addition, modeling patient behavior shows that the relative gains from various interventions are different than if we assume that patients visit the nearest facility. Broad improvements in the quality of care lead to smaller gains in overall access when we take into account behavior. On the other hand, improving roads or instituting small reforms in a selection of facilities have significantly greater impact when patients bypass than when they are

assumed to only visit the closest facility (because they reach better facilities less expensively or travel to reach the reformed facilities).

Any study of health care quality that ignores health-seeking behavior (by looking, for example, at averages across all facilities) makes the same mistake as policies that assume patients do not travel beyond their nearest facility. Such studies underestimate the gains to policies that reduce travel times or induce travel and over estimate the gains to policies that improve every facility. Governments with limited resources should focus on policies that take advantage of health seeking behavior, rather than ignoring it.

References

- Abel-Smith, B. and Rawal, P., "Can the Poor Afford 'Free' Health Services? A Case Study of Tanzania," *Health Policy and Planning*, 1992, 7 (4) 329-41
- Acton, J.P. "Nonmonetary factors in the Demand for Medical Services: Some Empirical Evidence," *Journal of Political Economy*, 1975, 83 (3)565-614.
- Akin, J. S. and Hutchinson, P. "Health-care facility choice and the phenomenon of bypassing," *Health Policy and Planning*, 1999, 14 (2) 135-51
- Banerjee, A., A. Deaton, and E. Duflo, "Health care delivery in rural Rajasthan," *Economic and Political Weekly*, 2004, 944-950.
- Banerjee, A., A. Deaton, and E. Duflo, "Wealth, health and health services in rural Rajasthan," *American Economic Review*, 2004, 94 (2), 326-330.
- Banerjee, Abhijit and Esther Duflo, "Addressing Absence," *Journal of Economic Perspectives*, 2006, 20(1), 117-132.
- Birnbaum, Allan, "Some Latent Trait Models and their use in Inferring an Examinee's Ability." In *Statistical Theories of Mental Test Score*, ed Frederic M. Lord and Melvin R. Novick, 397-479. 1968, London: Addison Wesley.
- Bock, Richard and Maury Leiberman, "Fitting a Response Model for Dichotomously Scored Items," *Psychometrika*, 1970, 33: 179-97
- Chaudhury, Nazmul, Jeffrey Hammer, Michael Kremer, et al., "Missing in Action: Teacher and Health Worker Absence in Developing Countries," *Journal of Economic Perspectives*, 2006, 20(1), 91-116.
- Chaudhury, Nazmul and Jeffrey S. Hammer, "Ghost doctors: absenteeism in Bangladeshi health facilities," *World Bank Economic Review*, 2004, 18 (3), 423-441.
- Chomitz, K., G. Setiadi, A. Azwar, N. Ismail, and Widiyarti, "What do doctors want? Developing incentives for doctors to serve in Indonesia's rural and remote areas," Technical Report 1888, World Bank 1998.
- CIA, "Tanzania." *CIA World Factbook 2007*. Accessed 24 September 2007
<<https://www.cia.gov/library/publications/the-world-factbook/geos/tz.html#Trans.>>
- Collier, P., S. Dercon, and J. MacKinnon, "Density versus quality in health care provision: using household data to make budgetary choices in Ethiopia," *World Bank Economic Review*, 2002, 16 (3), 425-448.
- Darkaoui N, De Brouwere V, Zayyoun A, Filali H, Belouali R, Gouaima F, Van der Vennet J, Dujardin B., "Utilization of hospital emergency service for primary care (study at the Children's Hospital of Rabat, Morocco)", *Rev Epidemiol Sante Publique*. 1999, 47 (Suppl 2):2S53-64.
- Das, Jishnu and Jeffrey Hammer, "Which Doctor? Combining Vignettes and Item-Response to Measure Doctor Quality," *Journal of Development Economics*, 2005, 78, 348-383.
- Das, Jishnu and Jeffrey Hammer, "Money for Nothing: The Dire Straights of Medical Practice in India," *Journal of Development Economics*, 2007, 83(1): 1-36.
- Das, Jishnu, Jeffrey Hammer and Kenneth Leonard, "The Quality of Medical Advice in Low Income Countries," 2008, World Bank Policy Research Working Paper no. 4501.

- Das, Jishnu and Thomas Pave Sohnesen, "Variations in Doctor Effort: Evidence from Paraguay," *Health Affairs*, 2007, 26(3): w324-w337.
- de Savigny, Don, Charles Mayombana, Eleuther Mwageni, et al., "Care-seeking patterns for fatal malaria in Tanzania," 2004, *Malaria Journal* 3:27.
- Dor, A., P. Gertler and J. Van der Gaag "Non-price Rationing and the Choice of Medical Care providers in Rural Cote D'Ivoire," *Journal of Health Economics*, 1987, 6(4), 291-305.
- Duflo, E., R. Hanna, and S. Ryan, "Monitoring Works: Getting Teachers to Come to School," 2007, NBER Working Paper #11880.
- Fan, Shenggen, Peter Hazell and Sukhadeo Thorat, "Government Spending, Growth and Poverty in Rural India," *American Journal of Agricultural Economics*, 2000, 82(4), 1038-1051.
- Fan, Shenggen and Connie Chan-Kang, "Road Development, Economic Growth and Poverty Reduction in China," International Food Policy Research Report 138, 2005.
- Fan, Shenggen and Peter Hazell, "Returns to Public Investments in the Less-favored Areas of India and China," *American Journal of Agricultural Economics*, 2001, 83(5): 1217-1222.
- Fabricant, S. J., C. W. Kamara, and A. Mills, "Why the Poor Pay More: Household Curative Expenditures in Rural Sierra Leone," *International Journal of Health Planning and Management*, July 1999, 14, 179-199.
- Filmer, D., J. Hammer, and L. Pritchett, "Weak links in the chain: a diagnosis of health policy in poor countries," *World Bank Research Observer*, 2000, 25 (2), 199-224.
- Gauthier, B. and Wane, W., "The Quest for Better Price and Quality of Health Care in Chad," Policy Research Working Paper 4462, 2008, The World Bank.
- Gertler, P. and J. Van der Gaag, *The Willingness to Pay for Medical Care: Evidence from Two Developing Countries*, 1990, Baltimore, MD: Johns Hopkins University Press, for the World Bank.
- Hanson, K., W.C. Yip, W Hsiao, "The impact of quality on the demand for outpatient services in Cyprus," *Health Economics*, 2004, 13 (12) 1167-1180.
- Hettige, Hamamala, "When Do Rural Roads Benefit the Poor and How?" Asian Development Bank, 2006.
- Kloos, H. "Utilization of Selected Hospitals, Health Centers and Health Stations in Central, Southern and Western Ethiopia," *Social Science and Medicine*, 1990, 31(2) 101-14
- Lebo, Jerry and Dieter Schelling, "Design and Appraisal of Rural Transport Infrastructure: Ensuring Basic Access for Rural Communities," 2001, World Bank Technical Paper No. 496.
- Leonard, K. L., "When States and Markets Fail: Asymmetric Information and the Role of NGOs in African Health Care," *International Review of Law and Economics*, 2002, 22 (1), 61-80.
- Leonard, K. L., "African Traditional Healers and Outcome-Contingent Contracts in Health Care," *Journal of Development Economics*, 2003, 71 (1), 1-22.
- Leonard, K. L., "Learning in Health Care: Evidence of learning about clinician quality in Tanzania," *Economic Development and Cultural Change*, 2007, 55 (3), 531-555.

- Leonard, K. L. and M. C. Masatu, "Comparing vignettes and direct clinician observation in a developing country context," *Social Science and Medicine*, 2005, 61 (9), 1944-1951.
- Leonard, K. L. and M. C. Masatu, "Outpatient process quality evaluation and the Hawthorne effect," *Social Science and Medicine*, 2006, 63 (9).
- Leonard, K. L. and M. C. Masatu, "Variation in the quality of care accessible to rural communities in Tanzania," *Health Affairs*, 2007, 26(3): w380-w392 2007.
- Leonard, K. L., M. C. Masatu and A. Vialou, "Getting Doctors to do their best: the roles of ability and motivation in health care," *Journal of Human Resources*, 2007, 43 (2).
- Leonard, K. L. G. Mliga, and D. Haile Mariam, "Bypassing Health Facilities in Tanzania: Revealed Preferences for Observable and Unobservable Quality," *Journal of African Economies*, 2002, 11 (4), 441-471.
- Lindelow, Magnus, "Understanding Spatial Variation in the Utilization of Health Services: Does Quality Matter?," *Centre for the Study of African Economies Working Paper Series*, Paper 212, 2004.
- London, L and Bachmann, O. M. "Paediatric utilisation of a teaching hospital and a community health centre. Predictors of level of care used by children from Khayelitsha, Cape Town." *South African Medical Journal*, 1997, 87(1):31-6.
- Platteau, Jean-Philippe, "Physical infrastructure as a constraint on agricultural growth: the Case of sub-Saharan Africa," *Oxford Development Studies*, 1996, 24(3):189 - 219
- Stock, R. "Distance and the Utilization of Health Facilities in Rural Nigeria," *Social Science and Medicine*, 1983, 17 (9) 563-70
- United Republic of Tanzania, "Health Statistics Abstract; Vol. II: Inventory Statistics," Technical Report, Planning and Policy Department, Ministry of Health, Dar Es Salaam 2002.
- van de Walle, Dominique and Dorothyjean Cratty, "Impact Evaluation of a Rural Road Rehabilitation Project," World Bank, 2002.
- Wagstaff, A., "Poverty and health sector inequalities," *Bulletin of the World Health Organization*, 2002, 80, 97-105.
- Wagstaff, A., Bustreo, F., J. Bryce, and M. Claeson, "Child health: Reaching the poor," *American Journal of Public Health*, May 2004, 94, 726-736.
- World Health Organization, "Declaration of Alma-Ata," International Conference on Primary Health Care, Alma-Ata, USSR, 6-12 September 1978.
- World Bank, "Kingdom of Morocco Impact Evaluation Report: Socioeconomic Influence of Rural Roads," OED Report 15808, 1996.

Table 1 Absenteeism in Facilities Accessible to Patients in Rural Arusha				
Broken down by		# Scheduled	Present	Absent
Facility Level	Dispensary	28	66%	33%
	Health center	19	52%	46%
	Hospital	46	84%	15%
Ownership	Public	63	71%	28%
	Private	15	80%	20%
	NGO	15	73%	26%
Location	Rural	19	63%	34%
	Semi-rural	38	74%	25%
	Urban	36	77%	22%
Cadre	MO	11	90%	10%
	AMO	15	87%	13%
	CO	45	76%	23%
	CA	22	45%	55%
Overall average		93	73%	26%

Table 2 Clinician Quality Summary Statistics

	Mean	Standard error
Clinician training (years)	4.89	2.68
Clinician competence (IRT score)	47.53%	17.76%
Clinician performance (IRT score)	40.96%	9.56%
Probability of correct diagnosis across clinicians (vignette score)		
Malaria with complications	10.38%	21.69%
Pelvic inflammatory disease	61.16%	28.85%
Infant diarrhea	58.31%	30.87%
Pneumonia	80.91%	20.20%
Malaria, worms, flu	82.38%	18.27%

Table 3 Determinants of the Correct Diagnosis in Vignettes

	Dependent variable: whether the diagnosis was correct (1), or not (0)						
	Random effects logit All illnesses	Random effects linear probability All illnesses	Logit				
	(1)	(2)	Malaria with complications (3)	Pelvic inflam. disease (4)	Infant diarrhea (5)	Pneumonia (6)	Malaria, worms or flu (7)
Competence (IRT score)	2.005 [0.768]***	0.304 [0.120]**	11.287 [3.834]***	0.433 [1.809]	5.815 [2.054]***	4.622 [2.206]**	-0.528 [1.386]
Years of training	-0.045 [0.062]	-0.007 [0.010]	0.226 [0.318]	-0.126 [0.146]	-0.1 [0.161]	0.154 [0.156]	-0.143 [0.104]
Rural	-0.437 [0.477]	-0.063 [0.073]	-0.237 [1.927]	-1.178 [1.135]	-0.753 [1.158]	-0.034 [1.214]	0.045 [0.783]
Urban	-0.693 [0.375]*	-0.102 [0.057]*	-1.721 [1.309]	-0.918 [0.771]	-0.472 [0.876]	0.914 [1.116]	-0.771 [0.636]
Dispensary	-0.07 [0.380]	-0.013 [0.061]	0.181 [1.771]	0.732 [1.015]	-0.723 [1.022]	0.302 [0.965]	-0.552 [0.594]
NGO	0.042 [0.585]	0.007 [0.094]	1.723 [3.315]	-1.111 [1.456]	0.59 [1.510]	1.526 [2.187]	-0.131 [0.857]
Private	0.351 [0.667]	0.054 [0.105]	3.38 [3.684]	-1.318 [1.675]	0.867 [1.754]	1.516 [2.384]	0.629 [1.000]
Public	0.072 [0.546]	0.014 [0.088]	3.042 [3.002]	-1.274 [1.391]	1.182 [1.476]	1.022 [2.114]	-0.451 [0.803]
Vignette type	Included						
Clinician random effect			2.577 [0.761]***	1.697 [0.358]***	1.771 [0.397]***	1.394 [0.394]***	0.972 [0.189]***
Constant	2.279 [0.930]**	0.86 [0.140]***	-13.104 [5.243]**	2.697 [2.147]	-1.742 [2.226]	-2.526 [2.734]	3.721 [1.488]**
Observations	714	714	106	106	106	106	290
# of unique clinicians	106	106	106	106	106	106	106

Standard errors in brackets; *, **, *** denote significance at 10%, 5% and 1%. The coefficients for the vignette dummy variables are jointly significantly different from zero in columns 1 and 2 (p-value < 0.000). Clinician random effects in column 1 are jointly significant (p-value = 0.027).

Table 4 Conditional Logit Determinants of Facility Choice

Dependent variables	Independent variable: Latent utility at each of 35 facilities	
Less absenteeism (% of clinicians present during site visit)	0.055	[0.057]
Clinician random effect	-0.108	[0.019]***
IRT competence	3.072	[1.166]***
IRT competence X responsiveness to effort	0.203	[0.129]
IRT performance	-0.401	[0.530]
IRT performance X responsiveness to effort	0.005	[0.121]
Training X net value of skill	0.003	[0.007]
Hospital X net value of skill	0.152	[0.024]***
IRT competence X net value of skill	0.153	[0.128]
Closest facility to the household	0.184	[0.021]***
Travel time (minutes) to facility	-0.026	[0.001]***
Travel time ² /1000	0.128	[0.010]***
Travel time ³ /100000	-0.024	[0.002]***
Public dispensary	0.073	[0.051]
Public health center	0.157	[0.049]***
Public hospital	0.551	[0.060]***
NGO 1—Dispensary	0.404	[0.060]***
NGO 2—Hospital	-0.336	[0.076]***
NGO 3—Dispensary	0.495	[0.069]***
NGO 3—Hospital	0.29	[0.059]***
Parastatal—Hospital	-1.1	[0.326]***
NGO 4—Dispensary	0.043	[0.076]
NGO 4—Hospital	0.323	[0.058]***
Private	-0.134	[0.082]
NGO 5—Dispensary	Omitted	
44 sub villages X net value of quality X IRT competence	Included	
1345 observations, 35 facilities		
Test of hypothesis that coefficients for performance are jointly equal to zero is rejected at p=0.06. Test of hypothesis that sub-village fixed effects are jointly equal to zero is rejected at p<0.0000.		
Standard errors in brackets; *, **, *** denote significance at 10%, 5% and 1%.		

Table 5 Impact of Policy Scenarios on Characteristics of Facilities

	Baseline (no policy change)			Policy simulation results									
	closest	chosen	predicted	<u>Eliminating absenteeism</u>		<u>Eliminating vacancies</u>		<u>Upgrading facilities</u>		<u>Upgrading roads</u>		<u>Public sector reform</u>	
				closest	predicted	closest	predicted	closest	predicted	closest	predicted	closest	predicted
Travel time	21.03 (0.00) (***)	28.55 (0.00) ***	27.12 (0.03)	21.03 (0.00) (***)	27.11 (0.03) (***)(+++)	21.03 (0.00) (***)	27.67 (0.09) +++	21.03 (0.00) (***)	28.15 (0.21) +++	9.65 (0.00) (***)	15.57 (0.02) (***)(+++)	21.03 (0.00) (***)	27.14 (0.03) +++
Years of training	4.00 (0.00) (***)	4.46 (0.00) (***)	4.55 (0.01)	4.03 (0.00) (***)	4.58 (0.01) (***)(+++)	4.64 (0.00) (***)	4.93 (0.02) (***)(+++)	6.54 (0.00) (***)	6.58 (0.01) (***)(+++)	3.96 (0.00) (***)	4.58 (0.01) (***)	4.00 (0.00) (***)	4.56 (0.01) (***)
Competence (% of items correct)	47.70 (0.06) (***)	48.60 (0.06) (***)	48.61 (0.03)	47.76 (0.06) (***)	48.72 (0.03) (***)	51.61 (0.34) (***)	51.99 (0.25) (***)(+++)	55.53 (0.96) (***)	55.38 (0.72) (***)(+++)	47.44 (0.06) (***)	49.55 (0.03) (***)	47.70 (0.06) (***)	48.64 (0.03) (***)
Performance (% of items correct)	39.64 (0.04) (***)	41.22 (0.04) (***)	41.37 (0.04)	39.67 (0.04) (***)	41.40 (0.04) (***)(+++)	40.72 (0.21) (***)	41.20 (0.18) (+++)	42.85 (0.58) (***)	42.55 (0.44) (***)(+++)	39.51 (0.04) (***)	41.86 (0.04) (***)	42.22 (0.04) (***)	44.38 (0.04) (***)

"Closest" refers to facility closest to the patient, "chosen" is the facility actually visited by sampled patients, and "predicted" gives the characteristics of facilities that patients are predicted to visit, based on the conditional Logit model of facility choice.

Bootstrapped standard errors (5,000 samples) in parentheses

*, **, *** denote significance of the difference from baseline (either closest or predicted) from each intervention (closest or predicted) at the 1, 5 and 10% level; parentheses indicate negative value in comparison, i.e. the policy leads to a fall in the examined characteristic.

+, ++, +++ denote significance level of the difference between the change from baseline to policy assuming patients visit the predicted facility and the change from baseline to policy assuming patients visit the closest facility at the 10%, 5% and 1% level; parentheses indicate negative value in comparison, i.e. the gain from policy is greater if patients visit the closest facility.

Table 6 Impact of Policy Scenarios on Probability of Correct Diagnosis (by Performance) for Seven Illnesses

	Baseline (no policy change)			Policy simulation results									
	closest	chosen	predicted	<u>Eliminating absenteeism</u>		<u>Eliminating vacancies</u>		<u>Upgrading facilities</u>		<u>Upgrading roads</u>		<u>Public sector reform</u>	
				closest	predicted	closest	predicted	closest	predicted	closest	predicted	closest	predicted
By Performance	8.57	9.11	9.08	8.57	9.08	9.48	9.15	36.23	26.10	8.56	9.04	10.11	10.60
Malaria with complications	(4.61)	(4.55)	4.55	(4.79)	(4.71)	(8.57)	(7.32)	(18.90)	(14.76)	(4.62)	(4.71)	(4.93)	(4.84)
Pelvic inflam. disease	67.45	66.02	66.38	67.45	66.52	78.49	74.53	85.23	76.78	67.59	67.72	67.61	66.69
	(7.04)	(6.08)	6.17	(7.07)	(6.20)	(6.25)	(5.60)	(8.16)	(7.21)	(7.05)	(5.96)	(6.67)	(5.72)
Infant diarrhea	50.17	54.84	55.16	50.17	55.29	64.36	65.08	77.96	71.09	50.21	56.71	52.26	57.75
	(5.86)	(5.14)	5.23	(5.88)	(5.26)	(7.90)	(6.45)	(10.98)	(8.53)	(5.92)	(5.30)	(5.64)	(4.92)
	(*)				+	***	***(+)	**	**(+)			***	***++
Child pneumonia	76.94	77.43	78.45	76.94	78.54	88.20	85.11	95.21	91.72	76.79	79.51	78.79	80.60
	(6.71)	(5.64)	5.60	(6.74)	(5.63)	(4.48)	(4.25)	(4.21)	(4.20)	(6.76)	(5.36)	(6.18)	(4.88)
		(**)				***	***(++)	***	***(+)			**	***+
Malaria, worms, flu	89.67	88.72	88.62	89.67	88.71	93.02	91.44	93.44	90.29	89.74	89.33	89.56	88.51
	(2.64)	(2.32)	2.34	(2.64)	(2.35)	(1.82)	(1.85)	(2.69)	(2.47)	(2.63)	(2.13)	(2.56)	(2.22)
					+	***	***	*	*				
By Competence	14.65	17.01	16.75	14.65	16.78	20.35	18.80	64.52	50.94	14.58	18.43	14.65	16.85
Malaria with complications	(5.36)	(5.56)	5.52	(5.56)	(5.68)	(10.55)	(8.81)	(19.19)	(14.66)	(5.37)	(6.11)	(5.36)	(5.55)
					*	**	**	***	***				
Pelvic inflam. disease	68.37	66.84	67.19	68.37	67.34	79.24	75.23	85.80	77.50	68.50	68.56	68.37	67.32
	(5.85)	(4.91)	5.02	(5.87)	(5.04)	(5.65)	(4.80)	(7.25)	(6.14)	(5.88)	(4.80)	(5.85)	(5.03)
					+	***	***(+)	**	**(+)				*+
Infant diarrhea	61.62	65.58	65.58	61.62	65.79	77.32	76.96	88.20	83.44	61.50	67.85	61.62	65.60
	(5.38)	(4.54)	4.56	(5.39)	(4.58)	(6.02)	(4.76)	(7.41)	(5.62)	(5.45)	(4.46)	(5.38)	(4.55)
	(*)				*++	***	***(++)	***	***(+)		*		
Child pneumonia	84.37	84.10	84.84	84.37	84.99	91.85	89.54	97.02	95.03	84.15	85.89	84.37	84.92
	(4.20)	(3.42)	3.42	(4.23)	(3.43)	(2.99)	(2.85)	(2.47)	(2.44)	(4.29)	(3.13)	(4.20)	(3.42)
						(+)	(+)		(+)				+
Malaria, worms, flu	89.11	88.12	88.04	89.11	88.13	92.55	90.94	92.96	89.62	89.17	88.74	89.11	88.08
	(2.48)	(2.12)	2.15	(2.48)	(2.16)	(1.83)	(1.76)	(2.81)	(2.49)	(2.47)	(1.94)	(2.48)	(2.15)
						**	**	*					

See Table 5 for description

Figure 1: Access to health facilities competent and not competent in diagnosing the causes of infant diarrhea in Arusha, Monduli and Arumeru

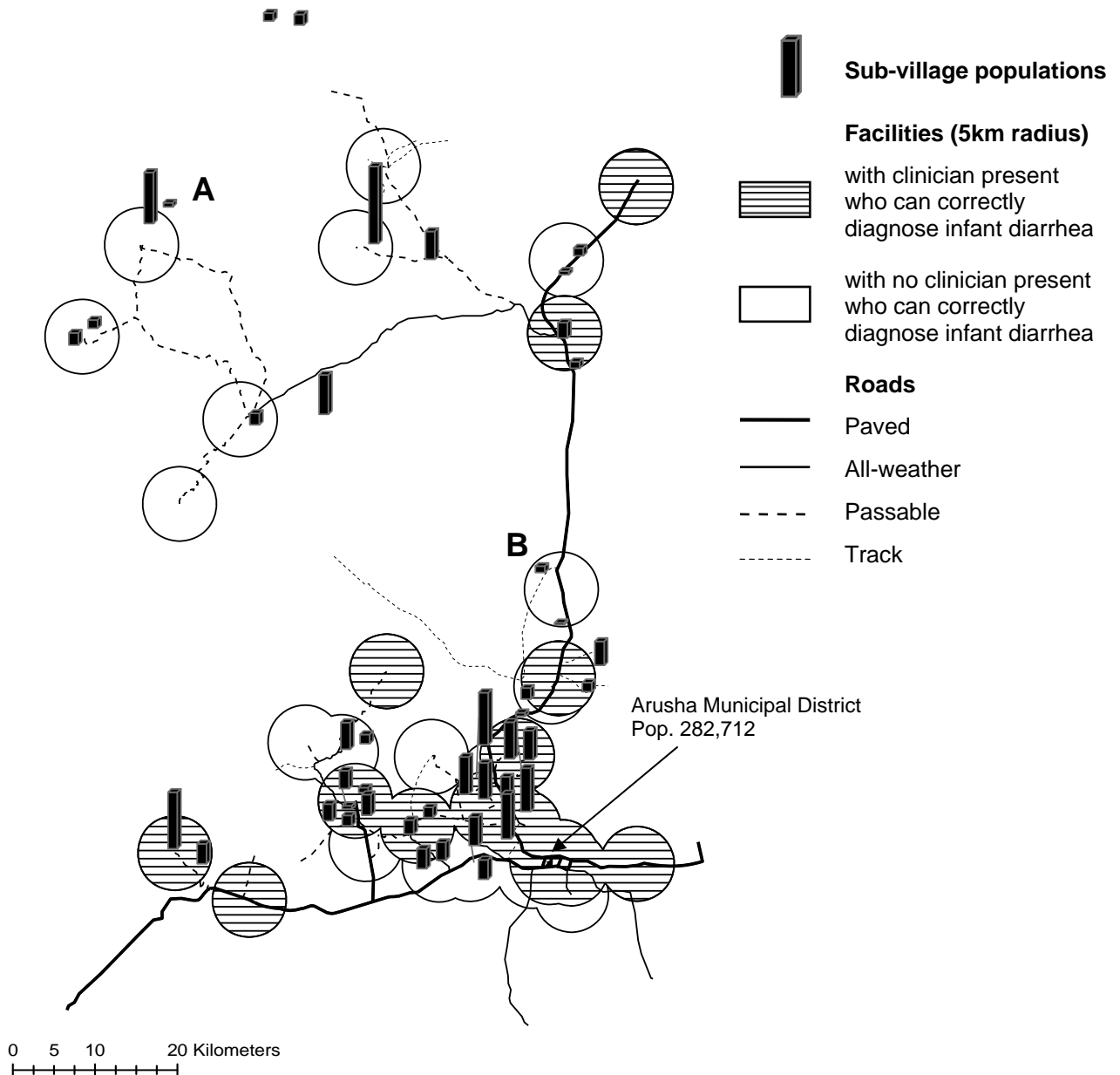


Figure 1 presents a map of Arusha, Monduli and Arumeru districts showing roads, sampled sub-villages and health facilities. Sampled sub-villages are represented by black bars proportional to their population. Each facility is represented by a 5km-radius circle, indicating the accessibility by foot travel to the surrounding population. Shaded circles signify health facilities with at least one clinician competent in diagnosing the causes of

infant diarrhea (defined as an 80% or greater probability of correct diagnosis based on vignette scores), and empty circles are health facilities without any clinician competent in infant diarrhea diagnosis.

Figure 2 Competence and performance scores for all clinicians in the sample observed on both scores (95 observations)

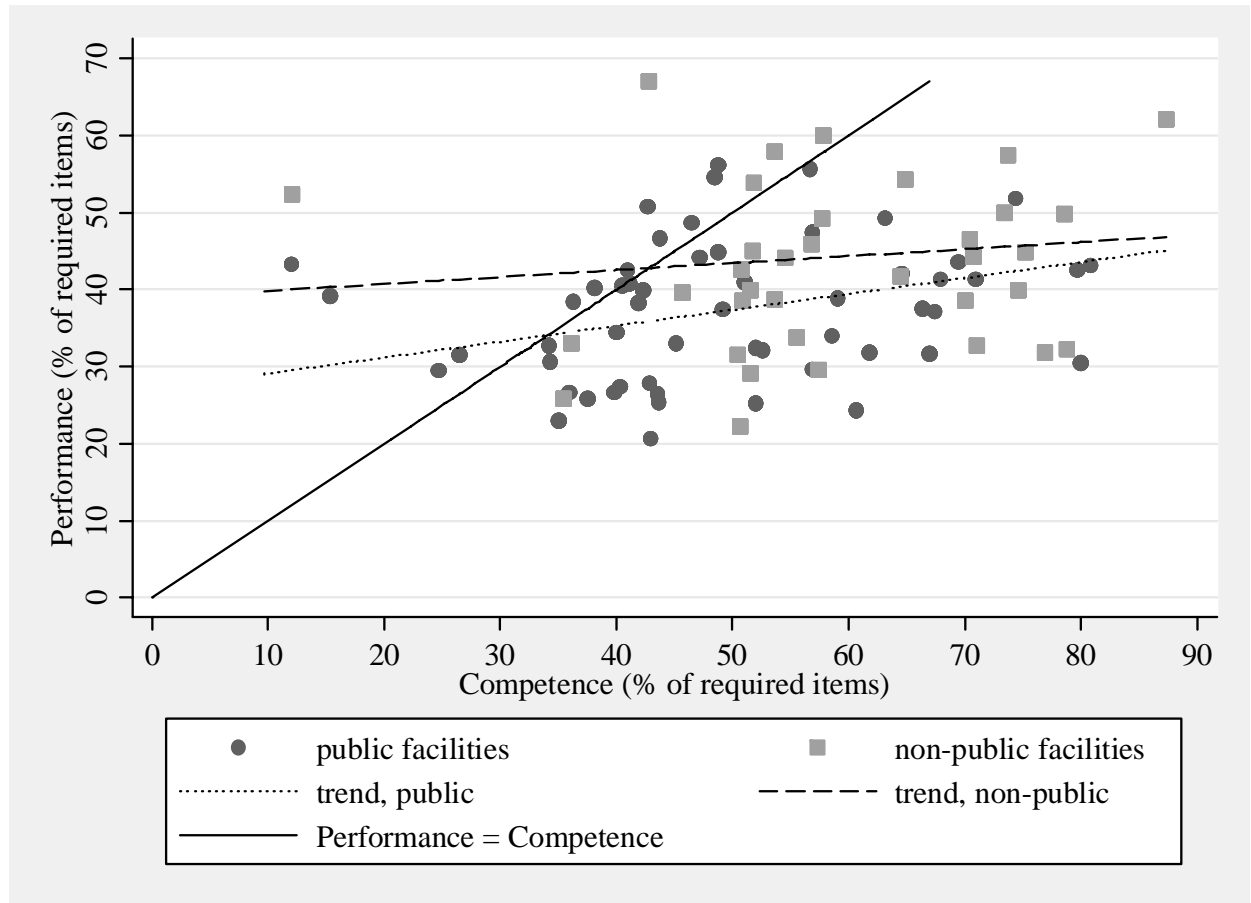


Figure 2 shows the levels of competence and performance (derived from IRT analysis as discussed in the text) for 95 clinicians who were observed with both vignette case studies and actual patients. The solid line represents the frontier where performance is equal to competence. The dashed line represents the trend line for clinicians at non-public facilities, and the dotted line represents the trend line for clinicians at public facilities.