European Bank
for Reconstruction and Development

Generosity norms and intrinsic motivation in health care provision: evidence from the laboratory and the field*

J. Michelle Brock
Andreas Lange
Kenneth L. Leonard

Abstract

We examine the correlation between the generosity of clinicians – as measured in a laboratory experiment – and the quality of care in their normal practices under three different intrinsic incentive schemes. Specifically, we observe clinicians in their normal work environment, when a peer observes them and six weeks after an encouragement visit from a peer. Clinicians who give at least half of their endowment to a stranger in the laboratory (generous) provide 10 per cent better quality care than those who do not. In addition, the average clinician provides about 4 per cent better quality when observed by a peer and 10 per cent higher quality care after the encouragement visit. Importantly, we find that generous clinicians react to peer scrutiny and encouragement in the same way as non-generous clinicians. Many clinicians are intrinsically motivated to provide higher quality care. However, most clinicians respond to increased intrinsic incentives in the form of scrutiny and encouragement from peers.

Keywords: intrinsic incentives, health care quality, altruism, professionalism, Tanzania, experimental economics, Hawthorne effect, Encouragement effect, Study effect

JEL Classification Number: I15, O19, C91, C93, J2

Contact details: J. Michelle Brock, One Exchange Square, London EC2A 2JN, UK

Phone: +44 20 7338 7193; Fax: +44 20 7338 6111; email: brockm@ebrd.com.

J. Michelle Brock is a Research Economist at the European Bank for Reconstruction and Development, London, UK. Andreas Lange is a Professor of Economics at the University of Hamburg, Department of Economics, Von Melle Park 5, 20146 Hamburg, Germany. Kenneth L. Leonard is an Associate Professor at 2200 Symons Hall, University of Maryland College Park, MD 20742

*This work was funded by a Maryland Agricultural Extension Station seed grant, a contract from the Human Resources for Health Group of the World Bank, in part funded by the government of Norway and the Eunice Kennedy Shriver National Center for Child Health and Human Development grant R24-HD041041 through the Maryland Population Research Center and support from the Knowledge Product Human Resources for Health group at the World Bank. We are grateful for the support of the Center for Educational Health, Arusha (CEDHA), specifically Dr Melkiory Masatu and Dr Beatus Leon. We thank participants at Michigan State, University of Minnesota and the Economic Science Association for their comments.

Working Paper No. 147 **Prepared in August 2012**

1. Introduction

Health care workers are commonly described as being intrinsically motivated; the literature on health care is full of references to terms such as professionalism, esteem and caring. At the same time, all health systems invest significant resources in regulation and quality assurance, thereby declining to leave quality up to the caring instincts of providers. Furthermore, where the regulatory capacity of governments is weak, quality is also often low and the effort levels of health workers in such settings is coming under increasing scrutiny (see Das et al., 2008, for a review). Particular attention is paid to the "know-do gap" – the gap between what health workers know how to do and what they actually do for their patients (Das and Hammer, 2007; Leonard and Masatu, 2010b; Maestad and Torsvik, 2008; Maestad et al., 2010; Rowe et al., 2005; WHO, 2005). Thus, the evidence suggests that most health workers are not sufficiently intrinsically motivated to provide the best care for their patients. The failure of "caring instincts" to provide high quality care has led to increased focus on extrinsic motivation and there is evidence that paying directly for increased output (Basinga et al., 2011; Meessen et al., 2006), better incentives within organisations (Das and Hammer, 2007; Leonard et al., 2007) and pressure from the community (Björkman and Svensson, 2009) can improve outputs and quality.

This paper takes a step back from extrinsic incentives to re-examine intrinsic motivation. We suggest that intrinsic motivation can be an important driver of quality care, but that the myth of the ubiquitously intrinsically inspired health worker with unlimited and context-free energy is an outdated and damaging view. In particular, the literature from experimental and behavioural economics suggests that most pro-social behaviour is context dependent. We partition the relationship between intrinsic motivation and effort by examining performance across types of health workers (in terms of altruism) as well as the performance of all health workers across working environments. In other words, we ask, first, if there are health workers who are more intrinsically motivated to provide high quality care and second, if changes in the work environment can provide intrinsic incentives for the average health worker to increase their effort.

To test the relative importance of types and environment we examine the behaviour of health workers who provide outpatient care (clinicians) in the urban and peri-urban Arusha region of Tanzania. We look at four different settings, each with different implied intrinsic returns to effort. First, we examine the performance of these same clinicians under normal situations in their workplace. Second, we measure their performance when there is a peer present to observe their activities. Third, we measure their effort after participation in a trial in which a Tanzanian medical doctor (MD) read an encouragement script and asked them to improve their performance on five specific items. Finally, we examine clinicians in an economic laboratory experiment, measuring their generosity to strangers. We use the laboratory experiment to identify clinicians who are generous to strangers and compare their performance in their normal work situation to clinicians who are non-generous in the laboratory experiment. In addition, we can measure the response of all clinicians to the increasing intrinsic incentives implied by peer scrutiny and encouragement. And finally, we can compare the differential response of generous and non-generous clinicians to the changes in intrinsic motivation implied by scrutiny and encouragement.

We find that clinicians who are generous in the laboratory perform better at work, as measured by adherence to protocol items required by the patients' symptoms. As such,

generosity and intrinsic motivation appear to be linked across different environments. That is, we find evidence that measuring pro-social preferences in the laboratory allows the classification of particular types of health workers who differ in their performance towards their patients. Also, in the field, the average clinician increased his or her effort significantly both when subjected to peer scrutiny and when encouraged to provide better care. In the latter case, the improvements are large and significant even eight weeks after clinicians received an encouragement visit. We emphasise that it is not just the generous doctors who respond to this other source of intrinsic motivation. The performance increases under scrutiny or encouragement are similar for generous and non-generous clinicians alike. This suggests two things: (i) those who are motivated by generosity are also motivated by peer scrutiny and encouragement; (ii) even those who are not categorised as "generous" or altruistic in the laboratory respond to intrinsic incentives in the field. This result thus supports the idea that social preferences (intrinsic motivation?) are multidimensional. Clearly, the behaviour in the laboratory reflects an underlying trait that is different between generous and non-generous doctors. This is not something that would have been identifiable in the field. But this characteristic is different from and can be complemented by responsiveness to other sources of intrinsic motivation, namely peer influence.

The changes in the quality of care observed in this investigation and intervention are relatively large. Generous clinicians are half a standard deviation better than non-generous clinicians. Encouragement also improves the average performance by half a standard deviation. These differences are about three-quarters of the difference between effective and ineffective organisations in a similar setting (Leonard et al., 2007) and significantly larger than the 0.14 standard deviation gain observed in the successful pay-for-performance scheme in Rwanda (Basinga et al., 2011).

In the following section, we outline the view of intrinsic motivation from the management and experimental economics literatures and present our hypotheses about behaviour in this setting. Section 2 outlines the data and empirical methodology for examining the data. Section 3 shows the results and section 4 discusses the implications and provides our conclusions.

2. Motivation and effort

Specifically for the health sector, Franco et al. (2002) define motivation as the "willingness to exert and maintain an effort towards organizational goals". For the sake of this document, we will assume that organisations participating in the health care market in Tanzania have the same goals as the profession and want to provide high quality care to their patients. Therefore maintaining organisational goals is the same as providing high quality and efficient services. Franco et al. discuss two sources of motivation: (i) unique personal factors that naturally align health worker incentives and organisational goals and (ii) the environmental factors that shape the health worker's personal goals so they are aligned with organisational goals. In this document, we focus on the intrinsic factors inherent in both of these sources of motivation.

In the management literature, extrinsic incentives refer to the possible enjoyment of rewards earned from doing the job and intrinsic incentives refer to the enjoyment of the job itself (Herzberg, 1959, 1987). Thus, extrinsic incentives would include not only salary and benefits, but also the possibility of promotion and praise or esteem earned for a job well done. Intrinsic incentives would include the opportunity to work with interesting people, learn new things on the job and the sense of accomplishment from a job well done.

The discussion of intrinsic motivation in the health care setting often focuses on Franco et al.'s first source of motivation: personal factors. Thus, the intrinsically motivated health worker has a utility function such that they automatically conform to organisational goals. Such a person can be described as altruistic and will do the best they can for the welfare of their patients and therefore conform to organisational goals to the degree that organisations also care about the welfare of their patients. For example, Serra et al. (2011) measure altruism in medical students in Ethiopia and find that altruistic students are more likely to choose to work in non-profit (charity) settings and earn lower wages. Thus, intrinsic motivation displayed before entering the workforce affects behaviour after entering the workforce, in this case, through the choice of workplace.

In reference to Franco et al.'s second source of motivation (environmental factors), an alternative view of intrinsic motivation is that all (or at least most) people respond to intrinsic incentives. We can think about intrinsic incentives as things that are non-monetary but impact people's views of themselves. This could begin with the perspective that most people do in fact care about other people, particularly the people over whom they have some measure of control. However, this "caring" is not unconditional. They care primarily about themselves and only about others to the degree that it increases the utility they obtain from their own activities. Thus, a health worker may care about his or her patients when he or she works in an environment that acknowledges his or her service and simultaneously provides them with a comfortable living.

Although caring about patients is unique to the health care sector, it is not the only source of intrinsic motivation. In particular, the self-satisfaction that comes from the recognition and appreciation of professional peers is also important.

2.1. Motivation and pro-social behaviour

The field of behavioural economics and psychology has directly investigated the behaviour of individuals in settings where their actions have implications on the welfare of others. Economists evoke the term social preferences to explain decision-making that goes beyond

self-interest. Charness and Rabin (2002) characterise a person with social preferences as "not maximizing own monetary payoffs when those actions affect others' payoffs", and consider people to be both self-interested and concerned with the payoff of others. Pro-social behaviour is the case where individuals voluntarily engage in activities that are costly to themselves but benefit others (Benabou and Tirole, 2003). Becker (1974), Sugden (1982) and Andreoni (1989, 1990) discuss social preferences in terms of contributions to a public good. According to Andreoni, individuals may contribute to a public good both because they benefit from the public good and because they get some private benefit from having given, which Andreoni refers to as a "warm glow." Rabin (1993) extends this view of giving and argues that individuals have conditional interest for the welfare of others; they are kind to people they believe are also kind. Feeling good about having done something nice for someone else will also depend on the identity of the recipient or how the recipient reacts to the gift. Benabou and Tirole (2003) enrich this approach by bringing in concerns for social reputation and self-respect, concentrating on happiness derived from others' perceptions. Importantly, the "other" can be the recipient of generosity or a witness to the act of generosity. Thus, people may be kind or generous to others because they like seeing the recipient receive something, they like the fact that the recipient knows they gave them something or they like being seen as having given to the recipient. In addition, giving may be conditional on the identity of the recipient or the individual observing the giving (Ellingsen and Johannesson, 2008); individuals may like being seen by specific people as being generous.

In the health care context, therefore, we can think of two potential sources of pro-social behaviour; patient- and peer-oriented motivation. If health workers care primarily about the welfare of their patients or if they care to be noticed by their patients, they will seek to do the best thing for their patients or to provide the care that patients seek, respectively. On the other hand, peer-oriented pro-social behaviour is a type of professionalism, in which health workers gain utility from following and being seen to follow the norms of their peer group.

Professionalism is common in settings where service to a greater good is advanced by dedication to group goals and values (Akerlof and Kranton, 2000, 2005; Cullen, 1978; Freidson, 2001). Leonard and Masatu (2010a) describe a form of latent professionalism in which individuals follow professional norms only when they believe their fellow professionals can observe or evaluate their behaviour. Whereas health workers who care about their patients always have patients present, these esteem-seeking health workers may or may not have the opportunity to garner the esteem of their peers, depending on their work environment. Thus, these two sources of intrinsic motivation define both a type of health worker who should always provide high quality care (patient-oriented health workers) as well as environmental conditions that would increase the intrinsic incentives to provide high quality health care (the presence or scrutiny of peers). Importantly, the first type of health worker should be less responsive to changes in the environment, whereas the second type is likely to alter their behaviour significantly when the environment changes.

2.2. A Model of effort with extrinsic and intrinsic incentives

Clinicians provide effort (a) for many reasons some of which may be described as social preferences. In order to illustrate the different motivational factors, we provide a descriptive model that uses a separable utility function, as in Levitt and List (2007). Motivations include monetary (wealth, W), altruistic (moral, M), and esteem-seeking (reflective, R) motivations. The level of effort is a function of three sources of stimuli, one for each source of utility: s_w ,

s_m and s_r . s_w can be thought of as the wage, s_m as the individual's level of social obligation to others and s_r as exposure to peers. Utility is defined as:

$$U(a, S) = U_w(a, s_w) + U_m(a, s_m) + U_r(a, s_r) - c(a) \qquad (1)$$

The health worker will choose effort to maximise utility such that:

$$\frac{\partial U_w}{\partial a} + \frac{\partial U_m}{\partial a} + \frac{\partial U_r}{\partial a} = \frac{\partial C}{\partial a}$$

It is reasonable to assume that increasing stimuli from any source (indexed i) increases effort ($\partial a^*/\partial s_i > 0$) and that these increases are smaller when stimuli are greater ($\partial^2 a^*/\partial s^2_i < 0$). We may further assume that when one form of stimuli is high – all else equal – the gain in effort from increasing another form of stimuli will be lower ($\partial^2 a^*/\partial s_i \, \partial s_{-i} < 0$), that is, the motivational sources are substitutes.

Therefore, when an individual has high incentives to provide effort (for example a high wage or high levels of social obligation) then increases in other stimuli (exposure to peers, for example) should not lead to large increases in effort. On the other hand, when a clinician faces low levels of stimuli overall, increases in any form of stimuli may be hypothesised to lead to large increases in effort.

In our empirical setting, we can observe effort (a^*) but we cannot observe the current levels of wealth, moral or reflective stimuli. However, we can observe the changes in effort due to exogenously increased levels of reflective stimuli ($\partial a^*/\partial s_r$) as a result of increased exposure to peer scrutiny.

In addition, we use behaviour in a lab experiment to define a set of clinicians who provide effort due to altruistic reasons; that is, have a higher moral stimulus (G: generous) than other clinicians (NG: non-generous). That is, we predict that the relative altruistic moral stimulus for agents is stable across different contexts and the stimuli they face in the field is also higher: $s_m^G > S_m^{NG}$

In our experiment, we test the following relationships:

Hypothesis 1: Clinicians with higher moral stimulus as measured in the lab ($s^G > s^{NG}$) will provide higher levels of effort in the field: $a^{*G} > a^{*NG}$.

Hypothesis 2: The average clinician will increase his effort when faced with increased reflective stimuli (exposure to scrutiny): $\partial a^* > 0$.

Hypothesis 3: Clinicians who provide relatively low levels of effort ($a^{*1} < a^{*2}$), will exhibit greater responses to additional scrutiny: $\partial a^{*1}/\partial s_r > \partial a^{*2}/\partial s_r$.

Hypothesis 4: By combining hypotheses 1 and 3, clinicians with high levels of moral stimuli (G) will respond less to increases in reflective stimuli than clinicians with low levels of moral stimuli (NG): $\partial a^{*G} < \partial a^{*NG}$

3. Methodology

We studied 103 clinicians who practice health care in the Arusha region by collecting data on the quality of care in the course of their normal practices. Sixty-three of these clinicians also participated in a laboratory experiment and this analysis focuses on these workers.

3.1. The laboratory experiment

The laboratory experiment took place in Arusha, Tanzania in July 2010. The subject pool consisted of 71 clinicians[1] and 78 non-clinician subjects. We ran one session each day for two days. We recruited non- clinicians subjects with printed advertisements distributed in major market areas in Arusha. While flyers were distributed to a variety of people, the group of non-clinician subjects was ultimately a convenience sample. All of the non-clinician subjects that arrived to participate each day were allowed into the experiment. Clinician subjects were given a *per diem* of TZS 35,000 in addition to what they earned in the experiment. Non-clinician subjects received a show-up fee of TZS 5,000. One US dollar is equal to approximately to 1,300 Tanzanian shillings.[2]

Clinician subjects gathered in a classroom and non-clinician subjects gathered on a lawn outside of the classroom, near enough that both groups could see each other but far enough that there was no communication or individual identification. This was done to preserve anonymity while ensuring that subjects understood the concept of being paired with another player. Subjects recorded decisions using paper and pen. We provided a hard copy of the experimental instructions to each participant and read them aloud before the experiment began. The instructions explained the basic guidelines of the experiment and how earnings were determined. Subjects were given the chance to ask clarifying questions after the instructions were read.

The experiment was a standard dictator game in which the dictator decides how to allocate money between him or herself and an anonymous partner. The dictator in each pair was always a clinician and the receiver was always someone drawn from the non-clinician pool. (We never used the terms doctor or patient in the experiment, but the clinicians knew they were in a group of clinicians). The only task for the clinician–dictator was to allocate 100 tokens between him- or herself and the receiver. True to the term "dictator", the receiver had no choice but to accept what was given. Each token was worth TZS 150, so that the clinician was choosing the allocation of TZS 15,000 (approximately US$ 12).

3.2. The field

We collected data on clinician performance for 103 clinicians and 4,512 patients in the semi-urban area of Arusha, Tanzania. The field data collection ran from November 2008 until August 2010.[3] The sample includes public, private and non-profit/charitable facilities. Clinicians were randomly sampled within each facility. The term "clinician" refers to primary

[1] Some of the clinicians in the laboratory experiment did not participate in the field study.
[2] The imbalance in the show-up fees was never highlighted to participants but could have been inferred. It does parallel the power and income imbalance in a typical clinical encounter.
[3] We sampled 100 per cent of the health care facilities in the area with outpatient departments, although some facilities were excluded based on convenience; they were either too difficult to reach for obtaining consent or had too small of a patient volume.

health workers who provide outpatient care. They fill the role of "doctor", although the majority of them do not have full medical degrees.[4]

The experiment was conducted as follows. First, each clinician was enrolled in the study by a member of the team.[5] Then we collected data on the quality of care in two different environments, which often took place on the same day. In the first setting, the baseline, the clinician knows he has been enrolled in a study (usually many days before) but does not know that we are collecting data because we interview patients he has already consulted. The second setting is what we call the peer scrutiny visit. For this visit, a clinician on the research team entered the consultation room and observed the clinician working. This visit was designed to increase peer scrutiny and invoke a Hawthorne effect and increased levels of effort (Leonard and Masatu, 2006).

Following this second data collection visit, the clinician was visited by a member of the research team in what we call the "encouragement visit," outlined below. The original idea was that this encouragement visit would invoke a smaller but more durable increase in effort than that seen in the peer scrutiny visit. The encouragement always occurred at least one day after the peer scrutiny. There were two more data visits and a follow up visit from a member of the research team after the encouragement visit.[6] Finally after all of these activities were completed and clinicians had reason to believe the study was finished, there were two additional data collection visits. We use these last two data collection visits to measure the medium-term response to peer encouragement.

Thus, we examine effort in the baseline, effort under peer scrutiny and effort after the study was completed. These environments are labelled the baseline, scrutiny and encouragement. The gap between enrolment and the final data collection was about six and half weeks on average.

On each day of data collection we interviewed all the patients seen in the four-hour window during which we visited the facility. The interviews with patients followed the Retrospective Consultation Review (RCR) instrument. It is a slightly modified version of the instrument used by Leonard and Masatu (2006) and it measures clinician effort. Immediately after their consultation, patients are asked a series of questions about their consultation based on the symptoms that they reported. The questions allow us to reconstruct the activities of the clinician, specifically the extent to which they followed protocol. Even though the interview took place within minutes of the consultation, patient recall is not perfect. It is, however,

[4] The four cadres of clinicians include assistant clinical officer (ACO), clinical officer (CO), assistant medical officer (AMO) and medical officer (MO). Each of these titles requires a specific degree. The medical training required for each depends on the degrees an individual already has. Typically, with no other degrees and four years of secondary school, it requires three years of training to become a CO. ACOs have less training. AMOs have on average 3.5 years of medical schooling. MOs have the equivalent of a United States Medical Doctor (MD) degree. None of the MOs in our sample participated in the laboratory experiments, so they are not featured in this paper.

[5] No one declined participation, and attrition was a minor problem for the individual doctors involved in the study. Only two clinicians that had originally consented opted out later. There was additional attrition as a result of clinicians taking their annual leave or attending compulsory continuing education seminars. Whenever possible, we maintained contact with these clinicians and continued to collect data when they returned to work. None of these types of attrition are correlated with observable clinician characteristics or quality of care.

[6] In between the encouragement visit and the post study visits, clinicians were randomised into four treatments in which they received gifts, prizes and follow-up visits at different times. These treatments are ignored in the current study and we examine only the long-run impact of having been encouraged and studied.

highly correlated with what actually took place (Leonard and Masatu, 2006). The questions used to establish protocol adherence are listed in Table 9 in the appendix.

3.2.1. The encouragement intervention

For the encouragement intervention, Dr. Beatus Leon, a Tanzanian M.D. and lecturer at a health research institution, visited each clinician and read the following script.

"We appreciate your participation on this research study. The work that you do as a doctor is important. Quality health care makes a difference in the lives of many people. Dedicated, hard working doctors can help us all achieve a better life for ourselves and our families.

One important guideline for providing quality care is the national protocol for specific presenting symptoms. While following this guideline is not the only way to provide quality, we have observed that better doctors follow these guidelines more carefully. Some of the protocol items that we have noticed to be particularly important are telling the patient their diagnosis, explaining the diagnosis in plain language and explaining whether the patient needs to return for further treatment. In addition, it is important to determine if the patient has received treatment elsewhere or taken any medication before seeing you, and to check the patient's temperature, and check their ears and/or throat when indicated by the symptom."

For this research, we look at clinician adherence to these specific protocol items. We chose specific items because our previous work shows that the best clinicians frequently perform these activities but most clinicians do not. Mentioning these five items also allows us to compare the performance on these items to performance on items not mentioned.

3.3. Empirical specification

An outpatient consultation with a clinician involves a series of discrete interactions, most of which are required by protocol. The RCR instrument is designed to measure whether the clinician did the clinical tasks he is required to do by asking patients if the clinician did those items (as soon after the consultation as possible). These items can involve greeting the patient and offering him or her a chair, asking the patient how long they have been suffering from particular symptoms, asking about additional symptoms, examining the patient and explaining the diagnosis properly. The list of discrete items required by protocol differs somewhat according to the presenting symptoms of the patient.

We have compiled lists of items required by protocol for four categories of presenting symptoms (fever, cough, diarrhoea and general) and two types of patients (older than or younger than five years). Overall, there are 74 different items (listed in subsection A.3), but only a subset will apply to any given patient. During the RCR interview, patients are only asked about items that apply to their symptoms and age category. Thus, we have data on a series of items for each patient and information on the age and gender of the patient and caregiver, where applicable.

The issue, of course, with measuring quality, is that not all items are equally important, not all clinicians are equally qualified to do each item, and the patients who are at one facility might be unobservably different from the patients at another facility. Thus, comparisons across doctors are difficult. We address this problem in four ways. First, wherever possible, we include clinician fixed effects, allowing us to compare each individual clinician to him or herself in different situations (baseline compared with peer scrutiny for example). This also

deals with the case mix and qualifications problem by avoiding the comparison of one type of doctor to another.

Second, we include dummy variables for each specific item, essentially asking if a clinician is more or less likely than the average clinician to provide a given item. The clinician who asks about the duration of a cough 80 per cent of the time is providing below average quality, whereas the clinician who asks about the history of vaccinations in infants 80 per cent of the time is providing above average quality. This helps to control for case mix by adjusting expectations for each type of patient; otherwise, a clinician who sees many infants will look worse than a clinician who sees fewer infants because his average score may be lower.

Third, because we observe a series of outcomes for each patient (corresponding to all of the required items) we can cluster the standard errors at the patient level or include a patient-level random effect.[7] This allows us to control for the fact that some patients may be quite different from others (they are more demanding or critically sick, for example), the distribution of these patients across clinicians may not be even and the probability of performing one item is likely to be correlated with the probability of doing another for the same patient.

Finally, in addition to examining the probability that a clinician would perform any individual required item, we examine the results looking at average adherence to protocol for each patient, reducing the number of observations to the total number of patients (not potential items).

Thus, rather than dealing explicitly with each of the potential issues in one specification, we include four specifications that deal with the expected problems in different ways. The first specification is a logit model of whether the doctor performed each required item with item-specific dummy variables. Since the standard errors are not corrected or adjusted, this specification always has smaller standard errors than the other specifications. The second specification is a logit model of whether the doctor performed each required item with item-specific dummy variables and patient random effects. The patient random effect captures the possibility that an unobservable patient characteristic might simultaneously increase (or decrease) the probability that a clinician did all of the required items.[8] The third specification is a linear regression of the discrete variable of whether the doctor performed each required time with item-specific fixed effects and standard errors clustered at the patient level. The fourth specification is a linear regression of the proportion of required items performed for each patient. Each model also controls for patient characteristics including age and gender. The patient-level regression also controls for the major symptoms reported, controls that are already embedded in the item-specific dummy variables for the other three regressions.

When comparing the quality of care in the baseline, we cannot control for individual doctor effects, but rather control for whether the doctor is generous. When comparing quality under scrutiny and encouragement to the baseline quality, we can include doctor level effects. For the first three models we can differentiate between items that were mentioned in the encouragement script and items that were not mentioned, since the analysis is at the patient-item level, as opposed to the item level as in the fourth specification.

[7] It is not possible to include patient level fixed effects, because each patient is only seen once in our data.

[8] We expect this affect to be uncorrelated with observable characteristics of the patient (age and gender) and therefore include these characteristics independently as dummy variables.

4. Results

4.1. Laboratory experiment

Table 1 presents a summary of giving in the dictator game. The average number of tokens given was just over one-third, but the mode in the data was half. Of the participants in the laboratory experiment 36.8 per cent gave at least half of their tokens to the stranger. Most of these gave exactly half but a few gave more. The fact that the mode was half suggests a fairness norm in which people simply divide their allocation. As a result, we create a dichotomous variable indicating the clinicians who gave at least half and call these clinicians generous.[9]

4.2. Effort in the field

Table 2 shows the basic statistics for the 63 clinicians who were involved in both the experiment and the field study. Of these 63 clinicians observed in the baseline, 59 were observed under peer scrutiny and 51 under the encouragement intervention. Doctors dropped out of the study for various reasons but attrition was not correlated with quality. The average clinician completed 74 per cent of the required items in the baseline and the standard deviation of average doctor quality was 16 percentage points.

4.2.1. Are generous clinicians different from other clinicians?

Table 3 shows that by all four ways of examining quality, generous clinicians provide significantly higher quality than non-generous clinicians. Laboratory behaviour is therefore informative of the relative performance of doctors in the field. This confirms hypothesis 1 that clinicians with greater moral stimulus as measured in the laboratory will provide greater effort under normal circumstances. The impact is between 7 and 9 percentage points, about half a standard deviation of quality.

4.2.2. Reactions to reflective stimuli (scrutiny and encouragement)

Table 4 examines hypothesis 2 for all of the clinicians who took part in the laboratory experiments (not differentiated by generosity). Unlike Table 3, each clinician is compared with his- or herself by the inclusion of clinician fixed effects. The average increase in quality due to scrutiny is between three and four percentage points, depending on the type of regression. The reaction to encouragement is almost twice the size on average, about 5 to 6 percentage points. We examine the additional impact of the encouragement on items that were specifically mentioned in the encouragement script—tracked items—and find that clinicians responded to these items an additional 4 to 7 percentage points. The overall reaction to encouragement of about 8 percentage points, seen in Column 4, represents an increase in quality that is about half of a standard deviation of quality. Thus, Table 4 confirms our hypothesis that the average clinician will respond to scrutiny and encouragement.

[9] The trends we observe below are not seen in any analysis of the continuous amounts given; those who give more than half are not higher quality clinicians than those who give exactly half, and those who give more than zero but less than 50 tokens are not higher quality clinicians than those who give zero.

Table 5 examines hypothesis 3, that clinicians who face low levels of motivation under normal circumstances also exhibit greater changes in effort when faced with additional scrutiny; the hypothesis inherent in our descriptive model of motivation. Table 5 includes a measure of the baseline performance for each clinician, transformed into a "performance gap" which is: (1 - average proportion of items completed in the baseline).

Thus, a clinician with high levels of effort in the baseline has a small gap and a clinician with low levels of effort in the baseline has a large gap. We examine the degree to which the gap explains the change in performance when a clinician is subject to peer scrutiny or encouragement. A coefficient of one suggests that the gap is fully closed, and a coefficient of zero suggests that the reaction to scrutiny is not correlated with the gap.

The coefficients (across all four regressions) suggest that the performance gap closes by about one-quarter under scrutiny and almost one-half under encouragement. The coefficient on the performance gap is significantly different from zero (and one) and implies that the reactions to scrutiny and encouragement differ across clinicians, with the better clinicians exhibiting a smaller reaction. The coefficients suggest that, after encouragement, a clinician with 75 per cent adherence in the baseline (a 25 per cent performance gap) will increase his or her effort by about 2 percentage points (.25*.3 - 0.055), whereas a clinician with only 50 per cent adherence will increase effort by 9.5 percentage points.[10]

4.3. Reactions to reflective stimuli for generous clinicians

Table 6 examines the impact of the two interventions – scrutiny and encouragement – and the differential response of generous and non-generous clinicians. In a second test of the impact of generosity, we examine hypothesis 4 and test whether the response to scrutiny or encouragement is similar for generous and non-generous clinicians. As with Table 5, these regressions include dummy variables for each clinician and examine changes in quality, not the level of quality.

The increase in quality due to the impact of peer scrutiny and encouragement for the average clinician is almost exactly the same as we found in Table 4. Since generous clinicians provide higher levels of quality overall, we expect that they will respond less to the additional stimuli inherent in peer scrutiny and encouragement. When we examine the marginal coefficients for generous clinicians we see no differential pattern for generous clinicians versus average clinicians for either the scrutiny effect or the encouragement effect. Thus, generous clinicians increase their effort at the same rate as non-generous clinicians when faced with additional scrutiny and encouragement and we find no evidence for hypothesis 4. Note that, not only are the coefficients not significantly different from zero, they are also small, but the data clearly indicate the hypothesis is false.

[10] A clinician with almost perfect adherence actually decreases effort. This result is driven by the asymmetry of measurement error in quality at the high end – it is difficult to overestimate quality when the baseline is 98 per cent but easy to underestimate it.

5. Discussion and conclusion

This paper examines two different ways of thinking about intrinsic motivation in the health sector: intrinsically motivated types of health workers and intrinsic incentives for all types of health workers. We isolate a type of health worker who is generous to strangers in a laboratory setting to proxy for altruistic types of health care workers. In addition, we measure the degree to which all health workers respond to intrinsic incentives in the field, with two exogenously assigned experimental interventions that change the exposure of health workers to their peers in the everyday workplace.

About a third of all health workers qualify as generous types; they conform to a generosity norm by sharing an allocation fairly between themselves and an anonymous partner. This is a higher percentage than usually found with the dictator game in other populations. It may be driven by observable income differences between clinicians and recipients who were recruited from the general population. However, it may also reflect the true distribution of characteristics among health workers. Importantly, those health workers who are generous and who conform to the fairness norm in the lab are better clinicians in their normal practices. The difference is large; almost half of a standard deviation in the distribution of quality. Our interpretation of this result is that both generosity in the laboratory and effort with actual patients are increased by the underlying preferences of individuals. To our knowledge, this is one of the few studies in any field that has demonstrated a strong link between behaviour in a laboratory and behaviour in the field.

The fact that generosity improves performance is both good and bad news for the health sector in countries with ineffective regulation. First, it suggests that some health workers will provide better care, even in difficult situations. However, it also suggests that if a health worker is not generous, quality will be low. Quality should not be up to the generosity of clinicians, it should be guaranteed by employers or regulators.

The good news, from a policy perspective, is that even non-generous clinicians respond to some types of intrinsic incentives. In this case, we look at the power of peer influences. The average clinician in our sample increased the quality of care they provided when observed by a peer and when encouraged and studied over a long period. Our original understanding of peer effects was that clinicians would respond more to the presence of a peer than to the encouragement of a peer. This turned out to be wrong. Even though the doctor doing the encouragement was never present in a consultation, clinicians worked harder after he visited them. Part of this finding reflects the fact that the encouragement effect combined both the encouragement visit and the repeated contact with clinicians due to on-going data collection.

We know, however, that the increase in quality is not something that only occurs when the research team is present at the facility because many of the patients we interviewed were consulted before our research team arrived at the facility (see subsection A.1 in the Appendix). On the other hand, some clinicians would find out that the research team had been present after the visit and this information likely played an important role in the effort they decided to exert. Encouragement combined with repeated visits over a long period is something clearly different from normal for these clinicians.

The impact of encouragement is large even six weeks after the encouragement. This study was designed to detect a fall in quality after the final official visit; we collected data for at least two weeks after that. However, not only is quality still higher at this point, but there is

no evidence of any decline. In fact, the data suggest that quality continued to increase over the entire duration of the study. Table 8 in the Appendix suggests that quality will continue to rise after we finished collecting data. This is hard to believe, but the evidence clearly suggests no fall in quality in the short to medium term.

By asking clinicians to work harder and by mentioning five clinical tasks, we were able to increase quality for the average clinician by at least half a standard deviation. Recall that this difference is only a little smaller than the quality differences between effective and ineffective organisations and about three times the size of the difference recorded in the pay-for-performance scheme in Rwanda.[11] Clinicians did more of the things we asked them to do, but they also did more of the things we never mentioned; there is no substitution away from one item towards the other. Clinicians were not paid more or promised any increases in pay. This is a large increase in quality with a simple and seemingly inconsequential intervention. It is tempting to discuss bringing such an intervention to scale. However, we know little about the separate roles of the encouragement script and the on-going monitoring and nothing about the generalisability of Dr. Beatus, who administered the script. We prefer to use these results to highlight the lost opportunities in the health care system. It should be natural and normal for all health workers to feel that their work is important and that they are accountable to their peers for the quality of care they provide. Tanzania has multiple supervision systems that are supposed to achieve exactly this aim. Yet clearly most health workers are neither encouraged nor held accountable.

As a final test, we examine the way that generous and non-generous clinicians respond to changes in peer scrutiny. By analysing the way generous clinicians respond to increases in peer stimuli, we have demonstrated that even intrinsically motivated types of health workers respond to changes in intrinsic incentives. Thus, patient-based pro-social behaviour and peer-based esteem-seeking exist side by side in the health field. Furthermore, the two types of motivation are not substitutes for each other: each one can increase the performance of health workers independent of the other.

These results suggest that a complex view of intrinsic motivation is warranted, but they also suggest the opportunity to improve the quality of care using changes in the types of motivation that clinicians face. The fact that clinicians who display social preferences provide higher quality has been alluded to in previous studies (Delfgaauw, 2007; Prendergast, 2007; Serra et al., 2011). However, no one knows how to make sure that only altruistic health care workers enter the workforce. Furthermore our research shows that, although these health workers are better than others, they are still far below their potential. In our study a generous clinician is half a standard deviation better than a non-generous clinician, but an encouraged generous clinician is also a half standard deviation better than a discouraged generous clinician. Thus, the focus on the right type of health care worker may offer less to policymakers than a focus on the right type of intrinsic incentives. In fact, the response to changes in intrinsic incentives may be just as promising as the response to changes in extrinsic incentives, while avoiding the problems of task shifting and motivation crowding out.

[11] The goal of that intervention was to increase output so quality was an indirect outcome, not the sole outcome.

1. Tables

Table 1: Laboratory experiment results

	tokens given				% of subjects with			
# subj.	mean	SD	median	mode	x=0	0≤x<50	x=50	x>50
63	35.03	19.77	40	50	3.17%	58.73%	30.16%	7.94%

Table 2: Summary statistics

variable	mean/count	std dev
Days between encouragement and post-study	45	(25)
Baseline visit		
Doctors observed	63	
Patients observed	408	
Possible items	8,992	
Average quality	74.4%	(16%)
Scrutiny visit		
Doctors observed	59	
Patients observed	338	
Possible items	8,142	
Average quality	74.3%	(15%)
Post-study		
Doctors observed	51	
Patients observed	841	
Possible items	17,937	
Average quality	81.5%	(11%)

Table 3: Generosity and provision of effort in the baseline

	Dependant variable			
	whether or not a doctor performs a particular item			protocol adherence
	(1)	(2)	(3)	(4)
Generous	0.092***	0.086***	0.082***	0.067***
	(0.010)	(0.335)	(0.020)	(0.020)
item effects	dummy variables	dummy variables	fixed effects	gender,age, symptoms
patient effects	gender/age	gender/age	gender/age	
N	7,994	8,002	8,002	358

Marginal effects reported; significant at 1 per cent (***), 5 per cent (**) and 10 per cent (*). (1) Logit regression with dummies for item effect, controlling for gender (G) and age (A) of patient; (2) logit regression with random effects at the unique patient level, with dummies for item effect, controlling for gender (G) and age (A) of patient; (3) linear regression with fixed effects for each unique item, controlling for gender (G) and age (A) of patient, clustered at the unique patient level; (4) linear regression of patient average, controlling for gender (G), age (A) and major symptom (MS).

Table 4: Changes in quality under peer scrutiny and encouragement

	Dependant variable			
	whether or not a doctor performs a particular item			protocol adherence
	(1)	(2)	(3)	(4)
Peer scrutiny (Scr.)	0.031***	0.027***	0.036***	0.029**
	(0.005)	(0.010)	(0.013)	(0.012)
Encouragement (Enc.)	0.058***	0.053***	0.057***	0.078***
	(0.005)	(0.010)	(0.010)	(0.010)
Enc. tracked (Trk.)	0.046***	0.043***	0.077***	
	(0.007)	(0.006)	(0.011)	
clinician effects	dummy variable	dummy variable	dummy variable	fixed effect
item fixed effects	discrim. and difficulty	dummy	fixed effect	none
patient effects	gender, age	gender, age	gender, age	gender, age, major symptom
patient order	-0.001**	0.00	-0.001	0.00
	(0.001)	(0.001)	(0.001)	(0.001)
N	35,296	35,296	35,296	1,595

Marginal effects reported; significant at 1 per cent (***), 5 per cent (**) and 10 per cent (*). (1) Logit regression with dummys for item effect, controlling for gender (G) and age (A) of patient; (2) logit regression with random effects at the unique patient level, with dummys for item effect, controlling for gender (G) and age (A) of patient; (3) fixed effect linear regression with fixed effects for each unique item, controlling for gender (G) and age (A) of patient, clustered at the unique patient level; (4) fixed effect linear regression of patient average, controlling for gender (G), age (A) and major symptom (MS).

Table 5: Changes in quality as a function of the baseline quality

	Dependant variable			
	whether or not a doctor performs a particular item			protocol adherence
	(1)	(2)	(3)	(4)
Peer scrutiny effect				
Peer scrutiny	-0.068***	-0.073***	-0.052**	-0.056**
	(0.017)	(0.031)	(0.024)	(0.024)
Proportion of gap closed	0.285***	0.29***	0.29***	0.295***
	(0.038)	(0.070)	(0.082)	(0.074)
Encouragement effect				
Encouragement	-0.071***	-0.082***	-0.082***	-0.103***
	(0.012)	(0.020)	(0.019)	(0.020)
Proportion of gap closed	0.432***	0.483***	0.527***	0.619***
	(0.033)	(0.059)	(0.070)	(0.060)
clinician effects	dummy variable	dummy variable	dummy variable	fixed effect
item effects	discrim. and diff.	dummy	fixed effect	none
patient effects	gender, age	gender, age	gender, age	gender, age, major sympt.
patient order	-0.001*	0.000	-0.001	0.000
	(0.001)	(0.001)	(0.002)	(0.001)
N	35,296	35,296	35,296	1,595

Marginal effects reported; significant at 1 per cent (***), 5 per cent (**) and 10 per cent (*). (1) Logit regression with dummies for item effect, controlling for gender (G) and age (A) of patient; (2) logit regression with random effects at the unique patient level, with dummies for item effect, controlling for gender (G) and age (A) of patient; (3) fixed effect linear regression with fixed effects for each unique item, controlling for gender (G) and age (A) of patient, clustered at the unique patient level; (4) fixed effect linear regression of patient average, controlling for gender (G), age (A) and major symptom (MS).

Table 6: Changes in provision by type and intervention

	Dependant variable			
	whether or not a doctor performs a particular item			protocol adherence
	(1)	(2)	(3)	(4)
peer scrutiny (Scr.)	0.033***	0.026**	0.038**	0.025*
	(0.006)	(0.012)	(0.017)	(0.015)
generous, Scr.	-0.006	0.002	-0.005	0.011
	(0.011)	(0.021)	(0.026)	(0.023)
encouragement (Enc.)	0.053***	0.054***	0.057***	0.08***
	(0.007)	(0.012)	(0.014)	(0.012)
generous, Enc	0.012	0.000	0.000	-0.006
	(0.010)	(0.019)	(0.020)	(0.020)
Enc. tracked (Trk.)	0.047***	0.042***	0.073***	
	(0.008)	(0.007)	(0.013)	
Enc. Generous Trk.	-0.002	0.004	0.008	
	(0.014)	(0.012)	(0.014)	
clinician effects	dummy variable	dummy variable	dummy variable	fixed effect
item effects	discrm. and diff.	dummy	fixed effect	none
patient effects	gender, age	gender, age	gender, age	gender, age, major sympt
patient order	-0.001**	0.000	-0.001	0.000
	(0.001)	(0.001)	(0.001)	(0.001)
N	35,296	35,296	35,296	1,595

Marginal effects reported; significant at 1 per cent (***), 5 per cent (**) and 10 per cent (*). (1) Logit regression with dummies for item effect, controlling for gender (G) and age (A) of patient; (2) logit regression with random effects at the unique patient level, with dummies for item effect, controlling for gender (G) and age (A) of patient; (3) fixed effect linear regression with fixed effects for each unique item, controlling for gender (G) and age (A) of patient, clustered at the unique patient level; (4) fixed effect linear regression of patient average, controlling for gender (G), age (A) and major symptom (MS).

A. Appendix

A.1 Do health workers react when they discover the team has arrived?

The data analysed in this study were collected from patients by enumerators who had not met the doctors they were studying. They would have no reason to falsify the answers of patients. Patients themselves could not have known what the study was about, and certainly could not have known the stage of the research. Thus, we believe the patients' assessments of quality are an unbiased (though noisy) reflection of what they have seen. Our study measures the impact of encouragement combined with monitoring or studying clinicians. It does not matter whether clinicians increased quality because they were encouraged or because they expected the team to collect data. However, one concern is that if clinicians knew what day we were coming or knew that our team had arrived at the facility, the gains we observe could be unrepresentative of the true changes in quality. In particular, clinicians might hear (from patients or nurses) that the research team had arrived and then change their behaviour in order that the patients might report improvements. If this is the case then our data do not capture real gains.

Since the first few patients we interviewed would have consulted with the clinician before the team arrived it is not possible to alter the true quality for these patients, but subsequent patients might see better (false) quality. To investigate the possibility of false increases, we look for trends in the quality of care with the order of patients on the same day. Over the course of a normal day, the quality of care declines slightly for the average clinician. This is probably due to the changing severity of illnesses reported; those who are very sick tend to queue early at the health facility. Thus, in the baseline – when clinicians knew nothing of the study – quality declines slightly over the course of the day. On the other hand, we know that quality increases significantly when a peer enters the room. Thus, if our enumerators were "discovered" we would observe an immediate increase in quality at the moment of discovery. This pattern should be observable in the quality of care provided with the order of patients. If the enumerators are discovered, the quality should increase with the number of patients in the post-study data collections.

Table 7 looks at the quality of care provided by all clinicians who were observed in both the baseline and post-study and measures the changes in quality of care with the order in which patients were seen. We examine a series of different windows that might capture the moment when a clinician is discovered from the first four patients up to the first eight patients, and also all patients seen on that day. All trends were negative, and there is no statistically significant difference between the trends in the baseline and post-study. This suggests that health workers did not know or care that we had arrived and that the increases seen in the data are representative of what clinicians do on days when we are not at the facility observing them.

Table 7: Quality by order of patients, comparing baseline to post study visits

	Dep Var: Whether a doctor provides a given item					
	patients included (by order)					
	all	1 to 8	1 to 7	1 to 6	1 to 5	1 to 4
Patient order by clinician and day						
baseline	-0.004**	-0.004*	-0.004*	-0.01***	-0.005	-0.011*
	(0.002)	(0.002)	(0.003)	(0.004)	(0.004)	(0.006)
end of study	-0.003***	-0.003***	-0.003*	-0.006	-0.005	-0.009
	(0.001)	(0.001)	(0.001)	(0.004)	(0.004)	(0.006)
time of day	0.001**	0.001	0.002*	0.005	0.006	0.001
	(0.001)	(0.001)	(0.001)	(0.007)	(0.008)	(0.010)
constant	0.797***	0.808***	0.798***	0.78***	0.761***	0.814***
	(0.007)	(0.008)	(0.012)	(0.061)	(0.073)	(0.095)
N	35,903	32,536	29,937	26,659	22,951	18,608

Each regression is a fixed effect regression (linear) with fixed effects for each doctor day.

A.2 Duration of the encouragement effect

In order to measure the true impact of duration, we examine the full dataset, including clinicians who were not present in the laboratory experiment and regress protocol adherence on whether the encouragement and follow-up have taken place, plus the number of weeks since the encouragement, follow-up and any contact with the research team. Contact with the research team includes the encouragement and follow-up as well as visits from the data collection team. This last variable captures the sense that any contact with the team "reminds" the health worker of the research.

Table 8 shows the coefficients as well as the coefficient representing the weekly increment at the limit where weeks since encouragement, follow-up and contact are approximately equal (the sum of each of these coefficients). In all four regressions the increment is positive and is significantly different from zero in all regressions except the second one. Thus, according to our data, quality increases the longer it has been since the research finished.

Thus, the results suggest that the impact of encouragement increases slightly with the length of time since the last contact. However, the 95th percentile of days since encouragement is 13 weeks and the 95th percentile of days since last contact is almost 7 weeks so these coefficients cannot be seen as evidence that quality will increase, or even stay constant in the longer run. However, during the life span of the research, there is no evidence that the impact declines.

Table 8: Duration of encouragement and study effect

	Dependant variable			
	whether or not a doctor performs a particular item			protocol adherence
	(1)	(2)	(3)	(4)
Has the encouragement taken place?	-0.004	0.005	0.003	0.005
	(0.003)	(0.007)	(0.008)	(0.008)
Has the follow-up taken place?	0.018***	0.017***	0.008	0.007
	(0.004)	(0.006)	(0.009)	(0.008)
Weeks since last contact	0.005***	0.003**	0.006***	0.005***
	(0.001)	(0.002)	(0.002)	(0.002)
Weeks since encouragement	0.005***	0.005***	0.008***	0.008***
	(0.000)	(0.001)	(0.001)	(0.001)
Weeks since follow-up	-0.006***	-0.004*	-0.007**	-0.006**
	(0.001)	(0.002)	(0.003)	(0.003)
Limit of per week increment	0.004	0.004	0.007	0.007
p-value	[0.020]**	[0.144]	[0.052]*	[0.021]**
N	84,918	84,918	84,918	3,849

Marginal effects reported in parentheses. Brackets are the p-value of the test that the sum of weeks since last contact, weeks since encouragement and weeks since follow-up is equal to zero. Significance reported: 1 per cent (***), 5 per cent (**), and 10 per cent (*) (1): Logit regression with dummies for item effect, controlling for gender (G) and age (A) of patient; (2) logit regression with random effects at the unique patient level, with dummies for item effect, controlling for gender (G) and age (A) of patient; (3) fixed effect linear regression with fixed effects for each unique item, controlling for gender (G) and age (A) of patient, clustered at the unique patient level; (4) fixed effect linear regression of patient average, controlling for gender (G), age (A) and major symptom (MS).†: The sum of the coefficients for weeks since encouragement, follow-up and last contact. At the limit this is the weekly change in quality.

A.3 Retrospective consultation review

Table 9: Baseline adherence by item and changes by peer scrutiny and encouragement

Item: "Did the doctor . . . "	Baseline	Scrutiny	Enc.
Greeting and receiving			
. . . welcome and greet you?	0.794***	-0.002	-0.001
	(0.166)	(0.027)	(0.021)
. . . listen to your description of the illness?	0.819***	-0.023	-0.021
	(0.166)	(0.027)	(0.021)
Did you have a chair to sit in?	0.814***	-0.024	-0.015
	(0.166)	(0.027)	(0.021)
General history taking and education			
. . . ask you how long you had been suffering?	0.715***	0.044*	0.062***
	(0.166)	(0.027)	(0.021)
. . . ask you if there were other symptoms different from the main complaint?	0.58***	0.080***	0.129***
	(0.166)	(0.027)	(0.021)
. . . ask if you already received treatment elsewhere or took medicine?	0.474***	0.091***	0.202***
	(0.166)	(0.027)	(0.021)
. . . give you a name for your illness?	0.550***	0.043	0.124***
	(0.166)	(0.027)	(0.021)
. . . explain your illness?	0.580***	0.040	0.118***
	(0.166)	(0.027)	(0.021)
. . . explain the treatment?	0.72***	0.034	0.034
	(0.166)	(0.027)	(0.021)
. . . give you advice to improve your health?	0.498***	0.061**	0.172***
	(0.166)	(0.027)	(0.021)
. . . explain if you need to return?	0.503***	0.098***	0.163***
	(0.166)	(0.027)	(0.021)
. . . explain what the drugs are for?	0.725***	0.023	0.046**
	(0.166)	(0.027)	(0.022)
. . . clearly explain instructions for the drugs?	0.67***	0.065**	0.091***
	(0.166)	(0.027)	(0.022)
. . . order a lab test?	0.481***	-0.053**	-0.095***
	(0.166)	(0.027)	(0.021)
If so, . . . explain why you would have this test?	0.765***	-0.014	0.01
	(0.166)	(0.033)	(0.027)
. . . explain why you were referred?	0.633***	-0.020	0.022
	(0.208)	(0.211)	(0.170)
. . . tell you what to do?	0.733***	-0.120	0.047
	(0.208)	(0.211)	(0.170)
Fever, history taking			
. . . ask you how long you had had a fever?	0.564***	0.035	0.095**
	(0.169)	(0.049)	(0.043)
. . . ask you if you had chills or sweats?	0.381**	-0.051	0.174***
	(0.169)	(0.049)	(0.043)
. . . ask you if you had a cough or difficulty breathing?	0.458***	0.011	-0.035
	(0.169)	(0.049)	(0.043)
. . . ask you if you had diarrhoea or vomiting?	0.474***	-0.015	0.087**
	(0.169)	(0.049)	(0.043)
. . . ask if you had a runny nose?	0.464***	-0.023	0.111***
	(0.169)	(0.049)	(0.043)
Fever, history taking, under 5			

. . . ask if the child had had convulsions?	-0.122	0.157**	0.192***
	(0.175)	(0.071)	(0.067)
. . . ask about difficulty drinking or breastfeeding?	0.122	0.214***	0.239***
	(0.175)	(0.071)	(0.067)
Listen to the child's breathing, or use a stethoscope?	0.389**	-0.026	0.160**
	(0.175)	(0.071)	(0.067)
. . . check the child's ear?	0.144	0.066	0.105
	(0.175)	(0.071)	(0.067)
. . . ask questions about the child's vaccinations?	0.078	0.111	0.155**
	(0.175)	(0.071)	(0.067)
Cough, history taking			
. . . ask the duration of the cough?	0.633***	0.069	0.058
	(0.171)	(0.058)	(0.050)
. . . ask if there was sputum?	0.433**	-0.013	0.104**
	(0.171)	(0.058)	(0.050)
. . . ask if you had blood in your cough?	0.180	0.080	0.149***
	(0.171)	(0.058)	(0.050)
. . . ask if you had difficulty breathing?	0.5***	-0.009	0.106**
	(0.171)	(0.058)	(0.050)
. . . ask if you also have a fever?	0.62***	-0.002	0.017
	(0.171)	(0.058)	(0.050)
Cough, history taking, under 5			
. . . ask about difficulty drinking or breastfeeding?	0.091	0.256***	0.233***
	(0.180)	(0.087)	(0.081)
. . . ask if the child had convulsions?	-0.038	0.02	-0.062
	(0.180)	(0.087)	(0.081)
. . . check the child's ear?	0.091	0.061	0.153*
	(0.180)	(0.087)	(0.081)
. . . ask if the child had diarrhoea or vomiting?	0.285	0.013	0.125
	(0.180)	(0.087)	(0.081)
. . . ask about the history of vaccinations?	-0.033	0.273***	0.201 **
	(0.180)	(0.088)	(0.082)
Diarrhoea, history taking			
. . . ask how long you have had diarrhoea?	0.633***	-0.144	-0.009
	(0.183)	(0.112)	(0.090)
. . . ask how often you have a movement?	0.473***	-0.166	0.145
	(0.183)	(0.112)	(0.090)
. . . ask about the way the stool looks?	0.553***	-0.126	0.019
	(0.183)	(0.112)	(0.090)
. . . as if there was blood in the stool?	0.313*	0.085	0.216**
	(0.183)	(0.112)	(0.090)
. . . ask if you are vomiting?	0.553***	-0.087	0.034
	(0.183)	(0.112)	(0.090)
. . . as if you also have a fever?	0.633***	-0.341***	0.005
	(0.183)	(0.112)	(0.090)
Diarrhoea, history taking, under 5			
. . . ask about difficulty drinking or breastfeeding?	0.333	-0.044	0.099
	(0.232)	(0.173)	(0.166)
. . . ask if the child had convulsions?	0.167	-0.174	-0.116
	(0.232)	(0.173)	(0.166)
. . . check the child's ear?	-0.167	0.294*	0.384**
	(0.232)	(0.173)	(0.166)
. . . ask if the child had diarrhoea or vomiting?	0.667***	-0.057	0.170

	(0.232)	(0.173)	(0.166)
. . . ask questions about the child's vaccinations?	0.167	0.099	0.027
	(0.232)	(0.173)	(0.166)
Fever, diagnostic			
. . . take your temperature?	0.481***	0.018	0.129***
	(0.169)	(0.048)	(0.042)
. . . check for neck stiffness?	0.110	0.060	0.061
	(0.169)	(0.049)	(0.042)
. . . ask if you felt weakness from lack of blood?	0.129	0.114**	0.104**
	(0.169)	(0.049)	(0.042)
. . . look in your ears or throat?	0.071	0.097**	0.122***
	(0.169)	(0.049)	(0.042)
. . . check your stomach?	0.129	0.050	0.032
	(0.169)	(0.049)	(0.042)
. . . ask for a blood slide?	0.548***	-0.022	0.026
	(0.169)	(0.049)	(0.042)
Fever, diagnostic, under 5			
. . . check if the child was sleepy, try to wake the child?	0.116	0.107	0.099
	(0.175)	(0.071)	(0.067)
. . . pinch the skin fold of the child?	0.100	0.065	0.121*
	(0.175)	(0.071)	(0.067)
. . . check both of the child's feet?	-0.033	0.032	0.200***
	(0.175)	(0.071)	(0.067)
. . . check the child's weight against a chart?	0.189	0.092	0.085
	(0.175)	(0.071)	(0.067)
Cough, diagnostic			
. . . look at your throat?	0.233	0.014	0.025
	(0.171)	(0.059)	(0.050)
. . . listen to your chest?	0.460***	0.078	0.112**
	(0.171)	(0.059)	(0.050)
. . . take your temperature?	0.407**	-0.011	0.057
	(0.171)	(0.059)	(0.050)
Cough, diagnostic, under 5			
. . . check if the child was sleepy, try to wake the child?	0.133	0.095	0.060
	(0.180)	(0.088)	(0.082)
. . . pinch the skin fold of the child?	0.059	0.157*	0.133
	(0.180)	(0.087)	(0.081)
. . . check the child's eyes, tongue and palms?	0.124	0.069	0.122
	(0.180)	(0.087)	(0.081)
. . . check both of the child's feet?	-0.07	0.113	0.149*
	(0.180)	(0.087)	(0.081)
. . . check the child's weight against a chart?	-0.005	0.194**	0.107
	(0.180)	(0.087)	(0.081)
Diarrhoea, diagnostic			
. . . take your temperature?	0.193	-0.094	0.282***
	(0.183)	(0.112)	(0.090)
. . . pinch the skin on the stomach?	0.073	0.139	-0.066
	(0.183)	(0.112)	(0.090)
Diarrhoea, diagnostic, under 5			
. . . check the child's eyes, tongue and palms?		0.182	0.170
		(0.176)	(0.166)

. . . check both of the child's feet?	0.000	-0.218	0.241
	(0.232)	(0.176)	(0.166)
. . . check the child's weight against a chart?	0.333	-0.303*	-0.116
	(0.232)	(0.176)	(0.166)
. . . offer the child a drink of water or observe breastfeeding?	0.000	-0.118	0.170
	(0.232)	(0.176)	(0.166)
If the child is under two years, . . . look at the child's head?	-0.125	-0.057	-0.048
	(0.184)	(0.117)	(0.091)
General, diagnostic			
. . . examine you?	0.538***	0.042	-0.026
	(0.166)	(0.030)	(0.024)

References

G. A. Akerlof and R. E. Kranton (2000), "Economics and Identity", *The Quarterly Journal of Economics*, 115 (3), 715-753.

G. A. Akerlof and R. E. Kranton (2005), "Identity and the Economics of Organizations", *Journal of Economic Perspectives*, 19 (1), 9-32.

J. Andreoni (1989), "Giving with Impure Altruism: Applications to Charity and Ricardian Equivalence", *Journal of Political Economy*, 97 (6), 1447-58.

J. Andreoni (1990), "Impure Altruism and Donations to Public Goods: A Theory of Warm-Glow Giving", *Economic Journal*, 100 (401), 464-77.

P. Basinga, P. J. Gertler, A. Soucat and J. Sturdy (2011), "Effect on maternal and child health services in Rwanda of payment to primary health-care providers for performance: an impact evaluation", *The Lancet*, 377 (9775), 1421-1428.

G. S. Becker (1974), "A Theory of Social Interactions", *Journal of Political Economy*, 86 (6), 1063-1093.

R. Benabou and J. Tirole (July 2003), "Intrinsic and Extrinsic Motivation", *Review of Economic Studies*, 70 (3), 489-520.

M. Bjorkman and J. Svensson (2009), "Power to the People: Evidence from a Randomized Field Experiment on Community-Based Monitoring in Uganda", *The Quarterly Journal of Economics*, 124 (2), 735-769.

G. Charness and M. Rabin (2002), "Understanding Social Preferences with Simple Tests", *The Quarterly Journal of Economics*, 117 (3), 817-869.

J. B. Cullen (1978), The Structure of Professionalism: a Quantitative Examination, New York: PBI.

J. Das and J. S. Hammer (2007), "Money for Nothing, The Dire Straits of Medical Practice in Delhi, India", *Journal of Development Economics*, 83 (1), 1-36.

J. Das, J.S. Hammer and K. L. Leonard (2008), "The Quality of Medical Advice in Low-Income Countries", *Journal of Economic Perspectives*, 22 (2).

J. Delfgaauw (January 2007), "Dedicated Doctors: Public and Private Provision of Health Care with Altruistic Physicians", Technical Report.

T. Ellingsen and M. Johannesson (2008), "Pride and Prejudice: The Human Side of Incentive Theory", *American Economic Review*, 98 (3), 990-1008.

L. M Franco, S. Bennett and R. Kanfer (April 2002), "Health sector reform and public sector health worker motivation: a conceptual framework", Soc Sci Med, 54 (8), 1255-1266.

E. Freidson (2001), Professionalism: The Third Logic, The University of Chicago Press.

F. Herzberg (1959), The Motivation to Work, John Wiley and Sons, New York.

F. Herzberg (Sept-Oct 1987), "One More Time: How do you motivate Employees?", Harvard Business Review.

K. L. Leonard and M. C. Masatu (2006), "Outpatient process quality evaluation and the Hawthorne Effect", Social Science and Medicine, 63 (9), 2330-2340.

K.L. Leonard and M. C. Masatu (2010), "Professionalism and the know-do gap: exploring intrinsic motivation among health workers in Tanzania", *Health Economics*, 19 (12), 1461-1477.

K. L. Leonard and M. C. Masatu (2010), "Using the Hawthorne effect to examine the gap between a doctor's best possible practice and actual performance," *Journal of Development Economics*, 93 (2), 226-243.

K. L. Leonard, M. C. Masatu and A. Vialou (2007), "Getting Doctors to do their best: the roles of ability and motivation in health care", *Journal of Human Resources*, 42 (3), 682-700.

S. Levitt and J. List (2007), "What do Laboratory Experiments Measuring Social Preferences Reveal about the Real World?", *Journal of Economic Perspectives*, 21 (2), 153-174.

O. Maestad and G. Torsvik (2008), "Improving the Quality of Health Care when Health Workers are in Short Supply", mimeo, Chr. Michelsen Institute.

O. Maestad, G. Torsvik and A. Aakvik (September 2010), "Overworked? On the relationship between workload and health worker performance", *Journal of Health Economics*, 29 (5), 686-698.

B. Meessen, L. Musango, J-P. I. Kashala and J. Lemlin (2006), "Reviewing Institutions of rural health centres: the Performance Initiative in Butare, Rwanda", Tropical Medicine and International Health, 11 (8), 1303-1317.

C. Prendergast (March 2007), "The Motivation and Bias of Bureaucrats", *American Economic Review*, 97 (1), 180-196.

M. Rabin (December 1993), "Incorporating Fairness into Game Theory and Economics", *American Economic Review*, 83 (5), 1281-1302.

A. K. Rowe, D. de Savigny, C. F. Lanata, and C. G. Victora (17 September 2005), "How can we achieve and maintain high-quality performance of health workers in low-resource settings?", *The Lancet*, 366, 1026-1035.

D. Serra, P. Serneels and A. Barr (2011), "Intrinsic Motivations and the Nonprofit Health Sector", Personality and Individual Differences, 51 (3), 309-314.

R. Sugden (1982), "On the Economics of Philanthropy", *Economic Journal*, 92 (366), 341-350.

World Health Organization, "Bridging the 'Know-Do' gap: Meeting on Knowledge Translation in Global Health", Technical Report WHO/EIP/KMS/2006.2 2005.