Overworked? On the relationship between workload and health worker performance

Ottar Mæstad [a,*], Gaute Torsvik [b], Arild Aakvik [c]

[a] Chr Michelsen Institute, Bergen, Norway
[b] Department of Economics, University of Bergen and Chr Michelsen Institute, Bergen, Norway
[c] Department of Economics, University of Bergen, Norway

ARTICLE INFO

ABSTRACT

The 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 high caseload reduces the level of effort per patient in the diagnostic process. We observed 159 clinicians in 2095 outpatient consultations at 126 health facilities in rural Tanzania. Surprisingly, we find no association between caseload and the level of effort per patient. 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 is unlikely to raise the quality of health services. Training has a positive effect on quality but is not in itself sufficient to raise quality to adequate levels.

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. 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.

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, 2010). 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[1]: "...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].

The neo-classical model – where all markets clear – has no notion of shortage. Any shortage will be due to non-market clearing prices or other market failures. One possible reason for a health worker shortage is thus that governments have fixed user fees below the market clearing price. However, WHO's conception of a critical health worker shortage has a different origin; it derives from the idea that governments are aiming to achieve certain health outcomes and that the current supply of health workers is insufficient for attaining these goals, basically because there are binding

---

* Corresponding author. Tel.: +47 4793 8081.
  E-mail address: ottar.mestad@cmi.no (O. Mæstad).

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

constraints on the amount of time and effort that health workers are able (or willing) to put into their work.

The alternative conceptions of a health worker shortage both imply that an increase in the number of health workers would improve health outcomes. The existence of a health worker shortage can thus be tested by studying the impact of health outcomes of an exogenous increase in health worker labour supply (while keeping the price of health services fixed). Such a test would require fairly lucky natural experiments. This paper proposes an alternative, and replicable, approach based on studying the relationship between process outcomes (assessment quality) and caseload. We use a new data set from rural Tanzania, a country defined by WHO (2006) to have a critical shortage of health workers. 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, 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. 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. Second, 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.[2] We propose to deal with this issue by using the catchment population of the health facility (per clinician) as an instrument for caseload (per clinician).

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 et 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). Traditional measures, focusing on physical inputs, have obvious shortcomings (Amin et al., 2008). In order to better capture the knowledge of health workers and the efforts they put into their practice, recent studies have measured the quality of care through observation of actual or 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, 2007; Leonard et al., 2007; Barber et al., 2007).[3]

We 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 2095 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. 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 describes the 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 data collection. Section 5 presents descriptive statistics and the results of the regression analyses. We present robustness analyses in Section 6, discuss our main findings in Section 7, and conclude in Section 8.

2. The context

Tanzania is a low-income country with life expectancy at birth at 51.9 years and infant mortality at 73.6 per 1000 live births (WDI, 2008). Major causes of premature deaths among children include acute respiratory infections, malaria, and diarrhoea, conditions that normally can be cured by simple, low-cost treatments (Black et al., 2003). 80% of the population lives in rural areas (Census, 2002).

The health care system includes 219 hospitals, 481 health centres and 4679 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).

The total number of health workers in the country is 1.4 per 1000 inhabitants. The number of doctors (physicians), nurses and midwives per 1000 is 0.4–0.6 (depending on whether assistant medical officers and clinical officers are included among the doctors), far below the threshold for a critical shortage defined by the WHO of 2.5 workers. In rural areas, clinical officers with 3 years of clinical training provide most clinical services, but it is also common that nurses and assistants, with little or no clinical training, carry out clinical work.

Our study area includes all nine rural districts in 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

---

[2] There is some 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).

[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).

The area has 440 health facilities owned by the government (81%) and Christian voluntary agencies (19%). The average health worker density in the area is 1.0 health workers per 1000 inhabitants, lower than the national average of 1.4, and also lower than the average of 1.1 health workers per 1000 inhabitants in rural districts.

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. 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

Consider a health worker choosing the time to spend at the clinic (l) and the effort (e) to exert in each consultation. Effort represents all 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. To ease exposition we assume that e is equal to the time spent per patient.

Effort will impact on the demand of 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 thus reduce reattendances and thereby the total number of consultations. We assume that demand (i.e., the number of patients showing up at the clinic per day) is given by $w(e, \theta) = w(e) + \theta$, where θ captures exogenous factors affecting the demand for health care.

Suppose health workers choose l and e to maximize their work satisfaction, $u(e,l)$, subject to the constraint that all patients must be attended, $ew(e, \theta) \leq l$. Exerting effort generates both benefits and costs for individual workers. The gains come as intrinsic and/or extrinsic rewards associated with the delivery of high quality health services, the costs from the fact that it is psychologically and physically demanding to provide high quality on a regular basis. Spending time at the clinic also generates costs (forgone income from alternative work) and benefits (fulfilling job description, improving chances for promotion, etc.). Hence, we assume that preferences are such that a health worker not facing the "all-patients-must-be-attended" constraint would choose work effort $e^* > 0$ and work time $l^* > 0$: $u_e(e^*, l^*) = u_l(e^*, l^*) = 0$. The level of e^* and l^* will vary among health workers, depending on their intrinsic motivation and extrinsic incentives. If the constraint does not bind, i.e. if $e^*w(e^*, \theta) < l^*$, the health worker has idle time at the clinic, and an exogenous increase in w will not impact on the quality of health care (effort is given by e^*).

If the constraint binds, i.e., if $e^*w(e^*, \theta) \geq l^*$, the health worker must adjust effort and work time to consult all patients. The time spent at the work place can then be written as a function of effort per patient; $l(e)$ defined by $l = ew(e, \theta)$. The first order condition for a constrained maximum is given by

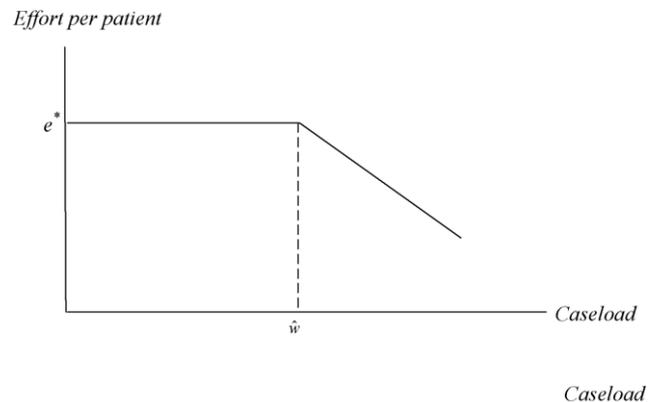$$g(e, \theta) = u_e + u_l \frac{\partial l}{\partial e} = 0.$$

Suppose there is an exogenous increase in the demand for health care. The effect on effort is

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

The second order condition for a maximum implies $g_e < 0$. The sign of $de/d\theta$ is therefore equal to the sign of g_θ. By differentiating and using the first order condition we get

$$g_\theta = \left(u_{el} - u_{ll} \frac{u_e}{u_l} \right) l_\theta + u_l l_{e\theta}.$$

Since l_θ is positive, the first term is negative as long as exerting quality-enhancing effort is a normal good. In the second term, u_l is negative, and with $w(e, \theta) = w(e) + \theta$ we have $l_{e\theta} = 1$ Hence, $de/d\theta < 0$ under reasonable assumptions; an exogenous increase in caseload will reduce effort per patient (*the intensive margin*). It may also cause an increase in time spent at the clinic (*the extensive margin*), but not to the extent that it will obviate the need to reduce the level of effort per patient.

Fig. 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.

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 randomly selected in each district from six strata defined by the type of facility (hospital, health centre and dispensary) and ownership (government and voluntary agencies) (see Table 1).

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 2 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

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

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

medical schools in Dar es Salaam were used as surveyors after a 1-week training session. 3494 consultations were observed in total. We measured assessment quality for the 2095 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.

During each consultation, surveyors noted which tasks – among a set of pre-defined relevant tasks – that were 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 A).

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 2 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.

5.1. Key variables: definitions and descriptive statistics

Following the approach outlined in Das et al. (2008), our dependent variable is the quality of health services, here defined as the level of effort exerted in the diagnostic process. Effort 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. Our data set does not contain this information, though.[5]

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 presents summary statistics on effort and caseload. On average, clinicians ask 2.94 relevant questions and undertake 1.26 physical examinations per patient. 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 min. This includes the time taken to fill prescriptions and patient cards.

Although we are primarily interested in 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 (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 workers from lower cadres, mostly nurses and assistants. Health workers trained as a clinical officer or above take care of 69% of the patients. 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

---

[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.

[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 (2010) provide an in-depth discussion of which questions that were asked and which examinations that were performed.

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

Variable	Variable definition	n	Mean (weighted)	Mean (unweighted)	S.D.	Min.	Max.
Questions	Number of history taking questions (a)	2095	2.94	2.92	1.88	0	12
Examinations	Number of physical examinations (b)	2095	1.26	1.13	1.35	0	15
Effort	(a) + (b)	2095	4.20	4.04	2.76	0	22
Time	Minutes per patient	1789	5.66	5.80	3.74	0	45
Caseload	Number of OPD patients per full time OPD health worker per day	2095	18.48	16.36	9.76	1	45

Table 4
Descriptive statistics. Control variables.

Variable	Variable definition	#Obs.	Mean	S.D.	Min.	Max.
Clinical officer	Health worker has at least 3 years of clinical training	2095	0.69	0.46	0	1
Male	Male health worker	2081	0.47	0.50	0	1
Age	Health worker's age (in years)	2095	40.1	9.99	22	70
Imci_child	Being trained in IMCI and patient is <5 years	2068	0.45	0.50	0	1
Government	Government-owned facility	2095	0.79	0.41	0	1
Drugs	Availability of seven drugs at the day of visit (0–7)	2095	4.47	1.73	1	7
Laboratory	Facility has a laboratory	2095	0.56	0.50	0	1
Child	Patient <5 years	2095	0.65	0.78	0	1
Patient weakness	Weakness of patient, observer's assessment (0 = not weak, 1 = moderately weak, 2 = very weak)	2051	0.31	0.50	0	2
Patient number	Patient's number in the order of observed consultations for health worker h (including both days 1 and 2)	2095	14.36	9.84	1	50

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. Furthermore, we 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. We also 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. Finally, we include a fixed effect specification to account for differences across levels of care (dispensaries, health centres, and hospitals).

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 health worker (*patient number*). 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 order to control for the possibility of a diminishing Hawthorne effect during the observation period, we included the patient number as a control.

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.

5.2. Relationship between caseload and effort

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

$$e_{ijh} = \alpha_1 + \beta_1 w_j + \delta_1(w_j - \hat{w})d_j + z_1\gamma_1 + \varepsilon_1 \qquad (1)$$

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 (cf. Fig. 1), d_j is a dummy that takes a positive value whenever $w_j > \hat{w}$ and z_1 is a vector of controls.

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.

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 then extend the analysis by including the possibility that there are some health facilities with a caseload above the threshold $\hat{w}$. Finally, we use the

---

[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)	0.018 (0.022)	0.010 (0.028)
Clinical officer		1.29** (0.50)	1.29** (0.51)
Male		0.13 (0.41)	0.15 (0.41)
Age		−0.04* (0.02)	−0.04* (0.02)
Imci_child		1.03** (0.45)	1.05** (0.45)
Government		−0.02 (0.45)	−0.01 (0.45)
Drugs		0.08 (0.14)	0.08 (0.14)
Laboratory		0.66 (0.40)	0.62 (0.42)
Child		1.25*** (0.37)	1.25*** (0.37)
Patient weakness		0.83*** (0.25)	0.82*** (0.24)
Patient number		−0.05*** (0.01)	−0.05*** (0.01)
Constant		1.67 (1.52)	1.79 (1.53)
Facility type fixed effect	No	Yes	Yes
Symptom fixed effect	No	Yes	Yes
n	2095	1806	1806
R^2	0.001	0.310	0.309

Sampling weights are used. Estimated standard errors take into account clustering at the facility level and stratification at the district level.

* $p < 0.1$.
** $p < 0.05$.
*** $p < 0.01$.

instrumental variable (IV) approach to control for possible endogeneity of the caseload variable.

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). The R^2 shows 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.

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, being trained as a clinical officer (or above) is associated with a 31% (95% CI: 7, 55) increase in effort per patient, while IMCI training is associated with a 25% (95% CI: 3, 46) increase in effort per patient in the IMCI target group. More effort is also exerted when the patient is a child; 30% (95% CI: 12, 48); or when the patient is very weak (as opposed to not weak); 40% (95% CI: 16, 62). 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 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 10 places down the queue is associated with a 12% 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.

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 0.246 ($p = 0.738$) with a threshold at 44 patients to 0.000 ($p = 0.996$) with a threshold at 13 patients.[9]

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 0.3049 to a maximum of 0.3083 at $\hat{w} = 12$. Hence, the inclusion of a threshold did not 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. Note, however, that the robustness analysis below will show some indication that health workers with more than 40 patients have substituted from more time-consuming tasks (examinations) to less time-consuming tasks (questions), which may suggest that they are (weakly) time constrained.

5.3. 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, the equilibrium values of these two variables are determined in a simultaneous equations model (SEM), where both caseload (w) and effort (e) are endogenous;

$$e_{ijh} = \alpha_1 + \beta_1 w_j + z_1 \gamma_1 + \varepsilon_1 \tag{2}$$

$$w_j = \alpha_2 + \beta_2 e_{ijh} + z_2 \gamma_2 + \varepsilon_2. \tag{3}$$

z_1 and z_2 are vectors of control variables not necessarily equal.[10] The clinician decides the effort level, while caseload is determined by the number of people who need health care as well as by patients

---

[9] Details are available upon request.
[10] We omit the threshold component in Eq. (1) for simplicity.

choice to visit the health facility or not, both of which may be affected by the level of effort. In this case, we cannot estimate Eq. (2) separately, because caseload will be correlated with the error term, thus violating an important assumption for unbiased OLS estimation.

In the absence of exogenous shocks in caseload, instrumental variable estimation (IV) provides one possible solution to the simultaneity problem. A valid instrumental variable (x) should (1) be uncorrelated with the error term in Eq. (2) (i.e., $Cov(x, \varepsilon_1) = 0$), and (2) affect the endogenous variable w (i.e., $Cov(x, w) \neq 0$ in Eq. (3)).

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 well-defined catchment population, varying 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 1000 to more than 400,000 (mean = 9520).

We observe 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. 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 Eq. (2). 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. Another possibility would be that large catchment areas are populated with more educated people, which might affect effort levels. However, the correlation between the caseload variable and the educational level of the patients is not significantly different from zero. A third possibility is that health authorities locate better doctors to places with large catchment populations, but we do not find any correlation between effort and catchment population per clinician in our data. 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 the 2002 Census 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 (only significant coefficients reported). Our instrument is a strong predictor of caseload (p-value < 0.001). The standard deviation of the instrumental variable is 2.79. Hence, a one standard deviation variation in the instrumental variable causes a 42% change in the caseload variable at the mean. We also notice that government facilities have higher caseload than voluntary agencies.

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

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

Variables	Coefficients (standard errors)	
log(catchment population) per health worker	2.91***	(0.22)
Age	0.11*	(0.07)
Government	4.87**	(2.33)
n	1806	
R^2	0.622	

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

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 ($\beta = 0.20$, p-value = 0.636), 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).

6. Robustness analysis

6.1. 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, we found no statistically significant association with caseload at the 5% level for 54 items. Four were negatively associated with caseload and four were positively associated with caseload (see Appendix A). Hence, even at the level of individual items, 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.

6.2. Alternative threshold analysis

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

Table 7
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	R^2	$\hat{\beta}_1 + \hat{\delta}_1$	p	R^2
37	0.03	0.798	0.2797	0.01	0.93	0.2109
38	0.07	0.533	0.2798	−0.02	0.83	0.2109
39	0.13	0.109	0.2803	−0.09	0.17	0.2114
40	0.17*	0.094	0.2803	−0.12	0.15	0.2114
41	0.23*	0.068	0.2803	−0.16	0.11	0.2114
42	0.30**	0.035	0.2803	−0.21*	0.07	0.2114
43	0.45**	0.035	0.2803	−0.31*	0.08	0.2114
44	0.88*	0.050	0.2803	−0.63*	0.08	0.2114

* $p < 0.1$.
** $p < 0.05$.

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 7).

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 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.

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 Eq. (1)). 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 $(\tilde{w} - \hat{w})d$. Again, we were unable to identify any association between caseload and effort per patient.

6.3. 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 hinge 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. The list of included variables 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 any of these variables did not have any impact on our results. Finally, we analysed an alternative model where we distinguished between "anticipated" and "unanticipated" caseload, hypothesising that only unanticipated changes in caseload affects effort. Unanticipated caseload was measured as deviations from the average historical caseload, but this variable was not significantly associated with effort.

7. 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 part of the story health personnel re-count 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 2 h of their workday with patients (18.5 patients, 5.7 min each). Other duties, including administrative work, would not be enough to fill the rest of the workday. 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 (Fig. 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; even the busiest clinics do not appear to have reached the threshold where caseload compromises quality. It is illustrative that in the 5% busiest clinics (>34 patients per worker), time use per patient was identical to the average, and the level of effort was 4.07, compared to the average of 4.04.

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, indicating that some 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%). If part of their contracted working time is being used for these activities, we cannot measure the real slack capacity with our data. However, as shown in Section 3, we would still expect a negative relationship between caseload and effort per patient. Our data thus do not support the hypothesis that health workers are "overworked" due to low effective working time.[11]

In our attempts to isolate the causal effect of caseload on 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 examinations – 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, 2010). Finally, an alternative explanation might be that high effort – in addition to a positive effect on demand due to higher clinical quality – has a negative effect due to longer waiting times, and that the net effect therefore is ambiguous.

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 themselves advance this view (Lindkvist, 2010)? 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 may 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.

Second, the perception of being overworked may relate to the lack of appointment systems. Patients typically queue up during 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*".[12]

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, 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 Paredes et al., 1996; Rowe et al., 2001, 2003).

One limitation of the study is that our pre-defined list of relevant questions and examinations might not be exhaustive and that effort therefore was underestimated in some cases. Furthermore, we have implicitly put the same "effort weight" in each question and examination, which may be a potential source of bias. Future studies should also consider collecting more data on labour supply and alternative job opportunities.

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 assessment quality. These conclusions are robust to numerous alternative specifications and test. 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, they suggest 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

---

[11] 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.

[12] 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.

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.

Acknowledgements

Appendix A.

See Table A1.

Table A1

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* = 3494)			
Welcoming			
Welcome the patient	0.745	0.0040	0.219
Greet the patient	0.632	0.0071*	0.071
All patients presenting with fever (*n* = 1228)			
History taking questions			
Duration of fever	0.841	−0.0036	0.127
Whether temperature has been taken	0.075	0.0011	0.494
Pattern (periodicity) of fever	0.187	0.0037*	0.098
Presence of chills, sweats	0.021	−0.0007*	0.063
Presence of cough, sore throat, pain during swallowing	0.339	0.0009	0.780
Presence of diarrhoea or vomiting	0.484	0.0041	0.212
Presence of convulsions	0.155	0.0017	0.463
Presence of running nose	0.061	0.0019*	0.067
Physical examinations			
Take temperature (with thermometer)	0.383	−0.0026	0.501
Check neck stiffness	0.021	−0.0015**	0.022
Look for palmar pallor (anemia)	0.245	0.0015	0.645
Check ear/throat	0.028	0.0007	0.345
Palpate for the spleen	0.025	−0.0000	0.979
Children < 5 years presenting with fever (*n* = 849). Additional items			
History taking questions			
Ability to drink or breastfeed	0.296	0.0060**	0.048
Difficulty in breathing	0.072	−0.0005	0.713
Presence of ear problems	0.083	−0.0040	0.101
Vaccination history	0.153	0.0008	0.756
Physical examinations			
Check for lethargy or unconsciousness (try to wake up the child)	0.034	−0.0023**	0.040
Check for visible severe wasting	0.040	0.0004	0.717
Look for oedema of both feet	0.017	0.0007	0.509
Check the child's weight (against a growth chart)	0.400	0.0064*	0.068
All patients presenting with cough (*n* = 820)			
History taking questions			
Duration of cough	0.785	0.0014	0.695
Sputum production or dry cough	0.134	0.0025	0.119
Presence of blood in sputum	0.033	0.0004	0.605
Presence of chest pain	0.092	0.0002	0.903
Presence of difficulty in breathing	0.156	0.0011	0.576
Presence of fever	0.461	0.0062	0.170
Physical examinations			
Count respiratory rate	0.111	−0.0065***	0.004
Observe breathing for lower chest wall indrawing	0.130	−0.0009	0.575
Examine throat	0.029	0.0003	0.752
Auscultate the chest	0.200	0.0002	0.958
Take temperature (with thermometer)	0.118	−0.0030	0.233
Children < 5 years presenting with cough (*n* = 526). Additional items			
History taking questions			
Ability to drink or breastfeed	0.245	0.0066*	0.055
Presence of convulsions	0.111	−0.0004	0.868
Presence of ear problems	0.076	−0.0028	0.244
Presence of diarrhoea or vomiting	0.337	0.0057	0.151
Vaccination history	0.166	0.0048*	0.090
Physical examinations			
Check for lethargy or unconsciousness (try to wake up the child)	0.025	−0.0007	0.499
Check for visible severe wasting	0.010	−0.0009*	0.034
Look for palmar pallor	0.111	0.0034	0.202
Look for oedema of both feet	0.008	−0.0001	0.611
Check the child's weight (against a growth chart)	0.255	0.0013	0.748
All patients presenting with diarrhoea (*n* = 381)			
History taking questions			
Duration of diarrhoea	0.760	0.0005	0.884
Frequency of stools	0.526	0.0014	0.698
Consistency of stools	0.237	−0.0084**	0.016
Presence of blood, and/or mucus in stools	0.353	0.0071**	0.044

Table A1 (*Continued*)

	Share of patients exposed	Coeff.	p-Value
Presence of vomiting	0.275	0.0131***	0.000
Presence of fever	0.471	0.0071	0.108
Physical examinations			
Assess general health status (alert or lethargic)	0.093	−0.0046*	0.069
Examine for sunken eyes	0.181	0.0041	0.287
Pinch abdominal skin to assess degree of dehydration	0.197	0.0008	0.835
Take temperature (with thermometer)	0.192	−0.0009	0.863
Children < 5 years presenting with diarrhoea (*n* = 298). Additional items			
History taking questions			
Ability to drink or breastfeed	0.300	0.0063**	0.040
Presence of convulsions	0.083	0.0031	0.178
Presence of ear problems	0.034	−0.0007	0.539
Presence of cough or difficulty in breathing	0.152	0.0001	0.971
Vaccination history	0.147	0.0035	0.354
Physical examinations			
Offer the child a drink of water or observe breastfeeding	0.050	−0.0002	0.881
Check for visible severe wasting	0.012	0.0001	0.823
Look for palmar pallor	0.160	0.0078*	0.089
Look for oedema of both feet	0.008	−0.0001	0.869
Check the child's weight (against a growth chart)	0.371	0.0057	0.210
All patients (*n* = 3494)			
Communication			
Look at the patient while talking	0.846	0.0014	0.399
Tell the patient his or her diagnosis (any name)	0.227	−0.0006	0.762
Explain the diagnosis (in common language)	0.163	0.0006	0.661
Explain the treatment being provided	0.292	0.0015	0.568
Give any health education related to the diagnosis?	0.160	0.0056***	0.001
Explain whether or not to return for further treatment	0.253	0.0039	0.103
Listen properly to the patient/caregiver	0.656	0.0073**	0.034
Allow the patient to talk	0.747	0.0055**	0.010
Ensure the patient had understood diagnosis, etc.	0.132	−0.0005	0.781

* $p < 0.1$.
** $p < 0.05$.
*** $p < 0.01$.

References

Amin, S., Das, J., Glodstein, M. (Eds.), 2008. Are You Being Served? New Tools for Measuring Service Delivery. The World Bank, Washington, DC.

Barber, S.L., Gertler, P.J., Harimurti, P., 2007. The contribution of human resources for health to the quality of health care in Indonesia. Health Affairs 26, w367–w379.

Black, B.E., Morris, S.S., Bryce, J., 2003. Where and why are 10 million children dying every year? The Lancet 361 (9376), 2226–2234.

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

Census, 2002. Tanzania 2002 Population and Housing Census. National Bureau of Statistics, Dar es Salaam.

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

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

El Arifeen, S., Blum, L.S., Hoque, D.M.E., Chowdury, E.K., Khan, R., Black, R.E., et al., 2004. Integrated Management of Childhood Illness (IMCI) in Bangladesh: early findings from a cluster-randomized study. The Lancet 364, 1595–1602.

Gouws, E., Bryce, J., Habicht, J.P., Amaral, J., Pariyo, G., Schellenberg, J.A., 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. Bulletin of the World Health Organization 82, 509–515.

Leonard, K., Masatu, M.C., Vialou, A., 2007. Getting doctors to do their best. The Journal of Human Resources 42, 682–700.

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

Leonard, K., Masatu, M.C., 2006. Outpatient process quality evaluation and the Hawthorne Effect. Social Science & Medicine 63, 2330–2340.

Lindkvist, I., 2010. Overworked or Under-motivated? Results from a Qualitative Study of Health Worker Performance and Clinical Quality in Tanzania. Mimeo.

Madden, J.M., Quick, J.D., Ross-Degnan, D., Kafle, K.K., 1997. Undercover careseekers: simulated clients in the study of health provider behavior in developing countries. Social Science & Medicine 45, 1465–1482.

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

Mwisongo, A., Mæstad, O., 2010. Missing the Essential? The Practice of Rural Clinicians. Mimeo.

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

Naimoli, J.F., Rowe, A.K., Lyaghfouri, A., Larbi, R., Lamrani, L.A., 2006. Effect of the Integrated Management of Childhood Illness strategy on health care quality in Morocco. International Journal for Quality in Health Care 18, 134–144.

Ofori-Adjei, D., Arhinful, D.K., 1996. Effect of training on the clinical management of malaria by medical assistants in Ghana. Social Science & Medicine 42, 1169–1176.

Osterholt, D.M., Rowe, A.K., Hamel, M.J., Flanders, W.D., Mkandala, C., Chizani, N., 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, 1147–1156.

Paredes, P., Pena, M., Flores-Guerra, E., Diaz, J., Trostle, J., 1996. Factors influencing physicians' prescribing behavior in the treatment of childhood diarrhoea: knowledge may not be the clue. Social Science & Medicine 42, 1141–1153.

Rowe, A.K., Hamel, M.J., Flanders, W.D., Doutizanga, R., Ndoya, J., Deming, M.S., 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, A.K., Onikpo, F., Lama, M., Cokou, F., Deming, M.S., 2001. Management of childhood illness at health facilities in Benin: problems and their causes. American Journal of Public Health 91, 1625–1635.

Rowe, A.K., Lama, M., Onikpo, F., Deming, M.S., 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, A.K., de Savigny, D., Lanata, C.F., Victora, C.G., 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., 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, 96–196.

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. National Bureau of Statistics and Macro International. MIT Press, Dar es Salaam.

Wedig, G., Mitchell, J.B., 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. The World Bank, Washington, DC.

WHO, 2006. Working Together for Health. World Health Report 2006. World Health Organization, Geneva.

Wooldridge, J.M., 2002. Econometric Analysis of Cross Section and Panel Data. MIT Press, Boston, MA.

Zurovac, D., Rowe, A.K., Ochola, S.A., Noor, A.M., Midia, B., English, M., Snow, R.W., 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.