WORKING PAPERS IN ECONOMICS

No. 02/09

OTTAR MÆSTAD, GAUTE TORSVIK AND ARILD AAKVIK

OVERWORKED? THE RELATIONSHIP BETWEEN WORKLOAD AND HEALTH WORKER PERFORMANCE IN RURAL TANZANIA

Department of Economics
UNIVERSITY OF BERGEN

Overworked? The relationship between workload and health worker performance in rural Tanzania

by

Ottar Mæstad[1]
Chr Michelsen Institute, Bergen

Gaute Torsvik
Department of Economics, University of Bergen and Chr Michelsen Institute, Bergen

Arild Aakvik
Department of Economics, University of Bergen

30 March, 2009

Abstract

The current shortage of health workers in many low-income countries poses a threat to the quality of health services. When the number of patients per health worker grows sufficiently high, there will be insufficient time to diagnose and treat all patients adequately. This paper tests the hypothesis that a high caseload reduces the level of effort per patient in the diagnostic process, using a new data set from rural Tanzania. Tanzania has a severe shortage of health workers, and previous research has pointed at high workload as a main reason for sub-standard clinical performance. We observed and evaluated the level of effort of 159 clinicians in 2,095 outpatient consultations at 126 health facilities with different levels of caseload per clinician. Surprisingly, we find no association between caseload and the level of effort per patient in the diagnostic process. In fact, clinicians appear to have ample amounts of idle time. We conclude that health workers are not overworked and that scaling up the number of health workers in this setting is unlikely to raise the quality of health services. A more promising measure for improved quality is to raise the level of formal clinical training among the clinicians, although training alone seems far from enough to raise quality to adequate levels.

---

[1] Corresponding author. E-mail ottar.mestad@cmi.no. We are grateful to Aziza Mwisongo, Ida Lindkvist, Magnus Hatlebakk and the rest of the MAP project team for their cooperation and support on this research, to Kenneth L. Leonard for valuable discussions about the survey instruments, and to Alexander K. Rowe for thoughtful comments on an earlier draft. We also thank seminar participants at the World Bank Human Development Forum, the Norwegian Annual Economics Conference as well as colleagues at CMI and the University of Bergen for useful comments and suggestions. Financial support from the Research Council of Norway is gratefully acknowledged.

1. Introduction

According to WHO (2006), 57 countries across the globe have a critical shortage of health workers. The claim is that the health workforce in these countries is too small to enable good coverage of even the most essential health interventions, including those necessary to reach the health-related Millennium Development Goals (MDGs). Besides reducing the range of services offered, a shortage of health workers may also diminish service quality. With few health workers, caseload per worker will grow high, and less time will be available per patient. A decline in the quality of the service is then likely, as the provision of high quality care requires health workers to spend sufficient time and effort with each patient.

The view that a shortage of health workers reduces the quality of health services accords well with recent research that has identified a *know-do gap* in clinical practice in low-income settings; what health workers do differs systematically from what they know they should do (Leonard et al, 2007; Das and Hammer, 2007). One explanation why health workers perform below their potential may be that they face an excessive workload. This account is also in line with how many health workers describe their current work situation. In focus group discussions with Tanzanian health workers, it was often acknowledged that inadequate quality of care is a problem in patient consultations (Lindkvist et al, 2009). For instance:

> *…once the patient arrives, the doctor will briefly listen to what the patient will have to say, and then … do a quick clinical investigation, and sometimes they don't even do investigations properly* [Clinical officer]

Furthermore, many health workers argued that high workloads are a major reason for the low quality of health services:[2]

> *…the workload becomes so big and as result the doctors decide to rush in order to catch up with the big number of patients waiting* [Doctor]

This paper tests the hypothesis that a high caseload reduces assessment quality, defined as effort per patient in the diagnostic process. We use a new data set from rural Tanzania, a country defined by WHO (2006) to have a critical shortage of health workers. The WHO

[2] See also Mæstad and Mwisongo (2007).

threshold for a critical shortage is 2.5 health workers (counting doctors, nurses and midwives only) per 1,000 inhabitants, while the figure in Tanzania is only 0.4 - 0.6 depending on definitions (see below). Even though the number of health workers per capita is low in most places in rural Tanzania, there is considerable variation in caseload per clinician across health facilities. In a situation with a general shortage of health workers, there will be – under reasonable assumptions – a negative relationship between caseload and effort per patient. We search for this pattern in the data.

Two methodological challenges are obvious: First, it may be difficult to identify a causal impact of caseload on assessment quality because of a potential simultaneity bias, as the quality of health care may have an impact on the demand for health services and thus on caseloads. Previous studies have found evidence that patients in Tanzania sometimes bypass their closest health facility and approach some other provider, suggesting that quality matters for the choice of provider (Leonard et al, 2002).

To deal with this challenge, we need a source of exogenous variation in caseload. We use the catchment population of the health facility (per clinician) as an instrument for caseload (per clinician). We expect the catchment population of a health facility to be highly correlated with the number of patients. We will also argue that there is little reason to believe that there is a direct association between catchment population per clinician and the quality of services. Hence, we anticipate our instrumental variable to perform satisfactorily.

A second challenge is that the relationship between caseload and quality may be highly nonlinear. Some health facilities may have such low caseloads per clinician that there will be no association between caseload and the quality of health services at the margin. By pooling such facilities together with facilities with a heavy workload, a linear model may bias our estimates of how caseload affects the level of effort per patient (positive bias for high caseloads and negative bias for low caseloads). We deal with this issue by estimating a nonlinear (kinked) relationship between caseload and effort, imposing alternative exogenous thresholds of caseload at which the time constraint starts to affect clinical practice.

The paper relates to two strands of the literature on quality of health care in low income countries. First, it builds on the public health literature on determinants of health worker performance (e.g., Rowe et al, 2000; Zurovac et al, 2004; Osterholt el al, 2006; Naimoli et al, 2006). Although the influence of caseload is not a major issue in this literature, it is discussed in several contributions. This paper adds to this literature by analysing the relationship between caseload and performance within a theoretical framework which takes into account both that the relationship may be nonlinear and that causality may run both ways. Moreover, as a secondary output of the analysis, we are able to identify a set of predictors – other than caseload – of health worker performance.

Second, the paper relates to a recent literature within economics on new ways of measuring and analysing quality of health care in low income countries (Das et al, 2008). A common way of assessing the quality of health services in such settings has been to register the availability of physical inputs (equipment, drugs, health workers, etc.) (see Amin et al, 2008). Such measures have obvious shortcomings, particularly because they do not capture the knowledge of health workers' and the efforts they put into their practice. These issues have more recently been dealt with by measuring the quality of care either through direct observation or through testing the knowledge of health care providers through vignettes (i.e., hypothetical patient-provider encounters). Quality scores have then been computed by comparing what health workers do with a checklist of essential procedures (e.g., Das and Hammer, 2005, 2007; Leonard et al, 2007; Barber et al, 2007).[3]

We use direct observation to measure assessment quality (i.e., effort in the diagnostic process) in outpatient consultations. The diagnostic process is time consuming and thus likely to be vulnerable to shortages of time. Effort in the diagnostic process is measured by the *number of relevant questions asked and examinations performed,* where the set of relevant questions and examinations follow from the symptoms of the patient as well as local clinical guidelines. We use data from 2,095 outpatient consultations, conducted by 159 clinicians at 126 health facilities with different levels of caseload per clinician.

We find that health workers perform only 22% of the diagnostic items prescribed by protocol. Clinicians ask 2.9 relevant questions and perform 1.3 relevant physical examinations per patient. We find no association between caseload and efforts in the diagnostic process, neither before nor after we control for simultaneity bias in a regression model. In fact, simultaneity does not emerge as a problem as there appear to be no effect on caseload of effort in the

[3] Other methods for measuring quality of health care are record reviews (see Ofori-Adjei and Arhinful, 1996) and simulated clients (see Madden et al, 1997).

diagnostic process. Estimation of the nonlinear (kinked) relationship between caseload and effort in the diagnostic process does not show signs of any associations at the margin either. On average, there seems to be considerable slack capacity. This finding has strong policy implications: Despite the low number of health workers in rural Tanzania, compared to international standards, a scaling up is not likely to improve the quality of the service. We do find, however, that quality enhancing effort is higher among more trained health workers. Hence, a change in the skill mix is a more appropriate policy measure than increasing the number of health workers.

The paper proceeds as follows: Section 2 provides a brief outline of our study area. A theoretical model of the relationship between workload and health workers' choice of effort follows in Section 3. Section 4 describes the data set and how data were collected. Section 5 presents descriptive statistics and the results of the regression analyses. We discuss our main findings in Section 6. Section 7 contains robustness analyses, and Section 8 concludes.

2. The context

Tanzania is a low-income country with a GNI per capita of 370 USD. Life expectancy at birth is 51.9 years, and infant mortality is 73.6 per 1,000 live births (WDI, 2008). Child mortality is on a remarkable downward trend (Masanja et al, 2008). Major causes of premature deaths among children include respiratory infections, malaria, and diarrhoea, conditions that normally can be cured by simple, low-cost treatments (Black et al, 2003).

The health care system consists of an extensive network of health facilities, including 219 hospitals, 481 health centres and 4,679 dispensaries. 70% of the population lives within a 5 km walking distance from a health facility. 64% of the health facilities are owned by the government; the remainder is run by voluntary agencies, private-for profit and para-statal providers (TSAM, 2007). Voluntary agencies, which run 40% of the hospitals, are typically located in rural areas, whereas private-for-profit providers are more common in the cities. As much as 80% of the population lives in rural areas (Census, 2002).

The total number of health workers in the country is 1.4 per 1,000 inhabitants. The number of doctors (physicians), nurses and midwives per 1,000 is 0.4, rising to 0.6 if we include assistant medical officers and clinical officers among the doctors. In rural areas, clinical officers with three years of clinical training provide most clinical services. However, it is also common in these areas for cadres with little or no formal clinical training, such as nurses and assistants, to carry out clinical work.

Our study area includes all nine rural districts in the Morogoro and Dodoma regions, located in central Tanzania. The total population in the area is 2.9 million, i.e., 9% of the country's total population (Census, 2002). There are 440 health facilities in the area owned by the government (81%) and Christian voluntary agencies (19%). In addition, there are a few para-statal and Muslim health facilities. The average health worker density in the area is 1.0 health workers per 1,000 inhabitants, lower than the national average of 1.4, and also lower than the average of 1.1 health workers per 1,000 inhabitants across all rural districts of the country. The number of health workers per capita varies across districts in the study area, from 0.6 per 1,000 in Kongwa to 1.9 per 1,000 in Kilombero (HRH Census 2001/2002).

At all three levels of care – dispensaries, health centres and hospitals – provide outpatient services, and the nature of the services does not differ much among them, except that higher-level facilities are more likely to have a laboratory. Health facilities provide drugs, but there is also a vibrant private pharmaceutical market. There is no appointment system in the outpatient departments; people queue as they arrive. Consultation is available for all who show up on the day; patients are usually not asked to return later.

3. A theoretical model

This section formalizes the relationship between caseload, the level of effort per patient and the quality of health services. In this paper, effort denotes actions taken by the clinician to improve the quality of the diagnostic process, such as history taking and physical examinations of patients. More generally, we may think of effort as all actions that improve the quality of health services, including activities that increase patients' feeling of convenience, comfort and knowledge about their medical conditions (Wedig et al, 1989). All such undertakings are time consuming. Hence, we assume that time use per patient increases with the level of effort.

Exerting effort generates both benefits and costs for the health workers. The gains come as intrinsic and/or extrinsic rewards associated with the delivery of high quality health services, the costs come from the fact that it is psychologically and physically demanding to provide high quality health care on a regular basis. Health workers with high levels of knowledge and skills may be able to exert quality-enhancing effort with greater ease – or smaller costs – than unskilled health workers. We capture these aspects in the following parameterisation of a health worker's utility of exerting effort

(1) $$u(e) = \alpha e - \frac{1}{2k}e^2,$$

where e denotes the effort per patient, k is the level of knowledge and skills and α captures the health worker's level of intrinsic and extrinsic motivation (or incentives) to exert effort. The latter parameter captures the impact of factors such as professional and altruistic attitudes, financial and non-financial incentives and the expectations of patients, colleagues and managers, etc.

We assume that health workers seek to maximize their utility subject to the constraint that all patients who show up on a given day must be consulted. Let w denote the caseload (i.e., the number of patients) faced by an individual health worker, let l be the total time that each health worker spends at the clinic, and let time use per patient (t) be given by the function $t = e$. Formally, utility is maximized subject to the constraint $ew \leq l$.

Caseload is likely to be an endogenous variable; the level of effort exerted by the health workers may affect patients' demand for health care. First, demand is likely to depend positively on patients' perceived quality of the services. Actions that improve the quality of the service, such as a higher level of effort, may therefore increase the number of patients. (Note, however, that actions that improve quality from a medical perspective will not necessarily translate into higher perceived quality from the patients' perspective.) Second, higher effort may increase the probability that patients are cured and may thus reduce reattendances and thereby the total number of consultations. Caseload is therefore a function of effort; $w = w(e)$. In our basic model, labour supply is exogenous ($l = \bar{l}$). We can then formulate the health workers' decision problem as

(2) $$\max_e u(e) = \alpha e - \frac{1}{2k}e^2 \quad \text{s.t.} \quad ew(e) \leq \bar{l}\ .$$

If the constraint does not bind, health workers can choose their first-best level of effort $e^* = \alpha k$. In this case, caseload will not affect effort, as the total time use on patients is lower than the amount of available time.

If the constraint binds, the health worker's choice of effort is implicitly given by the constraint; $\hat{e} = \bar{l}/w(\hat{e})$. In this case, it is easy to see that an increase in caseload must reduce the level of effort. That is, when more patients arrive at a clinic where the health worker's level of effort already is constrained from the demand side, the health worker has no choice but to reduce her effort further in order to take care of the additional patients. Formally, the effect of an exogenous shift in caseload on effort will be $d\hat{e}/dw = -\hat{e}/w(1+\varepsilon_{we})$, where ε_{we} is the elasticity of demand with respect to effort.

We show in Appendix 1 that the negative relationship between caseload and effort also holds when health workers optimally choose the total time *l* spent at the clinic. In this case, an exogenous increase in caseload induces health workers to spend more time at the clinic (an increase in *l*) but not to the extent that it will obviate the need to reduce the level of effort per patient.

Figure 1 illustrates. When caseload is lower than the threshold $\hat{w}$, the health worker can choose his or her preferred level of effort (e^*) and still have time to treat all of the patients that come to the clinic. In this "slack" region, variations in caseload will not affect effort. When the caseload exceeds $\hat{w}$, the health worker will reduce effort per patient in order to treat all patients who come to the clinic. Hence, if health workers are overworked, i.e., if a heavy workload is making health workers reduce the quality of the services, we ought to observe a negative relationship between exogenous shifts in demand (caseload) and the level of effort per patient.

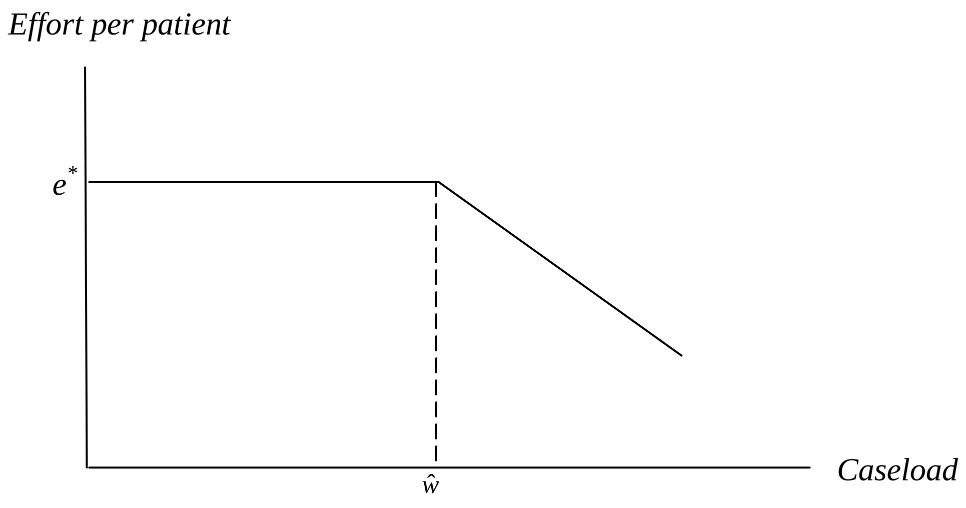

Figure 1: *The relationship between effort per patient and caseload.*

Note that heterogeneity among health workers and across health facilities (represented in our model by differences in α, k, and ε_{we}) implies that the positioning of the caseload / effort curve differs across health workers, although the basic shape will be the same.

4. Data

Our data was collected through the MAP (*Health Worker Motivation, Availability and Performance*) project in Tanzania in 2007. The MAP data set consists of a random sample of 159 health workers at 126 government and voluntary (Christian) health facilities in 9 districts. In the first stage, 14 health facilities were selected from each district. An updated list of facilities was provided by the Regional and District Medical Officers. Within districts, we randomly selected health facilities within six strata defined by the type of facility (hospital, health centre and dispensary) and ownership (government and voluntary agencies). Table 1 describes the sample of facilities by facility type and ownership.

Table 1: Number of health facilities in sample and in population.

Facility type	Number of health facilities			Population
	Government	Voluntary agencies	Total	*Total*
Hospitals	6	5	11	*12*
Health centres	24	1	25	*35*
Dispensaries	56	34	90	*393*
Total	86	40	126	*440*

At each facility, a maximum of two clinicians were randomly selected for observation among those who were working in the outpatient department (OPD) on the day of the visit. Visits were unannounced. If there was only one clinician at the health facility, he or she was observed over two days. All clinicians were observed from morning to around 1 pm (or earlier if more than 20 observations had already been made on that day). Graduate students from medical schools in Dar es Salaam were used as surveyors after a one week training session. 3,494 consultations were observed in total. We measured assessment quality for the 2,095 patients that presented with fever, cough, and/or diarrhoea. Reattendances were not included. Voluntary and informed consent from all patients and health workers was secured. No health workers and less than a handful of patients refused to participate. Table 2 summarizes the sample of consultations by primary symptoms and age of patient.

Table 2: Sample of consultations by symptom and age of patient.

Primary symptoms	Sample of consultations		
	Age < 5 years	Age > 5 years	Total
Fever, cough and/or diarrhoea	1371	724	2095
Other	359	1040	1399
Total	1730	1764	3494

During each consultation, surveyors noted which tasks – among a set of pre-defined relevant tasks – that were actually performed by the health worker. The set of pre-defined tasks included issues related to courtesy and communication and, for each of the focus symptoms (fever, cough, and diarrhoea), a list of relevant history taking questions and physical examinations. The list of relevant questions and examinations was adopted from Leonard et al (2007), who based their approach on the training curriculum of clinical officers in Tanzania. We expanded their framework by adding relevant items from the guidelines for Integrated Management of Childhood Illnesses (IMCI), which applies to children under the age of 5 years. Hence, the list of relevant items is longer for children under the age of 5 years than for others (see Appendix 2).

We conducted exit interviews with all adult patients and with the caretakers of the children. Background data on the observed health workers were obtained in interviews. Health facility data were obtained from interviews with the facility in-charge and from records. In particular, data on the number of patients are from facility records.

Since the actual number of consultations in the study area is unknown, sample weights were estimated. At each facility, we weighted the observations by the total number of consultations over the two days of observation, divided by the number of consultations observed. [4]

5. Analysis and results

Our aim is to test how variations in caseload between health facilities affect the quality of the clinical work conducted. We start by discussing in some detail how we measure the quality of work, the caseload and the various controls that appear relevant.

Key variables. Definitions and descriptive statistics

Following the approach outlined in Das et al (2008), our dependent variable is the quality of health services as measured by the level of effort exerted in the diagnostic process. Effort in the diagnostic process is measured as the *number of relevant history taking questions asked and physical examinations performed*. We focus on the diagnostic process, because this process is time consuming and thus likely to be vulnerable to shortages of time. Other aspects of quality, such as whether correct treatment is provided, are also likely to be affected by time constraints, both indirectly through the relationship between a thorough diagnostic process and the probability of providing correct treatment, and directly through the amount of time available for making careful judgements. Our data set does not contain such data, however.[5]

[4] For logistical reasons, we were able to correctly record the total number consultations only at the first day of observation. We use the number of consultations on the first day times two as our estimate of the total number of consultations over the two days. Moreover, since the sample of consultations for a given clinician is not a true random sample (observation normally ended when the number of observed patients per day exceeded 20), the use of consultation weights is based on the assumption that patients arriving later in the day are not treated systematically different from the observed ones. Our results suggest that this may be a strong assumption, but we nevertheless prefer to use these estimated weights over a non-weighted approach.

[5] Some tasks related to the explanation of diagnosis and health education, as well as courtesy, are time consuming. Sensitivity tests have been conducted where these tasks have been included in our measure of assessment quality (Section 7). Moreover, time use per patient is also a potential indicator of the level of effort. We have tried this approach in the sensitivity analyses, although our impression from the fieldwork is that this variable is not a good estimate of the level of effort as some clinicians spend a considerable amount of time talking to patients about issues unrelated to their medical condition. Finally, the effort variable does not necessarily account for all information spontaneously offered by the patient. If a person said "I had fever for two days, with chills, sore throat, diarrhoea, and a runny nose" the surveyors could in principle mark these items as non-applicable. We do not how accurately such information was recorded, though.

Caseload is calculated as the total number of outpatient consultations at the facility at the first day of observation, divided by the number of full time equivalent health workers in the OPD.[6]

Table 3: Summary statistics effort and caseload. Sampling weights are used to construct weighted averages.

Variable	Variable definition	n	Mean (weighted)	Mean (unweighted)	Std dev	Min	Max
Questions	Number of history taking questions (a)	2,095	2.94	2.92	1.88	0	12
Examinations	Number of physical examinations (b)	2,095	1.26	1.13	1.35	0	15
Effort	(a) + (b)	2,095	4.20	4.04	2.76	0	22
Time	Minutes per patient	1,789	5.66	5.80	3.74	0	45
Caseload	Number of OPD patients per full-time OPD health worker per day	2,095	18.48	16.36	9.76	1	45

Table 3 presents summary statistics on effort and caseload. On average, clinicians ask 2.94 relevant questions and undertake 1.26 physical examinations per patient. This is about one question and .25 examinations less than found in a comparable study from Arusha region in Tanzania (Das et al, 2008). The average level of effort – measured as the sum of the number of relevant questions and examinations – is 4.2, corresponding to 22% of all relevant tasks according to protocol.[7]

The average patient sees a clinician who counsels 18.5 patients in the OPD per day. There is considerable variation both in the effort and the caseload variable. Total time use per patient, including consultation time and follow up after laboratory testing, is 5.7 minutes. This includes the time taken to fill prescriptions and patient cards, if applicable.

Although we are primarily interested in examining the relationship between caseload and health worker effort, we also identify other predictors of effort. The analysis includes background variables at the health worker, health facility and patient levels (see Table 4). At the health worker level, we include variables for the level of training (*clinical officer*), sex (*male*) and age (*age*). The training variable is a dummy variable that distinguishes between health workers with clinical training at least at the level of a clinical officer and health

[6] Missing data on the number of patients on the day of observation at three facilities were replaced by the average number of patients per working day in August 2007.

[7] Mwisongo and Mæstad (2009) provide an in-depth discussion of which questions that were asked and which examinations that were performed.

workers from lower cadres, mostly nurses and assistants. Health workers trained as a clinical officer or above, i.e., workers with at least three years of clinical training, take care of 69% of the patients (Table 4). Within this group, those with more training than a clinical officer (i.e., medical officers (physicians) and assistant clinical officers) see only 2.5% of the patients. A large group of patients (31%) are consulted mostly by nurses and assistants with little or no formal clinical training. These cadres are not supposed to act as clinicians but do so due to lack of qualified workers. Finally, we included training in the Integrated Management of Childhood Illness (IMCI) as a control (*imci_child*). This is a dummy variable that takes a positive value when a patient in the target group of IMCI (i.e., children under the age of five) is treated by a health worker trained in IMCI.

At the facility level, we control for ownership with a dummy for government owned facilities (*government*). Government-owned facilities have a different governance structure from voluntary agencies, and this may result in different incentives to exert effort (Leonard et al, 2007). The variable may also control for selection effects insofar as health workers with different preferences (e.g., different levels of intrinsic motivation) are systematically (self-) selected into government facilities vs. voluntary agencies. We also control for the availability of drugs (*drugs*), as the lack of particular drugs may reduce the incentives for health workers to undertake careful diagnosis. We recorded the availability of seven essential drugs during our visit and have scored the variable from 0 through 7. Finally, we include a dummy variable for the existence of a laboratory (*laboratory*), because laboratory tests may to some degree substitute for a more comprehensive oral and physical examination.

At the patient level, we control for the patient being a child below the age of five (*child*), in which case the IMCI guidelines are applicable. Furthermore, the surveyors made a subjective assessment of the patients' general condition (*patient weakness*). The variable is scored as follows: 0 = not weak, 1 = moderately weak, 2 = very weak. Finally, we controlled for each patient's number in the order of observed consultations for each respective health worker (*patient number*). This is because we expect the presence of an external observer to raise the performance of the clinician (the Hawthorne effect). Leonard and Masatu (2006) have demonstrated, however, that the Hawthorne effect rapidly wears off in a situation almost identical to our study setting. They showed that after 10-15 consultations, clinicians are likely to return to their normal level of performance. In our sample, the average patient is the 14th

patient in the queue. In order to control for the possibility of a diminishing Hawthorne effect during the observation period, we included the patient number as a control variable.

Table 4. Descriptive statistics. Control variables.

Variable	Variable definition	#Obs	Mean	Std dev	Min	Max
Clinical officer	Health worker has at least three years of clinical training	2,095	0.69	0.46	0	1
Male	Male health worker	2,081	0.47	0.50	0	1
Age	Health worker's age (in years)	2,095	40.1	9.99	22	70
Imci_child	Being trained in IMCI & patient is <5 years	2,068	0.45	0.50	0	1
Government	Government owned facility	2,095	0.79	0.41	0	1
Drugs	Availability of seven drugs at the day of visit (0-7)	2,095	4.47	1.73	1	7
Laboratory	Facility has a laboratory	2,095	0.56	0.50	0	1
Child	Patient <5 years	2,095	0.65	0.78	0	1
Patient weakness	Weakness of patient, observer's assessment (0 = not weak, 1 = moderately weak, 2 = very weak)	2,051	0.31	0.50	0	2
Patient number	Patient's number in the order of observed consultations for health worker h (including both day 1 and 2)	2095	14.36	9.84	1	50

Finally, in order to account for the influence of case complexity we include symptom fixed effects throughout for all seven possible combinations of the three focus symptoms.

Relationship between caseload and effort

In order to identify the relationship between caseload and effort, we estimate the following equation:

(3) $$e_{ijh} = \alpha_1 + \beta_1 w_j + \delta_1 (w_j - \hat{w}) d_j + z_1 \gamma_1 + \varepsilon_1$$

where e_{ijh} is the level of effort for patient *i* at facility *j* consulted by health worker *h*, w_j is caseload per clinician at facility *j*, $\hat{w}$ is the threshold at which a further increase in caseload will imply a reduction in time use and effort per patient, d_j is a dummy that takes a positive value whenever $w_j > \hat{w}$ and z_1 is a vector of controls, capturing relevant characteristics of the health workers, their patients and the health facilities where they work.

Furthermore β_1 reflects the association between caseload and effort per patient when the time constraint does not bind, whereas $\beta_1 + \delta_1$ reflects this association in the case of a binding time constraint. If there is no simultaneity bias, i.e., if effort does not affect caseload, we expect that $\beta_1 = 0$ and $\beta_1 + \delta_1 < 0$, i.e., a negative association between caseload and effort if and only if the threshold level of workload is exceeded. We discuss simultaneity further below.

In the following, we first report results from the ordinary least square (OLS) regression[8], assuming that $\hat{w}$ is at a higher level than the maximum caseload we observe in our data. We do this because ultimately this regression is our preferred specification. We then use the instrumental variable (IV) approach to control for possible endogeneity of the caseload variable. The results show no signs of endogeneity. We therefore return to the OLS model, but now we extend the analysis by allowing for the possibility that there are some health facilities in our sample with a caseload above the threshold $\hat{w}$.

The univariate OLS regression with caseload as the single explanatory variable shows almost no association between effort and caseload; $\hat{\beta}_1$ is small and not significantly different from zero (Table 5, first column). From the R^2 reported in the same column, we see that caseload explains only 0.1% of the variation in the level of effort.

The low association between effort and caseload is robust to the inclusion of a number of controls at health worker, health facility and patient levels. There is no statistically significant association between effort and caseload in the multivariate OLS regression (Table 5, second column). This suggests that health workers are on average not constrained by high workloads in their practice. R^2 from this model is 0.3.

[8] All standard errors are adjusted for clustering at the facility level and stratification at the district level by using the svy-command in Stata 10.

Table 5. Regression results. Coefficients and standard errors. Dependent variable: Number of relevant diagnostic items performed.

	(1) OLS I	(2) OLS II	(3) IV
Caseload ($\hat{\beta}_1$)	0.010 (0.028)	0.016 (0.022)	0.015 (0.037)
Clinical officer		1.28** (0.51)	1.27** (0.50)
Male		0.25 (0.42)	0.25 (0.44)
Age		-0.03 (0.02)	-0.03 (0.02)
Imci_child		1.02** (0.48)	1.03** (0.49)
Government		-0.21 (0.43)	-0.21 (0.45)
Drugs		0.04 (0.13)	0.04 (0.13)
Laboratory		0.12 (0.45)	0.12 (0.48)
Child		1.26*** (0.40)	1.26*** (0.40)
Patient weakness		0.79*** (0.24)	0.79*** (0.23)
Patient number		-0.04*** (0.01)	-0.04*** (0.01)
Constant		2.84** (1.33)	2.86** (1.27)
Symptom fixed effect	No	Yes	Yes
n	2,095	1,806	1,806
R^2	0.001	0.300	0.300

**=p-value<0.1, **=p-value<0.05, ***=p-value<0.01. Sampling weights are used. Estimated standard errors take into account clustering at the facility level and stratification at the district level.*

We also observe that prescribers trained as clinical officers (or above) and/or who have IMCI training exert more effort per patient. At the mean level of effort (i.e., mean number of relevant diagnostic items performed), being trained as a clinical officer (or above) is associated with a 30% (95% CI: 6, 55) increase in effort per patient, while IMCI training is associated with a 24% (95% CI: 2, 47) increase in effort per patient in the IMCI target group. More effort is also exerted when the patient is a child; 30% (95% CI: 11, 49); or when the patient is very weak (as opposed to not weak); 38% (95% CI: 14, 60). The results also show greater effort when case complexity is higher, e.g., when patients present with more than one symptom (results not displayed). We find no significant associations between the level of effort and the sex and the age of health workers, facility ownership, availability of drugs, and the existence of laboratory. The magnitudes of the estimated coefficients are also quite small. Finally, we find that effort falls significantly the greater the number of patient consultations observed in advance. At the mean level of effort, for a patient to move ten places down the queue is associated with a 10% reduction in effort. This may indicate that there is a diminishing Hawthorne effect, but this pattern can also relate to tiredness or lower levels of concentration during the course of the day.

Reverse causality?

The apparent lack of association between effort and caseload can be due to a combination of a negative effect of caseload on effort and a positive effect of effort on caseload. If the causal relationship runs both ways as in this case, the equilibrium values of these two variables are determined in a simultaneous equations model (SEM), where both caseload (*w*) and effort (*e*) are endogenous. We thus have two stand-alone structural equations,

(4) $$e_{ijh} = \alpha_1 + \beta_1 w_j + z_1\gamma_1 + \varepsilon_1$$

(5) $$w_j = \alpha_2 + \beta_2 e_{ijh} + z_2\gamma_2 + \varepsilon_2$$

where z_1 and z_2 are vectors of control variables not necessarily equal.[9] The clinician decides the effort level, while caseload is determined by the number of people who need health care in the vicinity of the health facility as well as by patients choice to visit the health facility or not, both of which may be affected by the level of effort, as explained above. In this case, we cannot estimate equation (4) separately, because caseload will be correlated with the error term, thus violating an important assumption for unbiased OLS estimation.

Instrumental variable estimation (IV) provides a solution to the simultaneity problem. A valid instrumental variable (*x*) should 1) be uncorrelated with the error term in equation (4) (i.e., $\text{Cov}(x, \varepsilon_1) = 0$), and 2) affect the endogenous variable w (i.e., $\text{Cov}(x, w) \neq 0$ in equation (5)). A necessary and sufficient condition for identification of equation (4) is that we have a variable that is not included in equation (4) but that is important in equation (5). This is the rank condition for identification of a structural model.

We use the catchment population per full time health worker as an instrument for caseload per health worker in the OPD. Each health facility in Tanzania has a known and well-defined

[9] We omit the threshold component in equation (3) for simplicity.

catchment population, based on population data from the last Census. The catchment population may vary from a few thousand at the dispensary level up to several hundred thousands at the hospital level. In our sample, the catchment population varies from less than 1,000 to more than 400,000 (mean = 9,520).

In our data there is a strong positive relationship between catchment population and caseload. The relationship is nonlinear; the higher the catchment population, the smaller its impact on caseloads. This is as expected as the catchment area of hospitals, which typically have the largest catchment populations, will encompass the catchment population of lower level health facilities, because of their role as referral institutions. But for normal outpatient consultations, people will normally utilize the nearest facilities. Hence, for the type of consultations we consider here, the recruitment area of the largest facilities is not likely to include their entire catchment area. Moreover, in agrarian societies where population densities do not vary substantially, a high catchment population may be an indication that the catchment area is geographically large, implying longer averages distances and higher costs of seeking care. This may result in less demand. We take these nonlinearities into account by using the logarithm of the catchment population per health worker as our instrument.

There is little reason to believe that the catchment population itself affects the effort of clinicians and therefore correlates with the error term in equation (4). One possible reason for such an association would be that better clinicians seek areas with a large catchment population in order to establish profitable private clinics. This mechanism is not important in our setting as only 3% of the clinicians have external incomes from such practices. Indeed, we find no evidence of clinicians' selection into different areas based on factors correlated with the catchment population. Furthermore, it is unlikely that the quality of health services itself affects the catchment population. Migration in this setting is most likely determined by economic opportunities and family relations rather than by the quality of health service. In addition, differences in the quality of health services among health facilities are not likely to significantly affect catchment populations through differences in mortality rates. This is because the data on catchment population are from 2002 and because the dynamics involved in such an association will be very slow. Thus, we believe that catchment population is a valid instrument in our setting.

Table 6 displays the results from the first stage regression. We only report significant coefficients. Our instrument is a strong predictor of caseload (*p*-value < 0.0001). The standard deviation of the instrumental variable is 0.98. Hence, a one standard deviation variation in the instrumental variable causes about a 25% change in the caseload variable at the mean. The laboratory variable is also highly significant in this regression. Facilities with a laboratory have lower caseload.

Table 6. IV estimation. First stage regression. Dependent variable: Caseload.

Variables	Coefficients (standard errors)
log(catchment population per health worker)	4.87*** (0.86)
Laboratory	-5.04*** (1.78)
n	1,806
R^2	0.284

**=p-value<0.1, **=p-value<0.05, ***=p-value<0.01. Only coefficients with p-value < 0.1 are reported.*

Despite the strong statistical properties of our instrumental variable, the IV regression did not affect the result that there is no statistically significant association between caseload and effort (Table 5, third column). In fact, the estimated coefficient is almost exactly the same. We tested for endogeneity by including the residuals of the first stage regression into the second stage of a two-stage least squares (2SLS) estimation. The coefficient of the residual variable is not significant ($\beta = 0.001$, *p*-value = 0.979), suggesting that effort in the diagnostic process has no causal impact on caseload in our sample.

The result that there is no causal relationship between effort and caseload is robust to an expansion of our measure of effort to include tasks related to courtesy and communication (results not shown but available upon request).

Nonlinear relationship between caseload and effort?

We now return to our original specification from equation (3) with effort as a nonlinear function of caseload. Clinicians in our sample have a highly variable caseload, ranging between one and 45 patients per day. Although we have shown that there is no relationship between effort and caseload on average in our data, we know that when caseload becomes sufficiently high, clinicians will eventually have to compromise on effort per patient in order to be able to service them all. The question we ask here is whether we see any sign of such a threshold at the levels of caseloads reported in our data, or whether the threshold will kick in only at higher levels of caseload.

We ran successive regressions letting the threshold number of patients per clinician per day, $\hat{w}$, take on all integer values on the interval [1,45]. Our estimates of the slope of the effort / caseload function above the threshold $\hat{w}$, i.e., $\hat{\beta}_1 + \hat{\delta}_1$, were all close to zero and never statistically significantly different from zero. Estimates ranged from 0.041 (p = 0.753) with a threshold at 39 patients to -0.075 (p = 0.914) with a threshold at 44 patients (Table 7).

We calculated R^2 for each of the threshold levels in order to identify what threshold level fitted the data best. It is noteworthy that the model with a threshold was able to improve R^2 only from 0.2960 to a maximum of 0.2996 at the level of 12 consultations per health worker per day. Hence, the model with a threshold did not appear to provide any meaningful improvement in the model's fit.

We conclude that we cannot reject the null hypothesis that there is no association between caseload and effort, even at the margin.

Table 7. Estimates slope of effort / caseload function beyond the threshold level ($\hat{\beta}_1 + \delta_1$). OLS regression. Variables as in Table 5.

Threshold caseload per clinician per day ($\hat{w}$)	$\hat{\beta}_1 + \hat{\delta}_1$	*p*-value	R^2
1	0.016	0.456	0.2960
2	0.016	0.439	0.2965
3	0.016	0.455	0.2960
4	0.016	0.478	0.2960
5	0.014	0.546	0.2962
6	0.012	0.611	0.2967
7	0.011	0.668	0.2971
8	0.009	0.710	0.2974
9	0.008	0.766	0.2976
10	0.006	0.842	0.2977
11	0.001	0.965	0.2987
12	-0.003	0.938	0.2996
13	-0.003	0.935	0.2989
14	0.001	0.975	0.2975
15	0.004	0.913	0.2968
16	0.006	0.884	0.2965
17	0.008	0.852	0.2962
18	0.011	0.810	0.2961
19	0.013	0.774	0.2960
20	0.015	0.763	0.2960
21	0.016	0.757	0.2960
22	0.016	0.772	0.2960
23	0.014	0.809	0.2960
24	0.015	0.816	0.2960
25	0.011	0.875	0.2960
26	-0.002	0.977	0.2964
27	-0.003	0.976	0.2963
28	-0.003	0.972	0.2962
29	-0.005	0.962	0.2962
30	-0.007	0.955	0.2962
31	-0.007	0.953	0.2962
32	-0.002	0.991	0.2961
33	-0.003	0.987	0.2961
34	-0.003	0.988	0.2960
35	-0.001	0.997	0.2960
36	0.014	0.950	0.2960
37	0.017	0.939	0.2960
38	0.025	0.886	0.2960
39	0.041	0.753	0.2960
40	0.040	0.805	0.2960
41	0.029	0.883	0.2960
42	-0.015	0.947	0.2960
43	-0.030	0.930	0.2960
44	-0.075	0.914	0.2960
45	0.016	0.456	0.2960

6. Discussion

In a country with extremely few health workers per capita, it is reasonable to expect that the shortage of health personnel will have a negative impact on how thoroughly health workers examine and diagnose their patients. This assertion not only is intuitive and commonly held among health bureaucrats and analysts but is also an integral part of the story health personnel recount about their workdays.

Our data, however, tell a different story. Workload does not appear overwhelming, because patients are also few. In our sample, OPD clinicians spend on average less than two hours of their workday with patients (18.5 patients, 5.7 minutes each). Other duties, including administrative work, would not be enough to fill the rest of the workday. Admittedly, there are some clinicians who work only part time in the OPD, especially in hospitals, where they also may have duties in the in-patient wards. We have adjusted for part-time work by counting the number of full-time equivalent health workers in the OPD. It is also noteworthy that we found no significant difference between the average number of patients at the day of observation and the average number of patients recorded in March and August 2007, suggesting that there was nothing special about the time when we made our visits. Hence, there appears to be considerable slack capacity at the average health facility. This observation indicates that we should not be surprised to find a large portion of the clinicians on the horizontal segment of the caseload / effort curve in Figure 1.

There is substantial variation in caseload across facilities, from one to 45 patients per day. It is therefore conceivable that the estimated average impact of caseload on effort conceals a statistically significant impact at the margin. Our data rejects this hypothesis; no clinic appears to have reached the threshold where caseload compromises the quality of work. Even the busiest clinician in our sample would use less than 4 hours and 20 minutes per day on patient consultations at the average of 5.7 minutes per patient.

Our assertion that there is slack capacity follows from the fact that a normal working day is 7-8 hours and the assumption that health workers actually spend these hours at their workplace. However, several studies have shown that high rates of absenteeism in health facilities are common in many low-income countries (Chaudhury et al, 2006). One form of absenteeism is that health workers leave their duty posts early; for instance in order to conduct other businesses. In this case, the capacity constraint may bind even if the number of patients is low. Some 32% of the patients in our data told that they were not sure whether they would find a clinician at the health facility if they showed up at 1:00 pm or later. This indicates that some of the clinicians are working shorter hours than normal. Moreover, 75% of the health workers said that they conducted economic activities outside the health facility, most in the agricultural sector (61%). It is conceivable that they use part of the contracted working time at the health facilities for these activities. In this case, the real slack capacity is not possible to measure with our data. However, in Appendix 1 we show that if health workers are overworked due to low supply of working time, we would still expect a negative relationship between caseload and effort per patient. Our data thus do not support this explanation.[10]

In our attempts to isolate the causal effect of caseload and effort, we did not find any sign of a reverse causality. Several explanations are possible. First, caseload may respond to historic levels of effort rather than to the effort on the day of observation. We find this explanation unlikely, as the level of effort probably is highly correlated over time. A more likely explanation, in our view, is that patients do not necessarily take effort – measured by the number of relevant history taking questions and physical examination – as a signal of high quality. In interviews, health worker sometimes complained that patients tend to come with their own diagnosis and just want the doctor to prescribe drugs. Indeed, 95% of the health workers in our study agreed with the statement "*many patients prefer to get a confirmation of the diagnosis they think they suffer from*" and 84% agreed that "*most patients are dissatisfied if you do not prescribe drugs*". This suggests that our measure of effort is not necessarily what patients generally perceive of as high quality care. It is also interesting that 11% of the patients concede that patients who are waiting sometimes tell the doctor to hurry up. While the purpose of such behaviour is to reduce their own waiting time, it may at the same time display a lack of acknowledgement of the fact that it may take time to provide quality health services. Such attitudes have been reported in qualitative studies (Lindkvist et al, 2009).

It is not easy to reconcile our findings with the hypothesis that health workers underperform in their clinical work because they are overworked. But why then do the health workers

[10] It is possible that health workers who would normally leave their duty post early, behaved differently in our presence, staying for a longer time and treating their patients with the same level of effort as their colleagues with fewer patients and/or longer effective working hours. This could explain why we were unable to uncover real capacity constraints due to absenteeism. We would expect, however, that habits also play a considerable role for actual behaviour, making it difficult to mask normal practice completely.

themselves advance this view (Lindkvist, *op. cit.*)? One explanation could be that our research area is different, with lower demand for health services and/or a higher number of health workers per capita, from areas where Lindkvist et al conducted their research (i.e., Temeke - an urban district in Dar es Salaam, and Kisarawe - a rural district in the vicinity of Dar es Salaam). The general pattern in Tanzania is that urban districts have *more* health workers per capita than the rural areas (Munga and Mæstad, 2009). On the other hand, urban demand for health services could also be higher, for instance because of higher levels of education and shorter travelling distances. It is therefore conceivable that health workers in the urban areas have a heavier workload than their rural colleagues, but we do not have data to support this proposition.

Second, the perception of being overworked may relate to the lack of appointment systems. Patients typically queue up during the morning hours – which could be a rational response to their belief that health workers are more likely to be absent later in the day. When many patients show up more or less at the same time, there may be a pressure on health workers "*to rush in order to catch up with the big number of patients waiting*". As many as 48% of the clinicians said that other patients would complain if the doctor spent too long time with each patient. We expect, however, that health workers themselves would be able to locate this problem more precisely than to label it only as "*too many patients*".

Finally, the perception of being overworked could relate to the low effective supply of working time, for instance because health workers leave their job earlier than they are supposed to. But, as discussed above, our data does not support this explanation.

In a different context, it has been suggested that high caseload can have a *positive* impact on clinical performance, because more patients imply more training, which improves the skills and thus the performance of the doctor (Saxena et al, 2007). It is not likely that this mechanism is relevant for the type of work we observe in our study, as our focus is only on the most common symptoms that every doctor frequently encounters.

It is striking that only 22% of the assessment tasks required by guidelines are performed on average. We find that clinical officers perform significantly better than less trained health personnel and that IMCI training improves performance. These findings are in line with the classical paradigm that poor performance is caused by lack of knowledge and skills (e.g., Brugha and Zwi, 1998) and with evaluations of the impacts of IMCI (El Arifeen et al, 2004; Gouws et al, 2004; Tanzania IMCI Multi-Country Evaluation Health Facility Survey Study Group, 2004). However, clinical officers in our sample do not perform more than 25% of the items required by protocol, the same as the average for IMCI trained personnel. Hence, our results point in the direction that training alone is insufficient to achieve adequate levels of performance (Rowe et al, 2005; Leonard et al, 2007. See also Perades et al, 2006; Rowe et al, 2001, 2003).

7. Robustness analysis

Alternative measures of effort (OLS and IV models without threshold)

We employed three alternative specifications of effort per patient. First, we decomposed our measure of effort into a) the number of relevant history taking questions and b) the number of relevant examinations and ran the analysis separately for each category. Second, we constructed a measure of the number of time consuming tasks performed that were *not* part of the diagnostic process. These tasks include welcoming and greeting the patient, informing the patient of his or her diagnosis, explaining the treatment provided, providing health education related to the diagnosis, and explaining whether to return for further treatment. We used this variable both alone and in combination with our original measure of effort. Finally, we used time use per patient as an alternative measure of effort. We found no statistically significant association between caseload and effort per patient in any of these specifications, neither in the OLS nor in the IV model.

We also estimated the effect of caseload on each individual diagnostic item contained in our measure of effort. Among the 62 diagnostic items that were recorded, we found no statistically significant association with caseload for 46 items. Among the remaining 16 items, eight were negatively associated with caseload and eight were positively associated with caseload at conventional levels of significance (see Appendix 2). Hence, even at the level of individual items, the general pattern is that the association between caseload and effort is very weak. However, the fact that a few items appear to be positively associated with caseload while others are negatively associated with caseload may indicate that there is some degree of substitution between diagnostic items as caseload increases. Substitution from more to less time-consuming items would indicate that the time constraint is binding, though in a weaker sense than defined so far. We explore this possibility further below.

Exogenous variables (OLS model without threshold)

The MAP data set contains a large number of variables (other than those that we have used in our preferred specification) that may capture aspects of health worker knowledge, extrinsic and intrinsic motivations, and patient characteristics, and thus be related to the level of effort per patient. In an attempt to test whether our results hinges on our particular selection of exogenous variables, we ran regressions where we included each of the potentially relevant (yet excluded) variables in turn and investigated the effects on our estimates of the impact of caseload on effort per patient.

Appendix 3 provides a list of the included variables. They comprise:

- 3 variables describing education, knowledge and knowledge sharing,
- 9 variables capturing health workers' perceptions about patient expectations,
- 18 variables describing various aspects of management and supervision,
- 8 variables characterizing the relationship between effort and monetary incentives,
- 8 variables capturing aspects of intrinsic motivation, and
- 12 variables characterizing individual patients.

The inclusion of these variables did not have any impact on the result that there is no association between caseload and the level of effort.

Alternative threshold analysis

As discussed in Section 2, the threshold of workload at which health workers will begin reducing their effort per patient may differ across workers. Such heterogeneity may create a bias in the estimation of the slope of the effort/caseload curve. Assume that the individual thresholds vary between $\underline{\theta}$ and $\overline{\theta}$. It is easy to show that a uniform and exogenously imposed threshold, like the one we have used in our analysis, will bias the estimate of $\beta_1 + \delta_1$ upwards as long as the threshold is set lower than $\overline{\theta}$. Similarly, a threshold that is set above $\underline{\theta}$ will bias the estimate of β_1 downwards.

In order to mitigate this problem, we ran the threshold model (equation (3)) while omitting observations in the neighbourhood of the exogenously imposed threshold level. We omitted observations with a caseload of $\pm n$ patients relative to the defined threshold. We attempted $n = (2,4,6)$, but none of these specifications led to any significant association between caseload and effort.

We also ran the threshold analysis using IV estimation in order to check that our failure to identify reverse causality did not relate to model misspecification. This analysis requires two instrumental variables, as the caseload variable now enters the estimation equation both directly and indirectly through the threshold variable (see equation (3)). Following Wooldridge (2002, p. 237), we used as our second instrument a variable defined as a function of the predicted caseload variable from the first-stage regression. Let $\tilde{w}$ denote the predicted caseload of the first stage regression. As an instrument for $(w - \hat{w})d$ we used $(\tilde{w} - \hat{w})d$. We were also unable to indentify any association between caseload and effort per patient in this analysis.

Finally, in order to investigate whether there is any substitution between different types of diagnostic items as the caseload grows larger, we ran the threshold analysis separately for a) the number of relevant history-taking questions and b) the number of relevant examinations. It seems natural to expect that physical examinations on average are more time consuming than asking questions. In addition, some examinations and questions are partly substitutable. Therefore, a health worker who is (weakly) constrained by a high caseload may choose to ask more questions and to perform fewer examinations and still keep the aggregate number of diagnostic items at the same level as he or she would without such constraints. We observe this pattern in the data (Table 8).

As the caseload reaches about 40 patients, there is a tendency to reduce the number of physical examinations and correspondingly to increase the number of history taking questions. One interpretation of these findings is that the clinicians with the highest workloads are approaching the limits of what they can handle without reducing their level of effort per patient. They are effectively constrained, but only slightly so, implying that they are able to maintain their effort level by substituting from more time-consuming to less time-consuming diagnostic items. Note, however, that there are only five clinicians in our data with 37 patients or more and only two with 40 patients or more. Hence, the observed patterns may also result from specific health worker characteristics that we are unable to control for.

Table 8. Threshold model with effort measured as a) the number of relevant history taking questions and b) the number of relevant physical examinations.

Threshold caseload per clinician per day ($\hat{w}$)	*Dependent variable:* Questions			*Dependent variable:* Examinations		
	$\hat{\beta}_1 + \hat{\delta}_1$	*p*-value	R^2	$\hat{\beta}_1 + \hat{\delta}_1$	*p*-value	R^2
37	0.03	0.836	0.2753	-0.01	0.94	0.2008
38	0.06	0.541	0.2755	-0.04	0.71	0.2009
39	0.14	0.075	0.2760	-0.10	0.22	0.2013
40	0.17	0.074	0.2760	-0.13	0.15	0.2014
41	0.22	0.066	0.2760	-0.19	0.06	0.2015
42	0.27	0.040	0.2758	-0.28	0.01	0.2016
43	0.40	0.049	0.2758	-0.43	0.02	0.2016
44	0.78	0.059	0.2758	-0.85	0.02	0.2016
45			0.2753			0.2008

Alternative instrumental variable (IV model without threshold)

We tested an alternative specification of the instrumental variable by replacing the *log*-specification by 10 dummy variables to capture better the nonlinear relationship between caseload and catchment population. This alternative specification slightly reduced the R^2 of the first-stage IV regression from 0.2956 to 0.2942 and therefore was considered to be inferior to our original specification.

8. Concluding remarks

In our data, high workload does not reduce health workers' effort in the diagnostic process. Not even in clinics with the highest workload does the number of patients per clinician impact on the assessment quality. After examining our data using numerous tests, we are confident that this conclusion is robust. Another question is the external validity of our findings. Our data consists of a sample of outpatient consultations in the Morogoro and Dodoma regions, located in central Tanzania. These regions are representative of rural Tanzania in terms of the average number of health workers per capita. Whether the demand for health services differs from other regions is unknown.

Our findings have clear policy implications. First, our study suggests that sending more health workers to rural Tanzania – with the same qualifications as the workers already there – is unlikely to enhance the quality of the health care in outpatient consultations. On a more constructive level, our results suggest that it would help to increase the prescribers' level of training. We find that clinical officers provide significantly higher quality-enhancing efforts than less trained health personnel. However, the difference is not very large, suggesting that other measures are also needed to raise quality to an adequate level. One alternative would be to incentivize good performance. We do not have data to test this conjecture, but the know-do gap identified by Leonard et al (2007) and others indicates that this is a viable way to improve quality (see Meessen et al, 2006, on experiences from implementing monetary incentives in Rwanda). However, one should not downplay the difficulty of using standard (monetary) incentives in a profession where effort choices are influenced by ethical and professional standards and where it is difficult to find objective (observable and verifiable) quality measures that can be used as a basis for rewards. Alternatively, it is possible to use softer, more subjective criteria to reward health worker performance, and the rewards can be non-pecuniary encouragements that motivate by, for instance, enhancing the recognition and status of good performance.

Even if we find that workload is not a critical factor explaining the low quality of health care in rural Tanzania, the lack of health personnel could rapidly become a bottleneck. If the government chooses policies that effectively enhance the quality of health care, many more patients may seek assistance at health facilities, and the workload could soon become overwhelming. What we have shown is that we are not yet there, that there is substantial slack at rural health facilities, and that demand can increase quite a bit before the number of health workers becomes a constraint for the quality of health care.

References

Amin S, Das J and Glodstein M (eds.) 2008. *Are you being served? New tools for measuring service delivery*. Washington DC: The World Bank.

Barber SL, Gertler PJ and Harimurti P 2007. The contribution of human resources for health to the quality of health care in Indonesia. *Health Affairs* **26**:w367–w379.

Black BE, Morris SS and Bryce J 2003. Where and why are 10 million children dying every year? *The Lancet* **361**(9376):2226–2234.

Brugha R and Zwi A 1998. Improving the quality of private sector delivery of public health services: challenges and strategies. *Health Policy and Planning* **13**:107–120.

Chaudhury N, Hammer J, Kremer M, Muralidharan K and Rogers FH 2006. Missing in action: teacher and health worker absence in developing countries. *Journal of Economic Perspectives* **20**(1):91–116.

Das J and Hammer J 2007. Money for nothing: the dire straits of medical practice in Delhi, India. *Journal of Development Economics* **83**:1–36.

Das J, Hammer J and Leonard K 2008. The quality of medical advice in low-income countries. *Journal of Economic Perspectives* **22**:93–114.

Census (2002) *Tanzania 2002 Population and Housing Census*. Dar es Salaam: National Bureau of Statistics.

El Arifeen S, Blum LS, Hoque DME, Chowdury EK, Khan R, Black RE, et al 2004. Integrated Management of Childhood Illness (IMCI) in Bangladesh: early findings from a cluster-randomized study. *Lancet* **364**:1595–1602.

Gouws E, Bryce J, Habicht JP, Amaral J, Pariyo G, Schellenberg JA, et al 2004. Improving antimicrobial use among health workers in first-level facilities: results from the Multi-Country Evaluation of the Integrated Management of Childhood Illness strategy. *Bull World Health Organ* **82**:509–515.

HRH Census 2001/2002. *Tanzania Human Resources for Health Census 2001/2002.* Dar es Salaam: Ministry of Health and Social Welfare.

Leonard K, Masatu MC and Vialou A 2007. Getting doctors to do their best. *The Journal of Human Resources* **42**:682–700.

Leonard K, Mliga G and Mariam DH 2002. Bypassing health centres in Tanzania: revealed preferences for quality. *Journal of African Economies* **11**(4):441–471.

Leonard K and Masatu MC 2006. Outpatient process quality evaluation and the Hawthorne Effect. *Social Science & Medicine* **63**:2330–2340

Lindkvist I, Mwisongo A, Blystad A and Mæstad O 2009. Performance of health workers in Tanzania: results from qualitative research. *Unpublished draft.*

Madden JM, Quick JD, Ross-Degnan D, Kafle KK 1997. Undercover careseekers: simulated clients in the study of health provider behavior in developing countries. *Social Science & Medicine* **45**:1465–82.

Masanja H, de Savigny D, Smithson P, Schellenberg J, et al 2008. Child survival gains in Tanzania: analysis of data from demographic and health surveys. *Lancet* **371**:1276–83.

Meessen B, Musango L, Kashala J-PI, Lemlin J 2006. Reviewing institutions of rural health centres: the Performance Initiative in Butare, Rwanda. *Tropical Medicine and International Health* **11**:1303–1317.

Munga M and Mæstad O 2009. Measuring inequalities in the distribution of health workers: the case of Tanzania. *Human Resources for Health* **7**:4.

Mwisongo A and Mæstad O 2009. Missing the essential? The practice of rural clinicians. *Unpublished draft.*

Mæstad O and Mwisongo A 2007. Informal payments and the quality of health care in Tanzania: results from qualitative research. CMI Working Paper 2007:5.

Naimoli JF, Rowe AK, Lyaghfouri A, Larbi R and Lamrani LA 2006. Effect of the Integrated Management of Childhood Illness strategy on health care quality in Morocco. *International Journal for Quality in Health Care* **18**:134–44.

Ofori-Adjei D and Arhinful DK 1996. Effect of training on the clinical management of malaria by medical assistants in Ghana. *Social Science & Medicine* **42**:1169–76.

Osterholt DM, Rowe AK, Hamel MJ, Flanders WD, Mkandala C, Chizani N and Marum L 2006. Predictors of treatment error for children with uncomplicated malaria seen as outpatients in Blantyre District, Malawi. *Tropical Medicine and International Health* **11**:1147–56.

Paredes P, Pena M, Flores-Guerra E, Diaz J and Trostle J 1996. Factors influencing physicians' prescribing behavior in the treatment of childhood diarrhoea: knowledge may not be the clue. *Social Science & Medicine* **42**: 1141–53.

Rowe AK, Hamel MJ, Flanders WD, Doutizanga R, Ndoya J and Deming MS 2000. Predictors of correct treatment of children with fever seen at outpatient health facilities in the Central African Republic. *American Journal of Epidemiology* **151**:1029–1035.

Rowe AK, Onikpo F, Lama M, Cokou F and Deming MS 2001. Management of childhood illness at health facilities in Benin: problems and their causes. *American Journal of Public Health* **91**:1625–1635.

Rowe AK, Lama M, Onikpo F and Deming MS 2003. Risk and protective factors for two types of error in the treatment of children with fever at outpatient health facilities in Benin. *International Journal of Epidemiology* **32**:296–303.

Rowe AK, de Savigny D, Lanata CF and Victora CG 2005. How can we achieve and maintain high-quality performance of health workers in low-resource settings? *The Lancet* **366**(9490):1026–1035.

Saxena S, Car J, Eldred D, Soljak M and Majeed A. 2007. Practice size, caseload, deprivation and quality of care of patients with coronary heart disease, hypertension and stroke in primary care: national cross-sectional study. *BMC Health Services Researh* **7**:96–96.

Tanzania IMCI Multi-Country Evaluation Health Facility Survey Study Group 2004. The effect of Integrated Management of Childhood Illness on observed quality of care of under-fives in rural Tanzania. *Health Policy and Planning* **19**:1–10.

TSAM (2007) *Tanzania Service Provision Assessment Survey 2006.* Dar es Salaam: National Bureau of Statistics and Macro International.

Wedig G, Mitchell JB and Cromwell J 1989. Can Optimal Physician Behaviour be Obtained Using Price Controls? *Journal of Health Policy, Politics and Law* **14**(3):601–620.

WDI 2008. World Development Indicators 2008. Washington DC: The World Bank.

WHO 2006. *Working Together for Health. World Health Report 2006.* Geneva: World Health Organization.

Wooldridge JM 2002. *Econometric analysis of cross section and panel data.* Boston, MA: MIT Press.

Zurovac D, Rowe AK, Ochola SA, Noor AM, Midia B, English M and Snow RW 2004. Predictors of the quality of health worker treatment practices for uncomplicated malaria at government health facilities in Kenya. *International Journal of Epidemiology* **33**:1080–1091.

Appendix 1. Effort and workload with endogenous work supply.

We are interested in the effect of an exogenous change in workload on effort when work supply is endogenous. Let workload be defined by $w(e,\theta) = w(e) + \theta$, where θ captures exogenous factors that shift the demand for health care. With endogenous work supply, a health worker solves $\max_{e,l} u(e,l)$ s.t. $ew(e,\theta) \leq l$. Whenever the constraint is binding, l is a function of e, and the problem can be written as:

$$\max_{e} u(e, l(e)),$$

where $l = e[w(e) + \theta]$. The first-order condition for a maximum is:

$$g = u_e + u_l \frac{\partial l}{\partial e} = 0.$$

The effect on effort of an exogenous increase in the workload is given by:

$$\frac{de}{d\theta} = -\frac{g_\theta}{g_e}$$

The second-order condition for a maximum implies $g_e < 0$. Hence, the sign of $de/d\theta$ is equal to the sign of g_θ. By differentiating and using the first-order condition, we get $g_\theta = u_{el} - u_{ll}\frac{u_e}{u_l} + u_l$. Note that $u_l > 0$ is not consistent with a utility maximum, since the health worker then can increase her utility by increasing l and still fulfil the constraint. With $u_l < 0$, a quasi-concave utility function implies $u_{ll}\frac{u_e}{u_l} > u_{el}$. It follows that $g_\theta < 0$ and $\frac{de}{d\theta} < 0$.

Appendix 2.

Diagnostic and non-diagnostic items observed. Share of patients exposed to each item. Coefficient and p-values of the estimated relationship between caseload and each individual item. Control variables as in Table 5 have been used in all cases.

	Share of patients exposed	Coeff	*p*-value
All patients (n = 3,494)			
Welcoming			
Welcome the patient	0.745	0.0040	0.224
Greet the patient	0.632	0.0073*	0.064
All patients presenting with fever (n = 1,228)			
History taking questions			
Duration of fever	0.841	-0.0036	0.119
Whether temperature has been taken	0.075	0.0005	0.774
Pattern (periodicity) of fever	0.187	0.0030	0.169
Presence of chills, sweats	0.021	-0.0007**	0.034
Presence of cough, sore throat, pain during swallowing	0.339	0.0012	0.720
Presence of diarrhoea or vomiting	0.484	0.0042	0.199
Presence of convulsions	0.155	0.0012	0.617
Presence of running nose	0.061	0.0016	0.114
Physical examinations			
Take temperature (with thermometer)	0.383	-0.0030	0.442
Check neck stiffness	0.021	-0.0015**	0.021
Look for palmar pallor (anemia)	0.245	0.0012	0.710
Check ear/throat	0.028	0.0006	0.415
Palpate for the spleen	0.025	-0.0002	0.729
Children <5 years presenting with fever (n = 849). Additional items			
History taking questions			
Ability to drink or breastfeed	0.296	0.0061**	0.045
Difficulty in breathing	0.072	-0.0007	0.630
Presence of ear problems	0.083	-0.0042*	0.080
Vaccination history	0.153	0.0002	0.941
Physical examinations			
Check for lethargy or unconsciousness (try to wake up the child)	0.034	-0.0024**	0.038
Check for visible severe wasting	0.040	-0.0001	0.946
Look for oedema of both feet	0.017	0.0001	0.560
Check the child's weight (against a growth chart)	0.400	0.0060*	0.084
All patients presenting with cough (n = 820)			
History taking questions			
Duration of cough	0.785	0.0014	0.691
Sputum production or dry cough	0.134	0.0028*	0.083
Presence of blood in sputum	0.033	0.0004	0.594
Presence of chest pain	0.092	0.0002	0.879
Presence of difficulty in breathing	0.156	0.0010	0.599
Presence of fever	0.461	0.0061	0.174
Physical examinations			
Count respiratory rate	0.111	-0.0070***	0.002
Observe breathing for lower chest wall indrawing	0.130	-0.0012	0.470
Examine throat	0.029	0.0001	0.875
Auscultate the chest	0.200	0.0000	1.000
Take temperature (with thermometer)	0.118	-0.0028	0.264
Children <5 years presenting with cough (n = 526). Additional items			
History taking questions			
Ability to drink or breastfeed	0.245	0.0067*	0.051

Presence of convulsions	0.111	-0.0006	0.764
Presence of ear problems	0.076	-0.0030	0.220
Presence of diarrhoea or vomiting	0.337	0.0059	0.141
Vaccination history	0.166	0.0048*	0.086
Physical examinations			
Check for lethargy or unconsciousness (try to wake up the child)	0.025	-0.0011	0.308
Check for visible severe wasting	0.010	-0.0009**	0.033
Look for palmar pallor	0.111	0.0031	0.245
Look for oedema of both feet	0.008	-0.0001	0.583
Check the child's weight (against a growth chart)	0.255	0.0015	0.697
All patients presenting with diarrhoea (n = 381)			
History taking questions			
Duration of diarrhoea	0.760	0.0008	0.813
Frequency of stools	0.526	0.0014	0.690
Consistency of stools	0.237	-0.0078**	0.030
Presence of blood, and/or mucus in stools	0.353	0.0075**	0.038
Presence of vomiting	0.275	0.0127***	0.000
Presence of fever	0.471	0.0072	0.109
Physical examinations			
Assess general health status (alert or lethargic)	0.093	-0.0046*	0.076
Examine for sunken eyes	0.181	0.0038	0.340
Pinch abdominal skin to assess degree of dehydration	0.197	0.0006	0.877
Take temperature (with thermometer)	0.192	-0.0011	0.834
Children <5 years presenting with diarrhoea (n = 298). Additional items			
History taking questions			
Ability to drink or breastfeed	0.300	0.0067**	0.035
Presence of convulsions	0.083	0.0029	0.216
Presence of ear problems	0.034	-0.0007	0.519
Presence of cough or difficulty in breathing	0.152	0.0001	0.964
Vaccination history	0.147	0.0034	0.357
Physical examinations			
Offer the child a drink of water or observe breastfeeding	0.050	-0.0003	0.801
Check for visible severe wasting	0.012	-0.0002	0.775
Look for palmar pallor	0.160	0.0077	0.114
Look for oedema of both feet	0.008	-0.0001	0.806
Check the child's weight (against a growth chart)	0.371	0.0053	0.254
All patients (n = 3,494)			
Communication			
Look at the patient while talking	0.846	0.0018	0.290
Tell the patient his or her diagnosis (any name)	0.227	-0.0007	0.718
Explain the diagnosis (in common language)	0.163	0.0004	0.786
Explain the treatment being provided	0.292	0.0011	0.660
Give any health education related to the diagnosis?	0.160	0.0054***	0.001
Explain whether or not to return for further treatment	0.253	0.0037	0.122
Listen properly to the patient/caregiver	0.656	0.0073**	0.030
Allow the patient to talk	0.747	0.0057***	0.008
Ensure the patient had understood diagnosis, etc	0.132	-0.0007	0.678

Appendix 3. Alternative variables uses in robustness analysis.

Category	Variable
Knowledge	Primary and secondary education
	Degree of knowledge sharing at health facility
	Perceived lack of knowledge to form a correct diagnosis and treatment
Perceived patient expectations	Patients prefer the doctor to finish quickly
	Patients will complain if doctor spends much time with each patient
	Most patients do not like to be touched by the doctor
	Most patients do not like the doctor to ask many questions
	Most patients do not like to be educated on health issues
	Most patients are dissatisfied if the doctor does not prescribe drugs
	According to the patients, a good doctor spends a long time on each patient
	Many patients prefer to get a confirmation of the diagnosis they think they suffer from
	Patients complain to the community leaders if they are dissatisfied with the quality of health services
Management and supervision	Clinician has a clear job description
	There is a fair procedure for selecting health workers to attend workshops and seminars
	The workload at the facility is fairly distributed
	Performance is supervised by someone from the clinic (internal supervision)
	The internal supervisor regularly observes OPD consultations
	Which factors is the internal supervisor interested in observing in the OPD?
	The internal supervisor provides valuable advice on clinical issues
	The internal supervisor takes genuine interest in the clinician's work
	The internal supervisor assists in rectifying problems at your workplace
	The management treats workers respectfully
	Quality of management (assessed by health worker)
	The external supervisor sometimes observes OPD consultations
	Frequency of external supervision
	Which factors is the external supervisor interested in observing in the OPD?
	The external supervisor provides valuable advice on clinical issues
	The external supervisor takes genuine interest in the clinician's work
	The external supervisor assists in rectifying problems at your workplace
	Quality of management (assessed by surveyors)
Monetary incentives and effort	User fees for OPD consultations and tests
	Lower user fees may reduce health workers salaries
	User fees may be used to improve working conditions
	Large reduction in the number of patients may reduce or delay salary
	Careful diagnosis and treatment of patients may result in higher salary
	Those who provide quality care are chosen to attend seminars
	Doing a good is important for being promoted
	Good performance increases opportunities for further studies
Intrinsic motivation	Would rather work outside the health sector if I could increase monthly salary by 100,000 Tsh
	I speak of this facility as a great place to work to all my friends
	I never work longer than necessary
	I have donated blood
	I have helped carry a stranger's belonging
	I have let a neighbour whom I did no know that well borrow an item of some value to me
	I have helped a classmate whom I did not know well with a homework assignment
	I have before being asked voluntarily looked after a neighbor's children without being paid for it
Patient characteristics	Self-assessed level of sickness
	Sex
	Age
	Level of primary/secondary education

	Time travelled to health facility
	Household owns a radio
	First time visit at health facility
	Sometimes visits other health facilities
	Never uses an existing closer facility
	Has a personal relationship with clinician
	Prefers the doctor to be quick
	Prefers the doctor to perform a thorough examination

Department of Economics
University of Bergen
Fosswinckels gate 6
N-5007 Bergen, Norway
Phone: +47 55 58 92 00
Telefax: +47 55 58 92 10
http://www.svf.uib.no/econ