How to encourage service delivery to the poor: Intrinsic motivation, extrinsic incentives, and effort

Sheheryar Banuri (University of East Anglia); Damien de Walque (World Bank); Philip Keefer (Inter-American Development Bank); Paul Jacob Robyn (World Bank)[1]

February 2023

Abstract

Do pecuniary incentives compensate for lack of motivation when workers can choose between differently motivating tasks? We examine this question, focusing, for the first time, on intrinsic motivation to serve the poor. We use a lab-in-the-field experiment with nearly 400 health care workers in Burkina Faso and a novel "virtual clinic" task. The experimental design recognizes that poor patients are often harder to treat and service providers can choose who to serve. Pecuniary incentives trigger dramatically different behavioral changes in the presence of this choice compared to those observed in earlier research. Specifically, incentives have little effect on unmotivated workers but strongly affect the behavior of intrinsically motivated workers. These findings have significant policy implications. For example, in health, where disparities in access of the poor and non-poor are notorious and the difficulties of treating poor patients are well-known, the findings suggest that fixed salaries could be more effective than policies aimed at providing pecuniary rewards for treating poor patients: they better account for the counter-intuitive effects of pecuniary rewards on the behavior of more motivated health care workers.

JEL Codes: C91; H83; I14; I32; J45; M52
Keywords: public health, access, motivation, performance, inequality, socioeconomic status

[1] We are extremely grateful for a fruitful collaboration with the Ministry of Health in Burkina Faso and in particular with Philippe Compaore, Baly Ouattara, Abdoul Karim Ouedraogo, Boukary Tapsoba and Paul Toe. We are also very grateful to all participants in the experiment and those who facilitated it, in particular the direction, staff and students from the Faculté de Médecine de l'Université de Ouagadougou, the Ecole Nationale de Santé Publique and the Ecole de Santé privée Sainte Edwige, as well as the direction and staff of the health facilities in the regions of Boucle du Mouhoun, Centre-Nord, Centre-Ouest and Nord. The data were collected by Centre Muraz with special thanks to Hervé Hien, Nicolas Meda and all staff implementing the lab in the field experiment. We are also grateful to Nouna Health Research Centre, through Maurice Ye who developed the clinical-based scenarios for the videos. Special thanks to Bambio Yiriyibin for coordinating and supervising the experiment and to Boubakar Diallo and the team from Les Films du Dromadaire for producing the videos. We are also grateful to Ingmar Weber, Haidera Ousmane Diadie, Maurice Ye, Ellen Green, participants of the Measuring the Tricky Things Workshop (Oxford), participants of the Institute for Business Administration seminar series, and participants of the Centre for Behavioral and Experimental Social Science (CBESS) seminar series, the 2nd IUSSP Population, Poverty and Inequality Research Conference, the International Health Economics Association (iHEA), and the Essen Health Conference 2021, for helpful comments. This research is funded by the Health Results Innovation Trust Funds (HRITF) at the World Bank. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the World Bank, its Executive Directors, or the countries they represent.

Introduction

The challenge of delivering services to the poor is widely recognized, present everywhere, and linked to worse health and education outcomes.[2] We present results from a novel experiment that examines the role of pecuniary incentives in solving this problem, showing that worker discretion over tasks has important implications for how pecuniary and intrinsic incentives operate on motivated and unmotivated workers. The experimental design mirrors many service delivery contexts in addition to the health context: workers can choose between tasks, and one task (providing services to the poor) is both more difficult and offers more intrinsic rewards than the other (serving the non-poor). Public policy, guided by substantial prior research, frequently turns to pecuniary incentives in order to encourage unmotivated workers to undertake the more difficult task. The results of the experiment suggest, however, that in a context where motivated workers can choose between tasks, the efficacy of pecuniary incentives may be limited.

We explore these issues using a lab-in-the-field experiment[3] with 379 health care workers in rural Burkina Faso. The experiment features novel video vignettes – the "virtual clinic" task – and introduces an incentivized measure of intrinsic motivation to serve the poor. The experiment is the first to adapt video vignettes (see Banuri et al. 2018) to simulate poverty in order to study the impact of incentive contracts targeting poor populations in a healthcare service delivery context. Participants (professional health workers) were asked to review medical cases using these video vignettes. The vignettes featured poor and non-poor patients presenting symptoms associated with common maternal and child health conditions. The vignettes for poor cases included more complex information and took longer to view, reflecting real world differences in the costs of treating poor and non-poor patients.

Participants undertake the virtual clinic task under four different contracts (treatments) that are commonly used in health sectors around the world. One, the baseline contract, was an unconditional *salary* for engaging in the task (low powered incentive). Three conditional bonus contracts then vary the relative pecuniary incentives to serve the poor relative to the non-poor.

(i) The "*Non-poor bonus*" contract paid an unconditional salary plus a piece rate (bonus) for treating non-poor patients, but no bonus for the poor, reflecting the common situation where poor patients cannot afford user fees.

(ii) The "*Equal bonus*" contract paid an unconditional salary plus an equal piece rate for treating either poor or non-poor patients.

---

[2] On health outcomes for the poor, the focus of this paper, see Peters et al. (2008); Bindman et al. (1995); Ensor and Sam (1996); Haelterman et al. (2003); Joseph and Philips (1984); Ku and Matani (2017); Sudore et al. (2006); Raso et al. (2005); Weissman et al. (1991).

[3] The lab-in-the-field experiment begins by using dictator games to measure health workers' intrinsic motivation to serve the poor, which allows us to distinguish motivated from unmotivated workers. Next, workers view video vignettes in which poor and non-poor patients describe symptoms associated with common maternal or child health conditions (the "virtual clinic" task). This virtual clinic introduces a new way of simulating poor patients in a lab setting (see Banuri et al. 2018). The vignettes with poor patients present identical health conditions but include extraneous information and take 66 percent longer to view (100 seconds compared to 60 seconds), reflecting real world differences in the costs of treating poor and non-poor patients. After viewing a vignette, the health workers then answer a series of questions about patient diagnosis and treatment.

(iii) The "*Poor bonus*" contract was the same as the Equal bonus contract with one modification: the piece rate paid for treating poor patients was two times the piece rate paid for treating non-poor patients.[4]

Within each treatment (contract), healthcare workers choose how many and which type of cases (poor or non-poor) to review under a fixed time constraint.[5] Our main dependent variable is the proportion of poor cases reviewed relative to total cases.[6] We focus on two questions: First, how do pecuniary incentives affect services to the poor? Second, intrinsic motivation to serve the poor plays a significant role in accounting for different levels of effort directed toward poor patients. Therefore, what are the relative impacts of the introduction of pecuniary incentives on workers motivated to serve the poor, and workers unmotivated to serve the poor?

We find that pecuniary incentives to serve the poor are limited in their effectiveness overall: the Poor Bonus contract nominally (but not significantly) increases the proportion of poor cases reviewed relative to a fixed salary benchmark, and relative to the Equal Bonus contract benchmark. However, when the poor cannot pay user fees (the Non-poor Bonus contract), they are significantly under-served: health workers direct a much larger share of their effort towards non-poor cases. This yields an asymmetric effect, where the Non-poor Bonus contract yields a sharp reduction in service delivery for the poor, but a Poor Bonus contract yields a much smaller increase.

Second, and in contrast to the literature, we find that the change in service provision to the poor is due to differences in effort by *motivated* workers, while unmotivated workers exhibit no change. Motivation to serve the poor is a significant determinant of effort: over 40 percent of all cases that participants chose to review in the baseline (fixed salary) contract were poor cases. Under the status quo Non-poor Bonus, the reduction in effort towards the poor is disproportionately among *motivated* workers. Similarly, the increase in effort towards the poor under the Poor bonus contract is due to motivated workers. Unmotivated workers are relatively unresponsive to the contracts (treatments).

Nevertheless, we do find that the Poor Bonus contract is the only one that achieves equity between poor and non-poor cases, indicating the importance such contracts have for policy. In

---

[4] The status quo in many countries implicitly offers larger extrinsic rewards to serving the non-poor. For example, public sector service providers are paid little for seeing patients and may demand (informal) private payments that exceed poor patients' ability to pay, yielding a de facto bonus for serving non-poor patients. Other countries seek equity in the provision of health care by establishing provider payment mechanisms that compensate providers equally for serving poor and non-poor patients. This approach assumes that the effort or time required to serve poor and non-poor patients is the same. However, the poor often require greater effort, in which case an equal bonus system effectively provides incentives to treat non-poor patients. Finally, some countries (e.g., Burkina Faso, Cameroon, the Central African Republic, the Democratic Republic of Congo and Nigeria) have gone further and introduced extra financial incentives to service providers to increase utilization of essential health services by the poor.

[5] The experiment included incentives for healthcare workers to treat the virtual patients accurately: correct diagnosis and treatment resulted in benefits provided to the targeted group. Participants were shown pictures and given details about two actual schools, one poor and one non-poor. Every poor case they correctly diagnosed and treated yielded a donation to the poor school; reviewing non-poor cases resulted in the same donation to the non-poor school, simulating the benefits accrued to each group from correct diagnosis and treatment. The amount of the donation for reviewing each case was constant across all treatments.

[6] One natural question that emerges is the triage process that occurs in healthcare, which is (in itself) the subject of a large literature. We do not address this issue here, except to state that the health workers we interviewed report having considerable discretion in triaging and treating different types of patients. While there is little evidence for precisely how health workers prioritize patients in developing countries, our experiment uses the choice of patient ordering to proxy for effort exerted by healthcare workers for different types of patients.

sum, the evidence confirms that the *status quo* in many countries (*de facto* bonuses to treat the non-poor because the poor cannot afford to pay) have a large and detrimental effect on the treatment of the poor (relative to an unconditional salary baseline). This detrimental effect emerges because of the reaction of *motivated* workers to such a pay structure. The search for new pay schemes to correct these distortions is therefore well-founded. We find limited support for using pecuniary bonuses to correct these distortions, however.

Countries around the world use pecuniary incentives to motivate greater provision of services to the poor. Intuitive advice on how to structure these incentives comes from a burgeoning literature on worker motivation, which demonstrates theoretically, and empirically, that high powered incentives are effective for unmotivated workers, but inefficient for motivated workers (Frey and Oberholzer-Gee, 1997; Deci and Ryan, 1999; Delfgaauw and Dur, 2008; Carpenter and Gong, 2016; Banuri and Keefer, 2016). Workers who are intrinsically motivated are more likely to exert higher effort without high-powered incentives and additional pecuniary incentives yield diminishing returns or motivation crowding (Carpenter and Gong 2016; Banuri and Keefer 2016; Jones, Tonin and Vlassopoulos, 2018). The lesson for managers is (1) focus on organizational missions to attract motivated workers, or (2) provide high powered incentives for unmotivated workers.

The insights in the literature emerge from settings with single tasks that vary in the degree to which they match the mission motivation of the worker. However, in many mission-based organizations (healthcare providers, for example), workers face multiple tasks that differ both in how intrinsically motivating they are and the effort they require. We show that in such a setting, pecuniary incentives have the largest effect on the *most* intrinsically motivated, increasing their effort. Specifically, high-powered incentives to serve poor patients do not close the effort gap between motivated and unmotivated workers that routinely emerges under low-powered compensation schemes.

Our results have substantial implications for public policy when workers have discretion in the effort they allocate between more and less intrinsically motivating tasks. First, a focus on the organizational mission and the recruitment of motivated workers yields greater effort on behalf for the poor. Second, low-powered incentives may be more cost-effective since high-powered incentives to serve the poor yield increased effort only if they are large and expensive, since to be effective with unmotivated workers, they must more than compensate for the extra costs of serving the poor.

Contribution

Our analysis contributes to the large behavioral literature on intrinsic and extrinsic motivation. One theme in that literature is the interaction of extrinsic and pro-social motivations on effort: the effects of piece rate versus salary-based compensation, for example, when task or mission motivation vary (Frey and Oberholzer-Gee 1997; Deci and Ryan 1999; Delfgaauw and Dur 2008; Carpenter and Gong 2016; and Banuri and Keefer 2016). Jones, Tonin and Vlassopoulos (2018) extend this work to show intrinsic and extrinsic motivation interact to affect the quality and quantity of effort on a task when piece rate contracts provide incentives to product quantity but not

quality. Ashraf et al., (2014) find that agents who are offered non-financial rewards exert more effort than those offered financial margins or volunteer contracts.[7]

This study builds on prior work in two ways. First, it focuses on intrinsic motivation to serve the poor, specifically. Second, it examines a common task setting that has not previously received attention, in which workers have discretion over which task they undertake. One task is both more difficult and offers more intrinsic benefits to some workers, and the other is easier but does not offer these intrinsic benefits. We find a significant effect of intrinsic motivation to serve the poor and demonstrate a large and significant interaction between intrinsic and extrinsic incentives when workers have a choice of task.

The analysis and experimental design also contribute to a substantial literature on motivation among health care providers. Methodologically, the lab setting identifies incentive effects on health workers without risk to actual patients. It allows us to closely control differences in effort needed for poor and non-poor patients and to investigate the interaction of pecuniary and non-pecuniary (intrinsic) incentives of health workers. In contrast, most prior research in this area uses observational data. It does not examine the effects of bonuses in the presence of patient groups that are heterogeneous with respect to the costs of care, nor of health workers who are heterogeneous with respect to their intrinsic motivation to serve the poor.

Our work also differs along several notable dimensions from previous research that has employed laboratory experiments to study related issues. The body of research initiated by Hennig-Schmidt, Selten and Wiesen (2011) compares differences between capitation (a fixed payment per patient meant to cover all care the patient receives) and fee-for-service contracts, usually allowing for heterogeneity in patient health. They report under-provision of health services under capitation and over-provision with fee-for-service, but the differences are narrower than theory would predict (Hennig-Schmidt, Selten and Wiesen 2011; Green, 2014). They conclude that other-regarding preferences are responsible for the discrepancy between theory and actual behavior (Brosig-Koch et al., 2017).

We expand on this work in several ways. Subjects in prior research are university students and/or medical students in high and middle-income countries (Hennig-Schmidt, Selten and Wiesen 2011; Green 2014; Hennig-Schmidt and Wiesen 2014; Lagarde and Blaauw 2017; see Galizzi and Weisen, 2018 for an overview). We focus on the behavior of actual health practitioners, doctors, nurses and midwives currently deployed in rural areas. Our lab-in-the-field experiment is also the first to focus on equity issues, the treatment of a historically disenfranchised group (the poor). We explicitly measure subject preferences for serving the poor. The experiment incorporates an effort task that is closer to the context faced by health workers and does not abstract from the medical context. It implements simplified contractual arrangements that are typical of a developing country context. This is also the first attempt, whether related to a health or some other task, to examine the interaction of pecuniary and non-pecuniary incentives when subjects can choose among multiple tasks. Prior research has not addressed incentives to treat "difficult" (poor) patients versus "easy" (non-poor) patients.

Finally, although prior experiments distinguish the health status of patients, they do not vary the difficulty of dealing with patients – for example, how much time they demand – holding heath

[7] A related strand of literature examines selection into public service: Dal Bó, Finan and Rossi (2013), Banuri and Keefer (2016), and Barfort et al. (2019) examine how pay levels affect selection into public sector positions of more capable and mission-motivated workers.

status constant. In contrast to the tasks used in earlier experiments, our video vignettes allow us to identify separately the effort required to serve different patient types, in this case poor and non-poor patients.

Motivation

The disparities in health care treatment of the poor and non-poor are widely recognized; in addition, ample evidence suggests that the treatment of the poor is more difficult for health care providers. One well known indicator of potential disparities is the positive relationship between life expectancy and per capita income (see Figure 1 for this correlation across all countries in 2014). Figure 2 focuses on Burkina Faso. In 2010, child mortality in Burkina Faso was 97 per thousand live births in the richest wealth quintile but 175 among the poorest quintile; similarly, 93.2 percent of women coming from the richest quintile delivered in a health facility, but only 46.1 percent of those in the poorest quintile did so (Institut National de la Statistique et de la Démographie - INSD/Burkina Faso and ICF International, 2012).

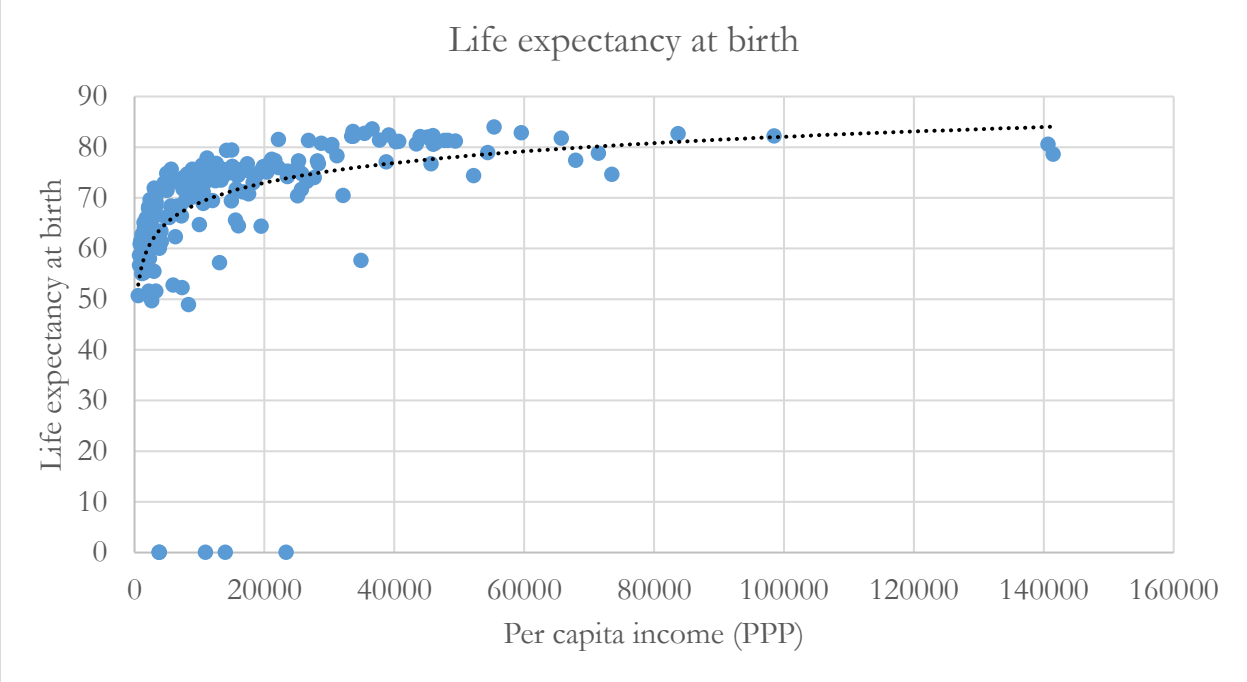


Figure 1: Relationship between life expectancy and per capita income in 2014 (source: World Bank Development Indicators)

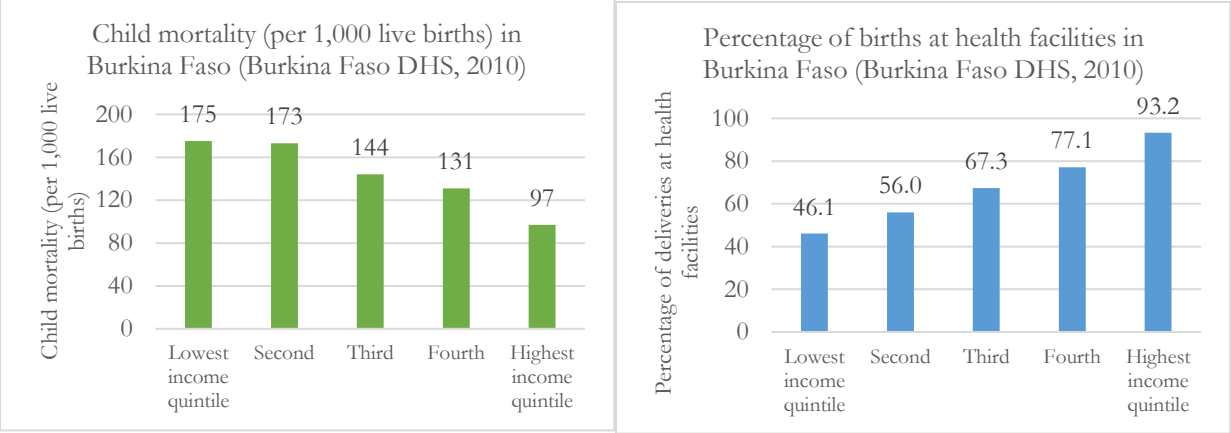


Figure 2: Income quintiles and health outcomes in Burkina Faso (source: Burkina Faso DHS, 2010)

In the past two decades, Burkina Faso has introduced several reforms that aim at improving financial access to health services for the poor[8], including the reduction of user fees for obstetric services in 2007, exemption of user fees for the worst off (les indigents) in 2009, Results-Based Financing, which introduced pro-poor incentives to health care providers in 2014, and generalized free health care for women and children in 2015 (*la gratuité*).[9] Globally, this (considerable) effort to improve health outcomes among the poor has had mixed impact. At a more micro-level, Das and Hammer (2007) and Das and Mohpal (2016) attribute a large share of the inequality in quality of care observed in India to the location and lack of information of poor patients, though Das and Sohnesen (2007) find little evidence of discrimination in health worker effort in Paraguay.

Two features of the disparity in health care access are central to our analysis. One is that health care workers have discretion over the patients they treat. If poor individuals require more resources, therefore, inequity in care is inevitable. Lipsky (2010) studies "street-level bureaucrats" (i.e., public servants who engage directly with the public), and attributes service delivery failure to corruption, misuse of discretion, and lack of resources. That is, health workers have limited resources and the discretion to allocate them as they see fit.

The other key feature is that improving health and education outcomes among the poor requires greater effort, in part because of the greater investment of time that is required (Ingersoll, 2004; Peters et al., 2008; Wagstaff, Bredenkamp and Buisman, 2014; Loignon et al. 2015; Willems et al. 2005; Street, 1992). In education, Ingersoll (2004) describes these as obstacles to staffing high-poverty schools. The poor also present more complex health conditions. Due to a lack of access, resources, and health-related knowledge, poor individuals tend not to invest in preventative care, yielding additional symptoms that make it harder for health care workers to arrive at the correct diagnosis (Wagstaff, Bredenkamp and Buisman, 2014; Peters et al., 2008; among others).

---

[8] See, for example, Robyn et al. (2012), and Fink et al. (2013).

[9] A recent study on the reduction and exemption of user fees showed that the majority of public sector health workers were unaware or at least did not understand these directives, and this contributed to weak implementation and adherence to the policy changes (Ridde et al., 2018).

Communication problems can arise because of differences in the use of language, but also in social standing: poor individuals might be shy or overwhelmed by the medical institution (Loignon et al. 2015; Willems et al. 2005; Street, 1992). Heinig (2009) emphasizes the reluctance of low-income patients to confide in medical professionals in maternal contexts like those in the experimental vignettes used in this paper. Loignon et al. (2015) report three main barriers to care for the poor: the complexity of the health care system, low standards of living among the poor (yielding greater complexity in cases), and poor quality of interaction between health workers and the poor, the latter rooted in problems of communication and social distance.

Researchers have argued that these obstacles affect health care. Willems et al. (2005) report the results of a systematic review that found that patients from low socio-economic backgrounds received less information, direction, and emotional support from health workers. Street (1992) analyzed audio recordings of 115 pediatric consultations and found that parents from weaker educational backgrounds received lower levels of support, due to weaker communication. These papers point to the importance of communication and complexity in treating poor patients, requiring medical professionals to spend greater time in serving the poor. Our experimental design incorporates these characteristics.

Since the discretion of health care workers is a critical element of effective care and is difficult to restrict (for example, by mandating levels of care per patient), the policy challenge is how to reduce inequity without reducing discretion. This is made even more difficult because, as researchers have noted, simply expanding resources for health care may widen these disparities. Gwatkin (2005) warns that because expanded health services typically reach wealthier groups before disadvantaged ones, poor people are unlikely to be the main beneficiaries of efforts to accelerate progress towards the health Millennium Development Goals (MDG) by providing additional resources to the health sector. Analyzing nationally representative surveys in 64 developing countries over the 1990-2011 period, Wagstaff, Bredenkamp and Buisman (2014), establish that the poorest 40 percent have made faster progress than the richest 60 percent on MDG intervention indicators. In terms of MDG outcome indicators, however, inequality has been growing in close to half of the countries. They conclude that inequalities remain substantial with poor children more likely to die and be malnourished and less likely to receive necessary health services.

Experimental design

The lab-in-the-field experiment has several features that allow us to analyze the effects of four compensation schemes (Salary, Non-poor bonus, Equal bonus, Poor bonus) on health worker effort towards the poor. We measure health worker effort across medical cases and types of patients; we incorporate health conditions that are common to the poor and non-poor; and we introduce plausible differences between poor and non-poor patients in the presentation of cases. The latter is a critical aspect of our experimental design that has not been attempted in the literature. We discuss these features in the remainder of this section.

Measuring effort and distinguishing poor and non-poor cases

Existing methodologies for measuring health worker effort are not easily adapted to testing the differential effect of different incentive schemes on effort for the poor and non-poor. Paper-based medical vignettes, for example, are a common tool for measuring performance (Glassman et al. 2000; Peabody, Luck et al. 2000 and 2004; Peabody, Tozija et al. 2004; Das and Hammer 2005;

Veloski et al. 2005). The characteristics that distinguish poor patients are difficult to incorporate into such vignettes, however.

To more accurately assess effort on behalf of actual patients, and to reduce bias induced by Hawthorne effects, researchers have used trained actors (Das et al., 2016; Green et al. 2017). This yields higher accuracy in measurements of health worker ability and effort when confronted with actual patients but is difficult to scale up to test the effects of multiple incentive schemes. Some researchers have used generic real effort and framed choice tasks (Hennig-Schmidt, Selten and Wiesen, 2011; Green, 2014; Hennig-Schmidt and Wiesen, 2014; Cox, Green, and Hennig-Schmidt, 2016; Lagarde and Blaauw, 2017) to measure effort, but these also do not lend themselves to simulating poverty. This approach, by construction, excludes the possibility of observing the effort effects of different patient types and minimizes the complexity of health worker decision making.

With the help of health researchers in Burkina Faso, we developed a series of video vignettes that offset the disadvantages of paper-based vignettes and reliance on trained actors.[10] The videos present a patient describing her symptoms, along with information relevant for arriving at a diagnosis (blood pressure, temperature, etc.). Health workers are shown the video and asked four multiple choice questions:

(i) What is the most probable diagnosis?
(ii) What is the most appropriate treatment?
(iii) When should you see the patient for a follow-up after the completion of the initial treatment?
(iv) What is likely to be the best alternative treatment for the patient (for example, if the patient's condition does not improve)?

Each question has five pre-determined responses, only one of which is correct.[11] To ensure that the cases were common in rural Burkina Faso and among both poor and non-poor patients, we focused on the domain of child and pre-natal care. To ensure that the cases would exhibit adequate variation in effort, we pretested cases with a sample of nursing students in the capital city, Ouagadougou. This process yielded a total of 20 cases that demonstrated adequate variation in responses (i.e., were neither too "easy" nor too "difficult").[12]

Simulating poverty

Once we had our pool of cases, the next challenge was to generate two presentations of each case to simulate poor and non-poor cases. We aimed to make the differences between our poor and non-poor cases as plausible and immersive as possible. The cases therefore differed in terms of speech patterns, time needed to gather the information necessary for diagnosis and treatment, and observable beneficiary characteristics.

Specifically, the non-poor scripts for each case were written to keep to a standard video length of 60 seconds, while still conveying all the information required to arrive at a diagnosis.

---

[10] See Banuri et al. (2018) and http://www.rbfhealth.org/resource/video-vignettes-lab-field-experiment-burkina-faso for more details on measuring ability and effort with video vignettes.

[11] The medical cases that formed the basis of the vignettes were developed with a health researcher in Burkina Faso, Dr. Maurice Ye.

[12] Please see Banuri et al. (2018) on more details on the construction of medical cases.

While there are many differences between poor and non-poor cases in the real world, three dimensions were particularly salient during interviews with health practitioners conducted prior to designing the experiment: (i) time, (ii) complexity, and (iii) communication.

(i) Time: Because of complexity and communication issues, poor cases take 40 seconds longer than non-poor cases.
(ii) Complexity: Poor cases are more complex. The scripts for the poor cases were modified to add additional symptoms that were not relevant for the diagnosis.
(iii) Communication: Especially in the case of the ultra-poor, substantial communication problems can arise. To mimic this, in the script for the poor vignettes the patient digresses from the medical problem and brings up non-medical, irrelevant subjects.

Below is the English translation of a script used in the study. The underlined sentences are used only in the poor version of this case.

> "Hello Doctor. My husband and I come from a village far from here. It is beyond the hill, just after the area with the thorny bushes. We had to walk for more than two hours in order to get your help for our child. He is 6 months old, and does not feel well at all. He has been coughing for more than 5 days. He has a runny nose and his body is very hot. My poor child, we can feel that he is suffering a lot. When he coughs, we can hear from a distance whistling sounds. My child is very tired and he is not breastfeeding as usual. Last night I did not sleep at all, because his breathing was heavy and fast. But it didn't stop my husband from snoring as usual. This morning, my baby seems a bit agitated; he cries incessantly, and his face is paler than usual. Help us Doctor. Save our child."

Care was taken to maintain consistency among all poor cases and all non-poor cases. The timing of the former adhered closely to the 60 second limit and the latter to the 100 second limit. The same actress was used in all videos. In addition, in poor cases the actress was dressed poorly (her clothes were muddy and tattered). In the non-poor version, the same actress was well-dressed.[13] Poverty was simulated using only these features. The instructions and the task itself never refer to the cases in terms of "poor" or "non-poor" as the term may carry different interpretations depending on health worker experiences. During the experiment, poor and non-poor cases are referred to with neutral labels "Type X" and "Type Y", respectively.

Our exit survey of the 1,113 health care students and 1,029 medical professionals who participated in the wider study further confirm the salience of these difficulties in treating poor patients in Burkina Faso.[14] We asked the participants:

(i) "Reflecting on why it is more difficult for health clinics to treat poor patients than non-poor patients, how important it is that the poor can have more complex health problems than non-poor patients?"

---

[13] Videos of the cases can be seen here: http://www.rbfhealth.org/resource/video-vignettes-lab-field-experiment-burkina-faso

[14] The results reported in this paper are for the subset of these subjects who participated in the four treatments that focus on the effect of quantity-based bonuses on the allocation of effort by health workers. We also conducted treatments focusing on the impact of quality-based bonuses, not reported here. Experiments with student samples focused on the impact of medical treatment-based profits on accuracy and on disentangling the effects of task and mission motivation. These results are in the process of being reported in separate articles (see Banuri et al. 2018; and Banuri, Keefer, and de Walque, 2018 as examples).

(ii) "Reflecting on why it is more difficult for health clinics to treat poor patients than non-poor patients, how important it is that the poor patients are more difficult to understand?"
(iii) "Reflecting on why it is more difficult for health clinics to treat poor patients than non-poor patients, how important it is that it takes longer for health clinics to treat poor patients than non-poor patients?"

Figures 3a, 3b, and 3c display the answers to each of these questions.[15] An overwhelming majority of both students and professionals agree that there are disincentives involved in treating poor patients: time, complexity, and communication. Policies meant to increase equity in public service delivery do not, however, regularly distinguish between the differences in difficulty of delivery to the poor and non-poor.

[15] Extensive interviews and discussions with a small set of healthcare workers prior to commencing the lab-in-the-field study had already indicated that the poor had more complex health problems, were more difficult to understand and that it took longer to treat them. We therefore included this information in the questions to the health workers who participated in the experiment in order to focus on whether they agreed that the three characteristics were important obstacles to treating poor patients. Those who had not previously thought of the poor in these terms should also be less likely to indicate that the characteristics, when told of them, are important obstacles. Those who are aware of these characteristics, but believe that they are not important obstacles to health care, should also, upon being reminded of the characteristics, not be more likely to say that they are important obstacles. Nevertheless, the responses are upwardly biased, towards finding that the characteristics are important obstacles, to the extent that some respondents, upon being told that the poor have these characteristics, also conclude that they are important.

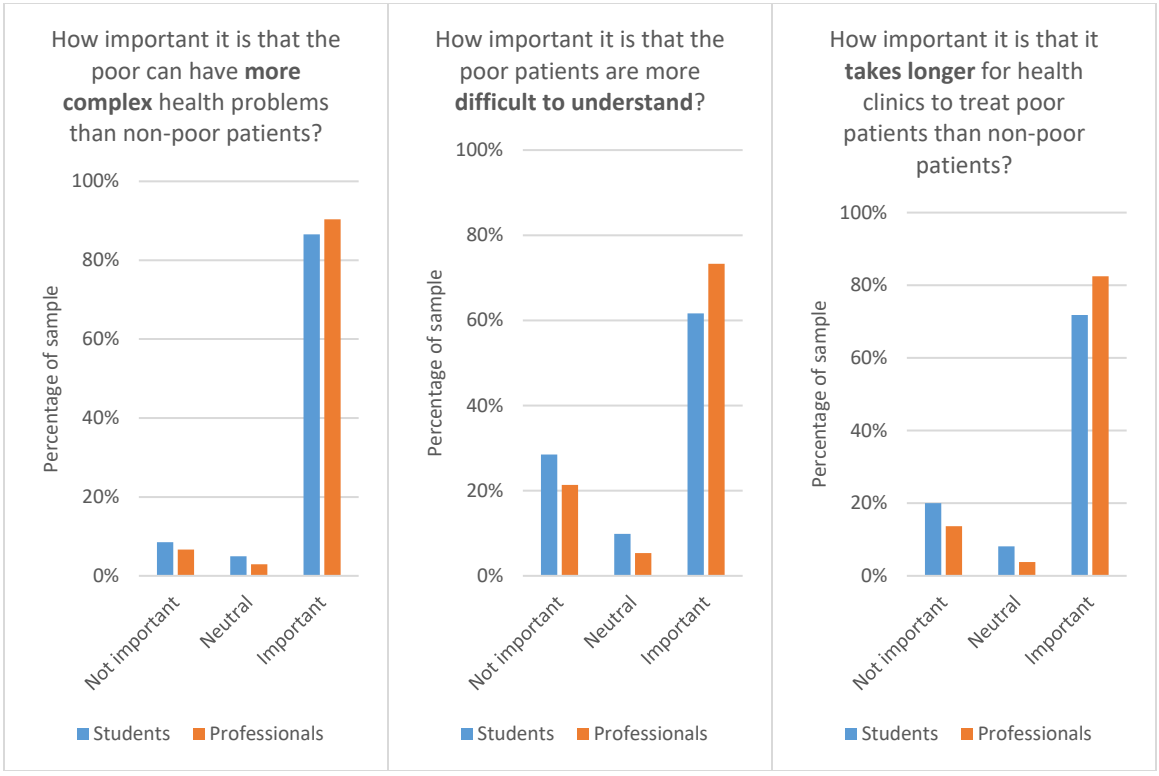


Figure 3 (a, b, c): The role of complexity (a); communication (b); and time (c) in serving the poor

It is not enough for the nature of the task to differ between poor and non-poor patients; it is also essential to account for how health workers internalize the welfare effects of their decisions on poor and non-poor patients. To capture those effects in the experiment, we implemented a link between effort on behalf of poor patients and the provision of benefits to actual poor people; and effort on behalf of non-poor patients and the provision of benefits to actual non-poor people. Accurate diagnoses and treatment of poor cases generated benefits for the poor: donations to a poor school. Similarly, accurate diagnoses and treatment of non-poor cases generated benefits for the non-poor: donations to a non-poor school.[16]

Specifically, the instructions informed subjects that correct responses to the questions in the medical cases would generate donations to two schools in the capital city, Ouagadougou. This reduces noise by holding constant key demographic variables (beneficiaries were all children, and all

[16] Ideally, effort in reviewing poor cases would have yielded donations to a health clinic that served the poor, and effort in reviewing non-poor cases to a health clinic that served the non-poor. This turned out not to be possible since we could not find a health clinic that specifically served the non-poor. However, the mismatch between actual beneficiaries of the donations (schoolchildren) and experimental beneficiaries (patients) is only problematic if school beneficiaries exaggerate the effects of intrinsic motivation to serve the poor compared to health clinic beneficiaries. This might happen, for example, if health care professionals had exhibited little variation in the share of donations that they might have given to poor clinics (as opposed to poor schools), preventing any identification of the effects of intrinsic motivation to serve the poor on effort. There is no behavioral or empirical reason to think that primary school beneficiaries would yield such a bias, however.

located in Ouagadougou) and context (all the children are in primary school). To make the poor and non-poor differences more salient to subjects, we showed them pictures of classrooms in the poor and non-poor schools (figures 4a and 4b). Subjects saw pictures of the poor and non-poor classrooms, both either taken from the front or the rear of the classroom (randomly assigned and from the same angle), and always capturing a similar number of children.[17]

To sum up, we simulated poverty in the experiment by implementing four differences between poor and non-poor cases: (i) Poor cases were more complex (complexity disincentive to serve the poor); (ii) they were more difficult to understand (communication disincentive to serve the poor); (iii) they took more time (time disincentive to serve the poor); and (iv) serving the poor generated benefits for members of poor groups in the real world (social preference incentive to serve the poor).[18]

Figure 4 (a, b): Non-poor (left) and Poor (Right) schools

The "Virtual clinic" task

The experiment put health workers in a virtual clinic where they needed to spend a single (virtual) day at work. They were asked to engage in the virtual clinic task for 11 minutes and were compensated depending on the treatment.[19] Figure 5 presents the virtual clinic flowchart, and figure 6 displays the virtual clinic case menu that health workers saw upon beginning the task. Health workers were shown a total of 16 cases on the menu, 8 poor and 8 non-poor, of which they were

[17] Besides the photographs, subjects were given some additional information about the two schools: school fees (2,000 CFA per year versus 300,000 CFA per year); class size (57 students versus 25); percentage of students who passed the primary school-leaving exam (CEP) in 2013 (92 versus 100 percent); and what the schools indicated they would do with the additional funds (health; rehabilitating and equipping classrooms; and rehabilitation of the water well versus increasing modernizing classroom equipment - introducing more technology - and improving sports facilities and equipment).

[18] Note that, although subjects received ample information that allowed them to distinguish poor and non-poor cases, at no point in the experiment were the labels "poor" and "non-poor" utilized. The schools were given neutral labels (school A and school B), as were the cases (case type X and case type Y).

[19] The choice of 11 minutes was so that subjects would not have enough time to review all 16 cases that were available to them, as cases lasted either 60 or 100 seconds, yielding a minimum amount of time needed to view all videos in their entirety to be 21.33 minutes. We could have selected any amount of time less that 21.33 minutes, though we chose 11 minutes do as to generate sufficient variation in our effort measure.

free to review any case in any order than they preferred.[20] The case menu organized the cases into poor and non-poor columns, labeled Type Y and Type X.[21]

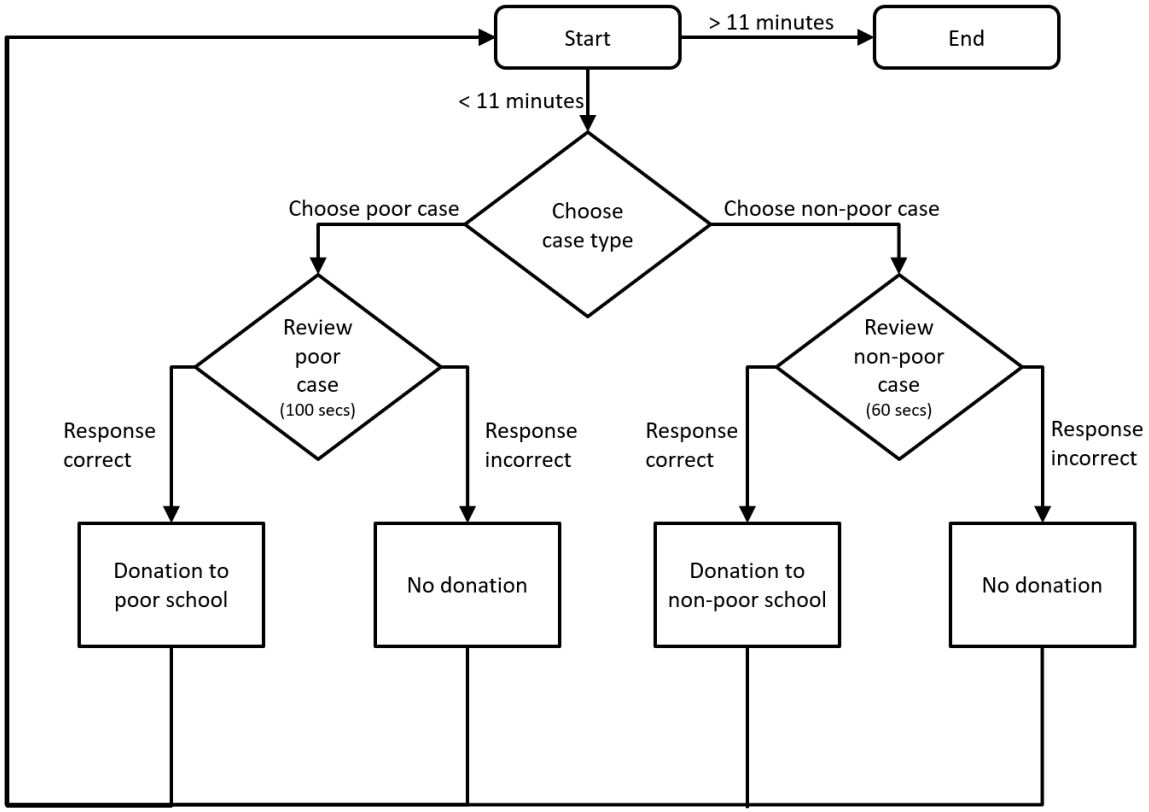


Figure 5: Virtual clinic process flowchart

Notes: Participants engage in the virtual clinic task for a total of 11 minutes, in which time they choose cases to review. Note that the time needed to review cases varies by case type.

Upon selecting a case, the video would play, and the questions would be displayed at the bottom (see figure 7 for a screenshot of the case screen).[22] Health workers had to answer all four questions for each case before being returned to the case menu and selecting their next case, each

[20] One important point to note here is that we generated poor and non-poor versions of all 20 cases in the casebank. Hence, at the start of each individual session, the computer would randomly select a poor or non-poor version of each case with 50% probability, with no case replacement. Hence, each subject could face different versions of the 16 cases used for the virtual clinic task. By doing this, we ensure that there are random distributions of case types in each treatment.

[21] The order of the columns was randomized such that poor cases either appeared on the left or the right side of the screen.

[22] It is important to note that watching the entire video was not necessary: participants had full control of the video, could stop it, pause it, rewind it, or forward through it. We do not find any evidence that they did so, however. All participants saw less than the maximum number of cases available, indicating that they viewed the entire video and made every attempt to diagnose and treat the case accurately.

time choosing between type X and type Y cases.[23] Health workers were not forced to watch the entire video and could answer questions at any time. This means that workers could simply skip through the videos and just respond to the questions based on information provided on screen, though this would not have been enough to diagnose and treat patients accurately. We find no evidence of such behavior; health workers typically watched the entire video prior to responding to the questions. We did not record the time spent on each individual case.

Treatments

The four incentive treatments are layered over this basic structure.

(i) "Salary" contract: The health worker receives a flat salary of 4,000 CFA (8.32 USD). This salary is completely unconditional.

(ii) "Non-poor bonus" contract: This corresponds to the *status quo* in rural health clinics in Burkina Faso and most other places. The health worker is paid a flat salary of 4,000 CFA (8.32 USD) and an additional 100 CFA (0.21 USD) bonus for each non-poor case seen. The bonus simulates consultation fees, which the poor usually cannot afford.

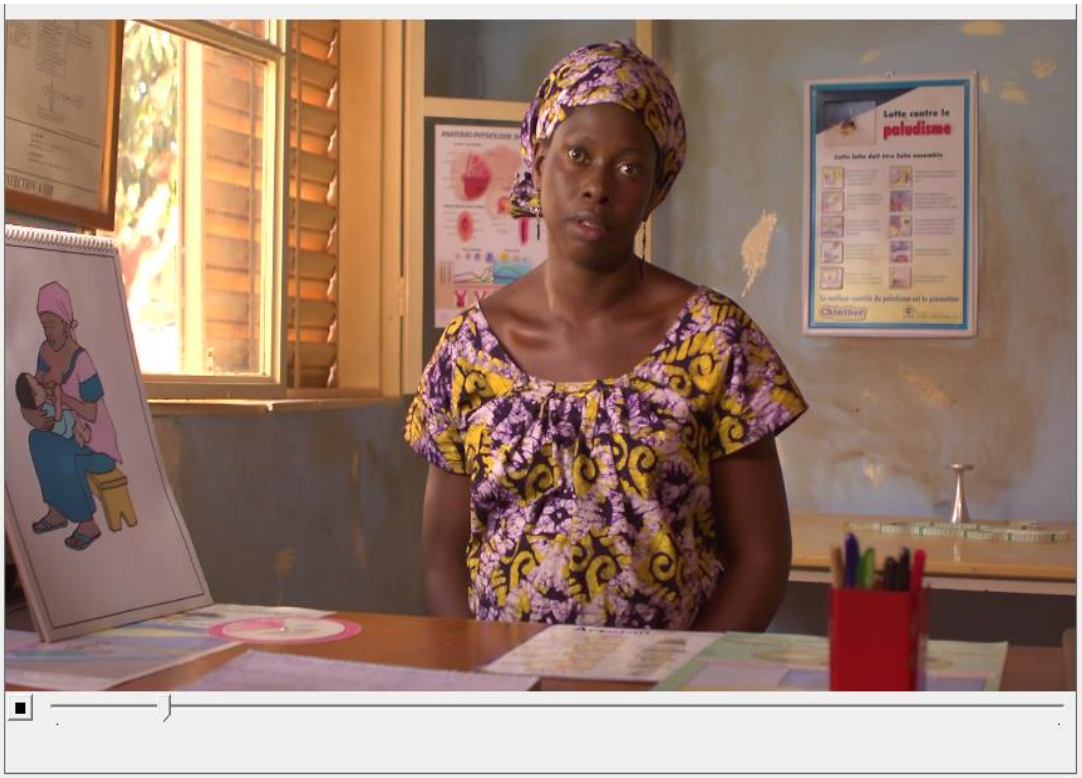
(iii) "Equal bonus" contract: The third treatment tests a policy where the state pays all consultation fees for the poor, so the health worker gets (in addition to their salary of 4,000 CFA) a 100 CFA bonus for treating a poor case, and a 100 CFA bonus for seeing a non-poor case.

(iv) "Poor bonus" contract: The final treatment simulates a condition where the government provides additional compensation for poor cases in order to combat the disincentives inherent in treating the poor. Health workers get a 100 CFA bonus to treat non-poor cases, and a 200 CFA bonus to treat a poor case.

Across all treatments, however, donations to the schools (for accurately treating cases) remain identical, such that the only factor changing is the pecuniary incentive for the health workers.

[23] Subjects knew that type Y cases lasted 40 seconds longer and were more complex than type X cases, and that that correct responses to type X cases benefitted the non-poor school, while correct responses to type Y cases benefitted the poor school. The case menu reminded them of these details.

Figure 6: Virtual Clinic case menu



Signes vitaux:

- Température : 37°6 C
- Tension artérielle: 110/80 mmhg
- Conjonctive : coloration normale
- Rythme cardiaque foetal : 130 battements/mn
- Absence d'oedèmes des membres inférieurs

Signes additionnels:

- L'examen gynécologique révèle des pertes évidentes mélangées avec du sang et une dilatation du col de l'utérus à 1 cm.
- Pas de signes objectifs d'infection vaginale
- Femme primipare

1) Quel est le diagnostic le plus probable?

A) Rupture utérine	B) Rupture prématurée des membranes	C) Eclampsie	D) Accouchement prématuré	E) Candidose vaginale
Répondre (A)	Répondre (B)	Répondre (C)	Répondre (D)	Répondre (E)

Figure 7: Non-poor case sample

Implementation

The overall structure of the main experiment is displayed in Figure 8. After health workers are given instructions and introduced to the two schools, we obtain incentivized measures of intrinsic motivation of serve the poor, and an incentivized measure of medical ability.[24] These are explained in the next section. Health workers then enter the virtual clinic. Finally, an exit survey records health worker demographics and other preferences.

The sessions were conducted in February-March 2014. The sample of health workers was drawn from participants in mandatory week-long training workshops conducted by the Ministry of Health in 5 regions of Northern Burkina Faso (Gourcy, Kaya, Koudougou, Nouna, and Ouahigouya). All rural health centers were required to send a single staff member to attend training. Hence, we have one worker per health center in our sample. Attendees are compensated for attending, providing strong incentives for health centers to send their representatives. The health centers in rural Burkina Faso typically have between 2-4 workers, usually nurses, or workers trained in providing basic health (see table A1 for a breakdown of worker qualifications). Our team invited the workers to participate in the experiment on random days during the workshop (beginning at 11am). Health workers could participate in the experiment at any time during the day for which they were invited. Hence, our results reported errors clustered by day rather than by session, since there were no sessions, *per se*. The data are fully anonymized to protect against potential bias, since workers may be concerned about responding to a study during an event conducted by the Ministry of Health.

Health workers earned an average of 6,215 CFA on average (approximately 12 USD, which was broadly in line with the per diem they received from attendance) from the session. The donations for the schools from both the dictator games and the virtual clinic task averaged 1,682 CFA per worker for the non-poor school (approximately 3.3 USD) and 1,440 CFA per worker for the poor school (approximately 2.9 USD).

[24] Note that our motivation measure precedes the virtual clinic task, and both are compensated. This raises concerns about the independence of these measures, in that the decision to donate may influence performance in the virtual clinic task. Two features mitigate this concern, however. First, since the motivation measure is identical across treatments, it is not clear whether the treatment effects we estimate could be influenced in different ways so as to bias our results. Second, the amount that is possible to donate for each school is fairly small (1,250 CFA) relative to the donation possible through the virtual clinic task (3,200). Nevertheless, it is still possible that those that donate more in the donation task might exert more effort in the clinic task for the sake of consistency, or less effort in the clinic task for the sake of satisficing. The treatment effects should be independent of these considerations, however.

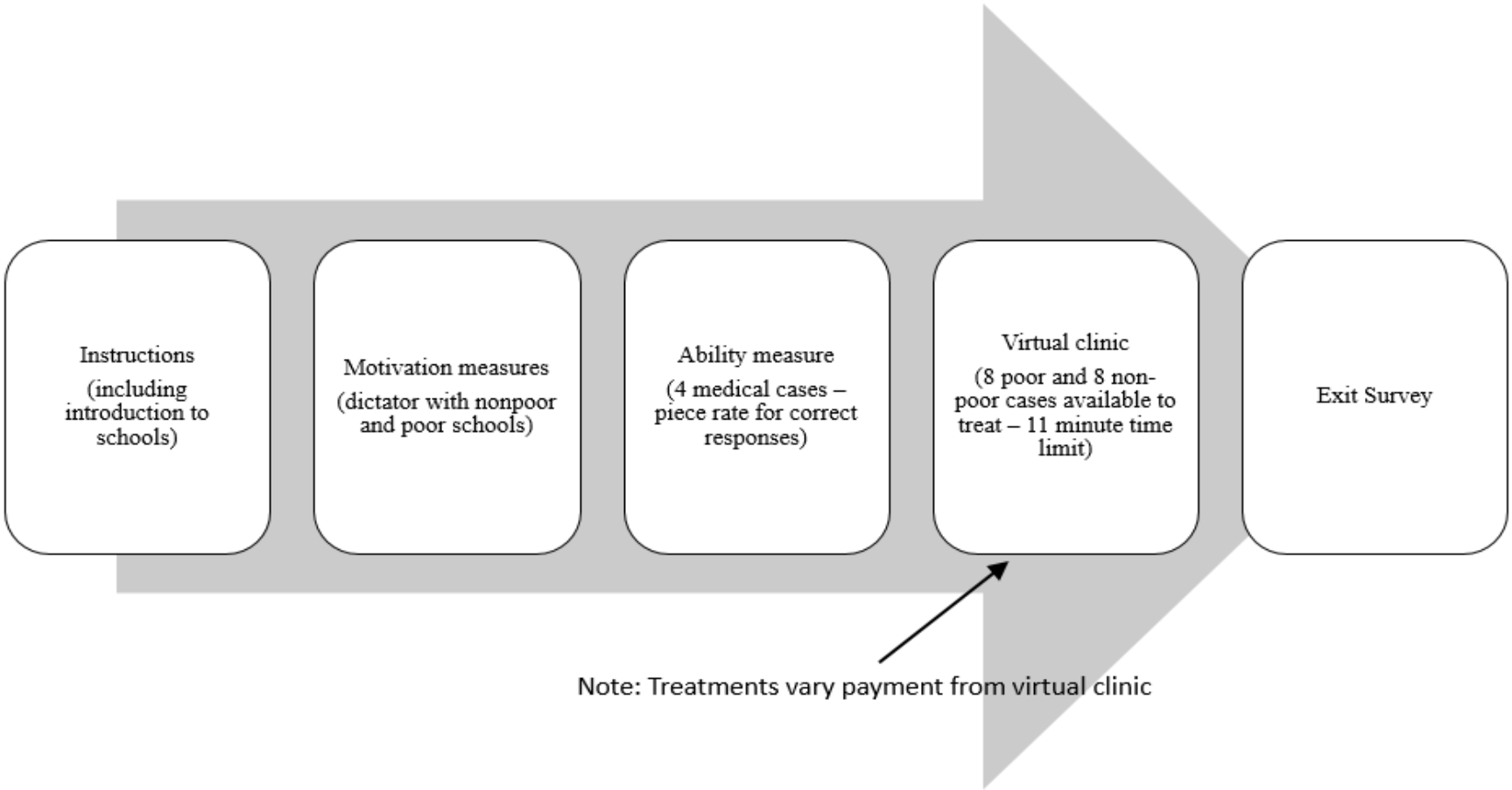


Figure 8: Main experiment overview

Variable Measurement

Intrinsic motivation to serve the poor

The key measure of intrinsic motivation to serve the poor comes from a modified version of the dictator "game". Subjects play two simultaneous dictator games, one with a non-poor primary school as beneficiary and one with a poor primary school as the beneficiary. These are the same non-poor and poor schools pictured in Figure 4 above. The games were always conducted prior to the virtual clinic task (as shown in figure 8).[25]

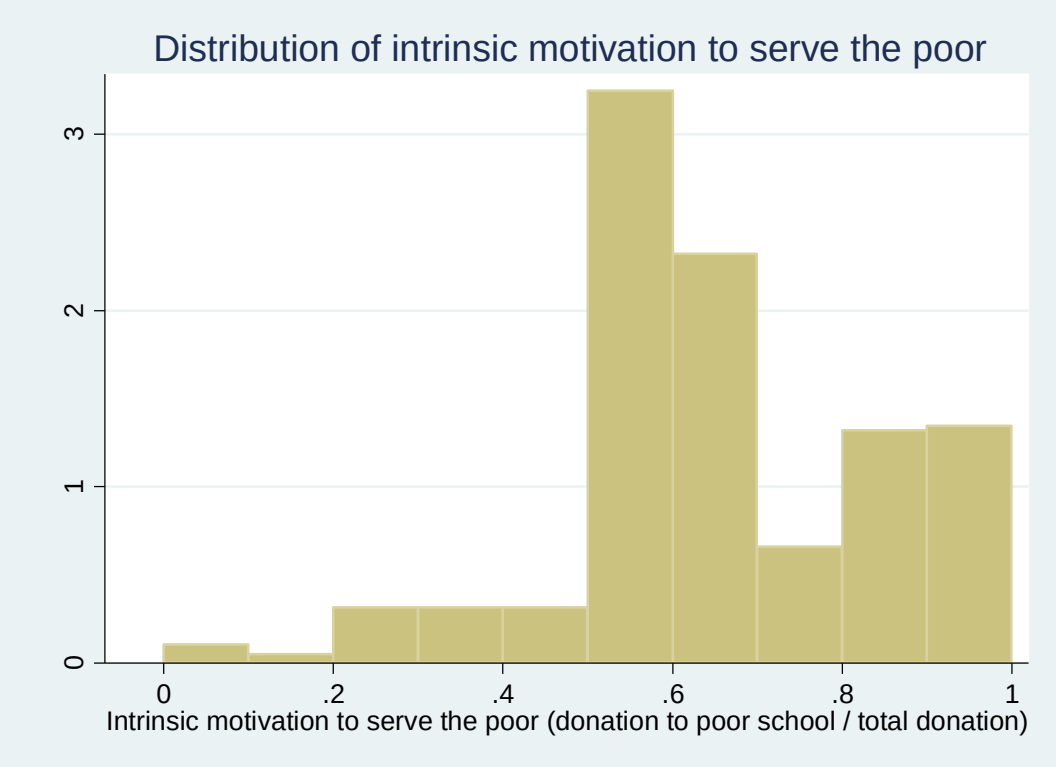


Figure 9: Distribution of Intrinsic motivation to serve the poor

Prior to playing the dictator games, health workers were informed about the size of the schools and the socioeconomic characteristics of their respective student bodies. Health workers were then given an endowment of 1,250 CFA Francs ($2.60) for each school and told that they could donate as much of the endowment as they wanted to each primary school, in any proportion that they wanted, keeping the remainder of the funds. As in Figure 4, half of the health workers,

[25] One issue may be that individuals that donate more to the school are less likely to treat poor patients in the virtual clinic task, reasoning that they have already acted generously (in line with moral licensing type behavior – see Brañas-Garza et al., 2013). We note, however, that the dictator donations were fairly small: each case could potentially generate 400 CFA in donations (maximum of 6,400 CFA for the virtual clinic task), while the maximum donation amount in the dictator game was 2500 CFA. There is no evidence of moral licensing behavior, since motivated workers treating far more poor patients on average (see figure 12).

randomly chosen, saw pictures of the poor school on the left side of their screen and the other half saw pictures of the poor school on the right.

The measure of intrinsic motivation to serve the poor is the proportion of the total donation that was allocated to the poor school. This measure carefully distinguishes prosocial preferences in general from intrinsic motivation to serve the poor: health workers who give larger total amounts to both schools, but a smaller proportion to the poor school, are less intrinsically motivated to serve the poor than health workers who give smaller amounts in total, but dedicate a larger proportion to the poor school.[26] In the results, "motivated workers" are those who gave more that 50% of their total endowment to the poor school (nearly two-thirds of our sample) and "unmotivated workers" are those who gave 50% or less. Figure 9 presents the distribution of the intrinsic motivation measure.

Health worker ability

Health worker ability can have a substantial influence on performance and may be correlated with intrinsic motivation. We control for ability in the estimates below using a measure derived from an early round of the experiment. Subjects reviewed four cases prior to the virtual clinic task and were paid 100 CFA - $0.21 – for each correct response. Unlike the virtual clinic, these cases were not time constrained: subjects spent 21.23 minutes (5.31 minutes per case) on average. Subjects had an accuracy rate of 48 percent and earned 765 CFA – $1.59 – on average. No feedback on their accuracy was provided until the end of the session (to prevent learning). Effort in this phase benefited the health workers alone; no donations to the schools were generated.

We construct two base measures of ability using the data from this component of the experiment: the time that health workers took; and their score – the number of questions, out of 16 (four questions times four cases), that they correctly answered.

By construction, poor cases took longer and were more complicated than non-poor cases. For this reason, two of the cases used in this component were non-poor and two were poor cases. We therefore construct two further ability measures that reflect relative ability in reviewing poor cases:

1. The "effort cost" ratio: this is the number of points that health workers scored in the two poor cases divided by their total point score. The average effort cost ratio was approximately 0.55.
2. The "time cost" ratio is the amount of time health workers spent on the two poor cases divided by the total time they spent on all four cases.

Because of the order of cases, the measures of ability to treat the poor are upwardly biased, but the bias is uniform across treatments. In the ability phase, health workers were presented with the four cases in the same order: non-poor, poor, non-poor, poor. Learning clearly took place: the first non-poor case took health workers an average of 383 seconds, compared to 267 seconds for the second non-poor case. The poor cases, with video vignettes that were 40 seconds longer, exhibited less improvement: 334 seconds for the first poor case versus 302 seconds for the second. Health

[26] These same schools were used as beneficiaries of effort during the virtual clinic phase of the experiment. However, at the time of the donation decision, health workers were not aware that the schools would be beneficiaries of effort later on in the experiment.

workers scored 1.41 correct answers on the first non-poor case and 2.04 on the second; 1.99 on the first poor case and 2.28 on the second, accounting for the average effort cost ratio of 0.55. Learning effects are distributed randomly across treatments and so do not bias the estimated effects of ability on effort.

Results

In the virtual clinic, health workers were given 11 minutes to complete as many cases as they could. They were provided with two types of cases, *non-poor* (type X in the instructions) or *poor* (type Y). Health workers were informed that type Y (poor) cases would contain longer videos and be more complex. They were free to choose any case in any order they preferred. Our main dependent variable is simple: the number of poor cases reviewed by the subject divided by the total cases reviewed, yielding a proportional measure. We ask: What is the impact of the different contracts on the proportion of poor cases reviewed? What is the effect of intrinsic motivation to serve the poor on the proportion of poor cases reviewed under each contract? And do motivated or unmotivated workers account for the shifts in behavior that we observe between contracts?

Effect of compensation schemes on services to poor patients

We first investigate how the different contracts affect services provided to the poor. There are two key results. The first is intuitive: the share of poor cases reviewed by health care workers is highest under the Poor bonus contract, followed by the Salary contract, and then the Equal-bonus contract (which effectively provides higher pecuniary rewards to reviewing easier, non-poor cases) and finally Non-poor bonus contract, offering the highest pecuniary rewards to reviewing non-poor cases. Second, however, the Poor bonus and the Salary contracts generate nearly the same levels of service to the poor. That is, results from contracts with bonuses are not symmetrical: moving from the Salary or Equal bonus contracts to the Poor bonus contract yields a relatively small increase in poor cases reviewed; moving from these contracts to a Non-poor bonus (*status quo*) contract yields a relatively large decrease in poor cases reviewed.

Figure 10 displays the impact of the treatments on the proportion of poor cases reviewed out of all cases (left axis and line graph) and the total cases reviewed (right axis and bar graph).[27] Under the Salary contract, when no bonus was paid, 42 percent of cases reviewed were poor cases.

---

[27] Table A.1 in appendix A presents summary statistics across the treatments reported in the paper. We randomized subjects to treatments within sessions using a between-subjects design (upon startup, the software randomly selected a treatment). Between treatment differences were calculated using a joint F-test (p-value reported in the final column in Table A.1). Overall, there are no significant differences across treatments in the reported variables. Across all treatments, subjects were in their mid-30s, on average. There were more women than men in all the treatments. However, though balanced on average, significantly more women participated in the equal bonus treatment than in the poor-bonus treatment. Subjects reported similar incomes on average, though subjects in the non-poor bonus treatment reported significantly higher incomes. Subjects in all treatments were similar in the amounts donated to either the poor or the non-poor schools. Health workers displayed similar average ability except in the salary treatment, where subjects scored half a point more than in the other treatments. However, relative ability on poor and non-poor cases, as measured by the effort cost and time cost measures, were indistinguishable across treatments. The remaining variables, except for one, were entirely balanced: those in the poor bonus treatment found the instructions significantly clearer than those in the non-poor bonus treatment. Results are entirely unaffected by controls for those variables that demonstrate imbalance.

This was less than 50 percent ($p<0.10$),[28] but significantly more than one might have expected if intrinsic motives to serve the poor were low. Under the Poor bonus scheme, the proportion of poor cases reviewed, 49 percent, was almost identical to the Salary contract ($p=0.216$).

Both the Salary and the Poor bonus contracts had a significantly higher proportion of poor cases reviewed than under the Non-poor bonus contract. (Under the Non-poor bonus contract, poor cases comprised 21 percent of all cases reviewed, significantly lower than the Salary and Poor bonus treatments ($p<0.01$ in both cases).[29] The Poor bonus contract achieves nearly perfect equity in the proportion of poor cases reviewed: 49 percent, statistically indistinguishable from 50 percent ($p=0.83$). Because of the higher costs of treating poor patients, the Equal bonus still offers higher pecuniary returns to reviewing non-poor cases. Consistent with this, the share of poor cases under the Equal bonus contract lies between the share under the Non-poor and Poor bonus contracts ($p<0.01$ and $p<0.10$ respectively).

The total number of cases seen under each contract follows an intuitive pattern. Since non-poor cases take significantly less time, total cases under the Non-poor bonus contract, 5.47, are higher than under the Salary contract, 4.91 ($p<0.05$). In the presence of an Equal bonus for both poor and non-poor cases, participants reviewed 4.99 cases on average, not significantly different from either the Salary contract ($p=0.79$), or the Non-poor bonus treatment ($p=0.14$). The Poor bonus also increases the number of cases reviewed to 5.39; this is significantly higher than the Salary contract ($p<0.10$), but not significantly different from the other bonus contracts ($p=0.77$ and $p=0.22$ for the Non-poor and Equal bonus treatments respectively).[30]

---

[28] Unless otherwise stated, p-values are based on two-tailed t-tests throughout.

[29] We show below that the fact that the percentage reaches 21 percent, significantly greater than zero, is due to the choices of health workers with stronger intrinsic incentives to serve the poor.

[30] It might seem surprising that, although the videos for poor cases are longer and the proportion of poor cases reviewed is significantly greater, the total number of cases reviewed remains largely the same in the Non-poor bonus (5.47) and Poor bonus (5.39) treatments. Accounting for this fact, on average, individuals who reviewed poor cases were of higher ability than those who reviewed non-poor cases. Low ability workers reviewed 0.67 fewer cases in the Poor bonus treatment compared to the Non-poor bonus treatment, while high ability workers reviewed 0.4 cases more. These choices lead to convergence in the total cases reviewed under the two contracts.

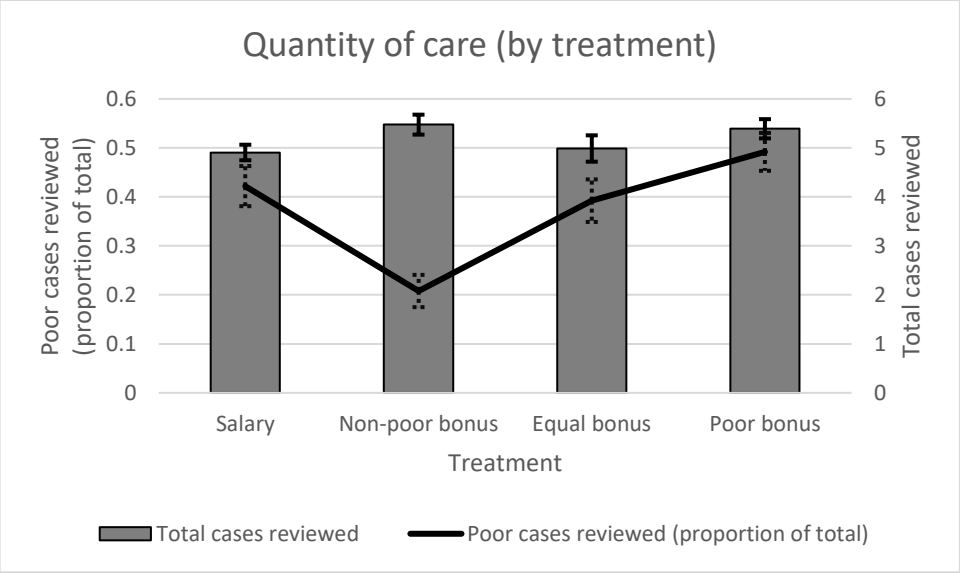


Figure 10: Cases reviewed, by type and treatment

The Poor bonus contract compensates subjects an extra 100 CFA for reviewing poor cases, double the payment for reviewing non-poor cases in the Equal contract. The Non-poor bonus contract reduces the payment for reviewing poor cases by 100 CFA (down to zero, consistent with the status quo, where the poor cannot pay the fees). The Poor bonus, though, increases the proportion of poor cases reviewed by 10 percentage points ($p=0.13$) while the Non-poor bonus reduces the proportion of poor cases reviewed by 18 percentage points ($p<0.05$).

Whose effort changes? The interaction of intrinsic motivation to serve the poor and pecuniary incentives

The results in the previous subsection demonstrate the novel effects of pecuniary incentives on services to the poor when workers can also choose to serve the non-poor. They do not, however, inform a central problem of behavioral research that is key to the recruitment decisions of organizations: is there a tradeoff between hiring motivated workers and providing high-powered incentives? In an organization with a single task that offers intrinsic motivation to some, the literature already answers in the affirmative. Prior research finds that motivated workers exert greater effort in tasks that offer intrinsic motivation than unmotivated workers, but high-powered incentives narrow the gap by encouraging effort by the unmotivated (e.g., Carpenter and Gong 2016).

The four panels of Figure 11 and the regression results in Table 1 illustrate the relationship between motivation and case choice under each of the four contracts. Since we are looking at the proportion of poor cases reviewed as a fraction of total cases reviewed, the relevant indicator of intrinsic motivation to serve the poor is our proportional measure from the dictator game: the fraction of donations to the two primary schools that went to the poor school.

The first panel of Figure 11 indicates that workers with more intrinsic motivation to serve the poor review a much larger proportion of poor cases under the Salary ($p<0.01$) and Poor bonus ($p<0.05$) contracts. However, the influence of intrinsic motivation dissipates under the Non-poor

bonus contract ($p=0.197$) and the Equal bonus contract ($p=0.733$), both of which offer greater pecuniary rewards for reviewing non-poor cases.

Table 1 presents a formal test of these relationships. Model 1 of the table adds the measure of intrinsic motivation. It has a modestly significant and positive effect on the proportion of total cases reviewed that were poor ($p<0.10$). Model 2 adds interaction terms with the treatment dummies (contracts) and intrinsic motivation measure. In the presence of the interaction terms, the linear intrinsic motivation coefficient therefore captures the effect of intrinsic incentives in the omitted salary treatment. The interaction terms capture the difference in effects of intrinsic incentives under the other contract treatments.

The coefficient on the linear motivation term in Model 2 is .58, meaning that compared to subjects who directed none of their donations to the poor school, the subjects who directed all their donations to the poor reviewed 58 percentage points more poor cases under the salary contract. The Non-poor bonus interaction is significant and negative (the negative coefficient on the interaction term is significant at $p<0.01$): the intrinsically motivated reduce the proportion of poor cases served under the Non-poor bonus scheme significantly more than the unmotivated. This coefficient is larger in magnitude than the Equal bonus interaction, which is also significant and negative ($p<0.05$), recalling that the Equal bonus effectively offers higher pecuniary compensation for reviewing non-poor cases because of the difficulty of viewing poor cases.

Models 3 and 4 show that the results are robust to numerous controls. In particular, Model 3 controls for ability in reviewing poor cases along two dimensions, accuracy (the ratio of scores on poor cases to total score) and speed (the ratio of time spent on poor cases to total time), using the incentivized (piece rate) ability score from the round that just preceded the virtual clinic task. It also controls for the professional qualifications of the health worker (generally categorized as nurse, midwife, doctor, or other).

These controls have little effect on the estimated coefficients in Model 2. However, ability matters: those who score higher while reviewing poor cases (relative to their total score) review a higher proportion of poor cases in the virtual clinic ($p<0.05$). The same is true in Model 4, which includes demographic controls (age, gender, monthly income, and current state of wealth); and experiment-specific controls (confidence that schools are paid in line with the instructions, and task motivation for the virtual clinic task). The small number of doctors in our sample review a significantly higher proportion of poor cases across all treatments.

In sum, within those contracts that provide pecuniary incentives to review more difficult, poor cases, the effect of intrinsic motivation to serve the poor is almost identical to its effect under the Salary contract: the Poor bonus does not narrow the gap between motivated and unmotivated workers in the share of poor cases reviewed, in contrast to what one might expect in a single-task setting. In contrast, within contracts that provide pecuniary incentives to serve the non-poor, the gap between motivated and unmotivated workers narrows, but only because the *motivated* review fewer poor cases, not because the *unmotivated* review more non-poor cases.

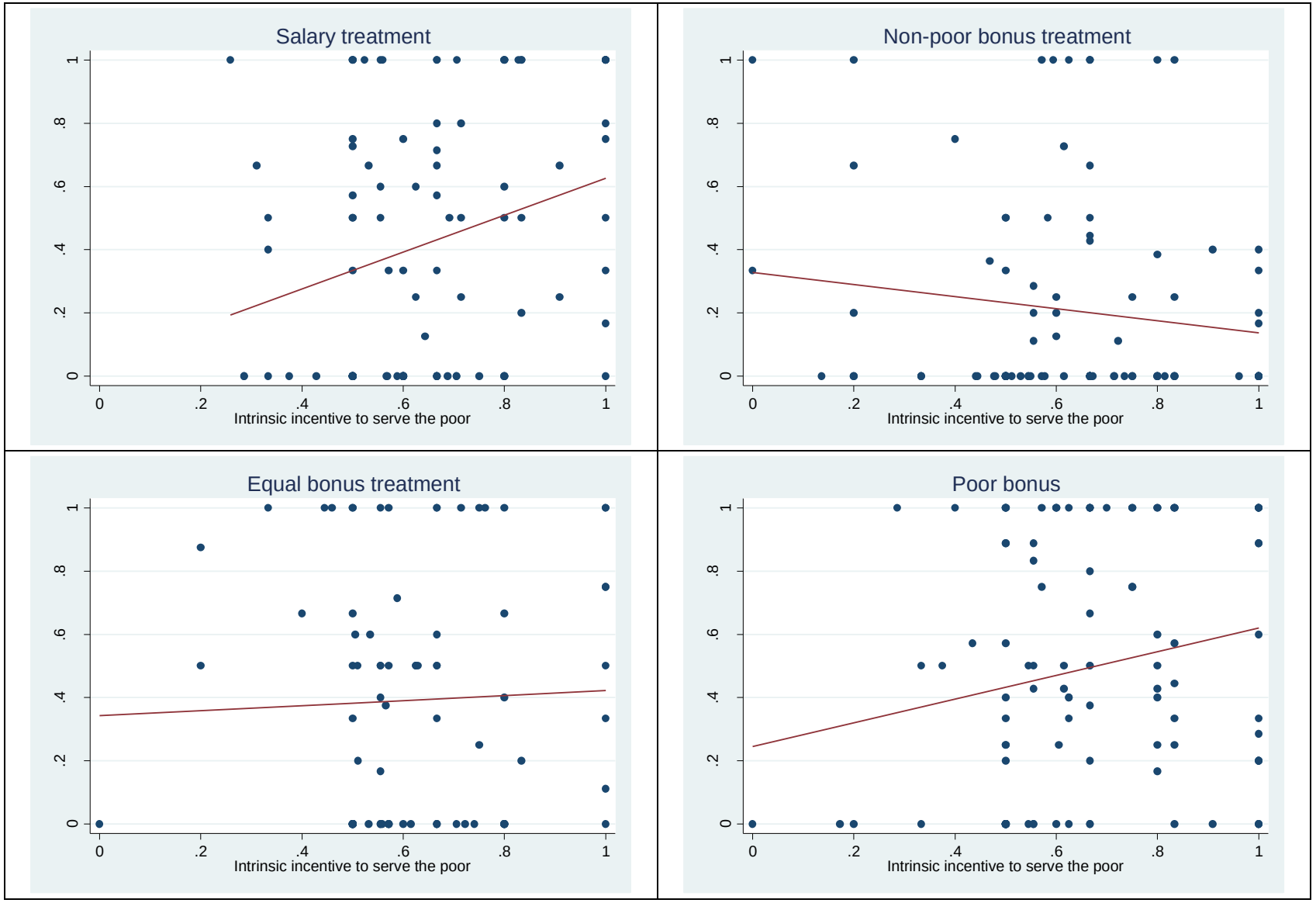


Figure 11(a-d): Intrinsic motivation and allocation of effort to the poor, by treatment

(Note: solid lines represent lines of linear fit)

Table 1: Treatment effects on poor cases as a proportion of total output

Dependent variable: Poor cases as a proportion of total cases				
	I	II	III	IV
Treatment: Non-poor bonus	-0.210***	0.286	0.294*	0.314
	(0.05)	(0.17)	(0.16)	(0.19)
Treatment: Equal bonus	-0.025	0.300*	0.281*	0.280
	(0.07)	(0.17)	(0.16)	(0.16)
Treatment: Poor bonus	0.068	0.203	0.198	0.236
	(0.05)	(0.18)	(0.19)	(0.21)
Intrinsic motivation	0.189*	0.584***	0.587***	0.619***
Donation to poor school / total donation	(0.09)	(0.16)	(0.15)	(0.17)
Interaction: Non-poor bonus X		-0.775***	-0.804***	-0.833***
Intrinsic motivation		(0.25)	(0.23)	(0.27)
Interaction: Equal bonus X		-0.505**	-0.486**	-0.488**
Intrinsic motivation		(0.23)	(0.22)	(0.22)
Interaction: Poor Bonus X		-0.210	-0.206	-0.274
Intrinsic motivation		(0.26)	(0.27)	(0.30)
Effort cost ratio (ability round)			0.245**	0.256**
Score with poor cases / Total score			(0.10)	(0.11)
Time cost ratio (ability round)			0.209	0.220
Time spent with poor cases / Total time spent				(0.22)
Occupation: Midwife			-0.004	0.013
			(0.04)	(0.05)
Occupation: Doctor			0.197***	0.176**
			(0.05)	(0.06)
Occupation: Other			0.018	0.041
			(0.05)	(0.05)
Female				-0.054
1 = Female				(0.04)
Age (years)				-0.002
				(0.00)
Income (in CFA)				0.000
				(0.00)
Constant	0.299***	0.042	-0.204	-0.175
	(0.09)	(0.12)	(0.13)	(0.23)
Pseudo R2	0.086	0.110	0.123	0.130
Log Likelihood	-165.7	-160.5	-157.7	-156.2
P	0.000	0.000	0.000	.
Observations	379	379	379	379

Notes: OLS regressions. The salary treatment is the omitted, baseline category. Dependent variable is the proportion of output that are poor cases (number of poor divided by total cases). * 10%, ** 5%, *** 1% significance level. Clustered standard errors (by day) in parentheses. The specification in column V also includes controls for current state of personal finances, confidence in payment to schools, and interest in task. None of these are significant. Results in

column I are robust to using overall generosity (total donation) instead of intrinsic motivation (share of total donations made to the poor school).

Intrinsic motivation and between-contract changes in worker behavior

Next, we split our sample into "motivated workers", those who gave a larger share of their donation to the poor school, and "unmotivated workers", those who gave an equal or a lower proportion. Approximately two-thirds of the sample fall into the motivated category (see figure 9).

Figure 12 presents treatment effects by motivated and unmotivated workers. Consistent with the importance of intrinsic motivation, under the salary contract, motivated workers review a higher proportion of poor cases: the difference is large (approximately 15 percentage points), although not statistically significant ($p=0.109$). Comparing the differences between the Salary and Non-poor Bonus columns, the proportion of poor cases seen by motivated workers falls by more than one-half or more than 20 percentage points ($p<0.01$), but by less than one-third among the unmotivated ($p=0.16$). Comparing the Salary and Poor Bonus columns, the motivated again react more strongly than the unmotivated, but their response is much less than in the Non-poor bonus case. The proportion of poor cases reviewed by the unmotivated rises 2 percentage points under the Poor bonus contracts compared to the Salary contract ($p=0.81$); the motivated review nearly 10 percentage points more ($p=0.15$), a far smaller increase than the decrease observed in the movement to the Non-poor bonus. These results reflect the sharper non-pecuniary tradeoffs that the more motivated confront when switching from non-poor to poor cases.

Figure 12 shows that, as with the Salary baseline, motivated workers again respond more to both the Poor and Non-poor bonuses. Now, however, the responses of the motivated are symmetrical across the two cases, with a 17 percentage point increase from the Non-poor to the Equal bonus contract ($p<0.01$), and another 17 percentage point increase from the Equal bonus to the Poor bonus contract ($p<0.01$). This is not the case for unmotivated workers, who display a 22 percentage point increase from the Non-poor bonus to the Equal bonus contract ($p<0.05$), but then show a decline of 6 percentage points from the Equal bonus to the Poor bonus contract, though this is not significant ($p=0.593$).

Past research examining performance on a single task indicates that increasing extrinsic incentives to serve poor patients would close the effort gap between the motivated and unmotivated. Instead, Figure 12 demonstrates the opposite: if anything, the gap between the motivated and unmotivated increases in the presence of pecuniary incentives to serve the poor under the Poor bonus contract ($p<0.01$).[31] Only bonuses to serve the less difficult cases, with fewer non-pecuniary benefits, closed the performance gap between the motivated and unmotivated, but of course in a downward direction. Both motivated and unmotivated workers review a lower proportion of poor cases under the Non-poor and Equal bonus contracts, but the differences between the two groups are both small and insignificant ($p=0.696$ and $p=0.843$ respectively).

[31] Motivated workers review 9.55 percentage points more poor cases in the Poor bonus contract, relative to the salary contract, though this difference is not statistically significant ($p=0.148$).

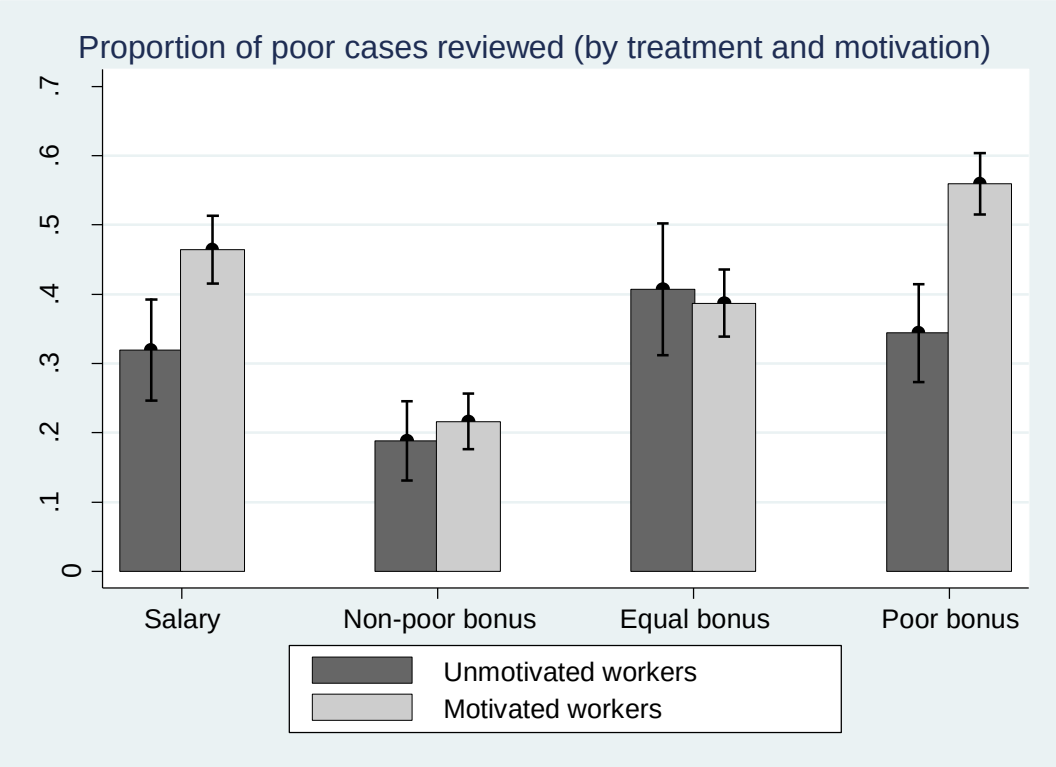


Figure 12: Poor cases reviewed, by treatment and motivation

Table 2 presents a formal test of the results observed in Figure 12. We run a model with just the treatment dummies, and with controls (identical to those used in table 2), on motivated workers (that directed a larger share of their donation to the poor school) and unmotivated workers (equal or lower donations to the poor school). Models 1 and 2 focus on motivated workers, while Models 3 and 4 focus on unmotivated workers. From the first two models, we immediately observe that, in contrast to earlier research, *motivated* workers respond to extrinsic incentives. Relative to the salary contract, they review a significantly lower proportion of poor cases (around 25 percentage points lower) under the Non-poor bonus contract ($p<0.01$). The effect of the Non-poor bonus on the unmotivated is much smaller (around 14 percentage points) and not statistically robust. The motivated respond much less strongly to the Poor bonus, but again the magnitude of their response is two to three times greater than the response of the unmotivated.[32]

Taken together, these results indicate the importance of compensating workers for the disincentives involved in treating poor patients, and to potentially recruit workers that are motivated to serve the poor.

Table 2: Motivated and unmotivated workers

Dependent variable: Poor cases as a proportion of total cases		
	Motivated workers	Unmotivated workers

[32] While the treatment effects are significant for motivated workers, and not for unmotivated workers, the experiment was not powered to detect whether the coefficients are different from each other. Indeed, the estimates for the effect of the equal bonus contract (relative to the salary contract) comes closest (p=0.103).

	I	II	III	IV
Treatment: Nonpoor bonus	-0.248***	-0.264***	-0.131*	-0.152
	(0.05)	(0.06)	(0.07)	(0.10)
Treatment: Equal bonus	-0.077	-0.073	0.087	0.072
	(0.07)	(0.07)	(0.11)	(0.12)
Treatment: Poor bonus	0.096	0.069	0.025	0.030
	(0.06)	(0.06)	(0.11)	(0.14)
Intrinsic motivation		0.189		-0.341
Donation to poor school / total donation		(0.16)		(0.39)
Effort cost ratio (ability round)		0.035		0.833**
Score with poor cases / Total score		(0.12)		(0.29)
Time cost ratio (ability round)		0.13		0.104
Time spent with poor cases / Total time spent		(0.24)		(0.42)
Occupation: Midwife		0.0532		-0.125
		(0.06)		(0.12)
Occupation: Doctor		0.219*		0.056
		(0.12)		(0.18)
Occupation: Other		0.005		0.103
		(0.06)		(0.10)
Female		-0.048		-0.085
1 = Female		(0.05)		(0.11)
Age (years)		-0.001		-0.007**
		(0.00)		(0.00)
Income (in CFA)		0.000		0.000
		(0.00)		(0.00)
Current state of personal finances		-0.022		-0.028
4 = Excellent		(0.05)		(0.06)
Confidence in payment to schools		0.006		0.005
5 = Strongly Agree		(0.01)		(0.03)
Interest in task		0.009		0.048
5 = Very interesting		(0.03)		(0.04)
Constant	0.464***	0.292	0.320***	0.090
	(0.04)	(0.32)	(0.06)	(0.41)
Pseudo R2	0.109	0.125	0.041	0.172
Log Likelihood	-112.6	-110.2	-49.66	-41.44
P	0.000	0.000	0.108	0.000
Observations	267	267	112	112

Notes: OLS regressions. The salary treatment is the omitted, baseline category. Dependent variable is the proportion of output that are poor cases (number of poor divided by total cases). The first two models are for motivated workers, while the last two are for unmotivated workers. * 10%, ** 5%, *** 1% significance level. Clustered standard errors (by day) in parentheses. Motivated workers are defined as those that directed a higher proportion of their donation to the poor school, while unmotivated workers are the rest. Results are robust to including overall generosity (total donations) instead of the intrinsic motivation measure used.

Conclusion

Past research has suggested two solutions to the problem of under-treatment of poor patients: increase the role of intrinsic motivation, by hiring more health workers who are intrinsically motivated to serve the poor; or offer extrinsic incentives to treat poor patients. These solutions emerge from the analysis of single-task settings where workers have no discretion over task choice. Public policy has emphasized the second: responding to situations in which, *de jure* or *de facto*, the health sector offers greater pecuniary rewards to seeing non-poor patients, it has usually provided equal pecuniary rewards to treating either type of patient. We provide evidence bearing on two questions raised by past research and current public policy initiatives.

First, are pecuniary incentives effective in increasing service delivery to the poor? The question is relevant because, while the insights derived from the single-task setting have been fundamentally important, both theoretically and in practice, in many service delivery settings, including health, service providers have discretion. Variations across tasks in both difficulty and motivation should be relevant for worker behavior. Our experimental results indicate that they are. It is the *motivated* who are responsive to changes in pecuniary incentives. Moreover, whereas past work shows that pecuniary incentives to undertake tasks with non-pecuniary benefits significantly narrow the performance gaps between motivated and unmotivated workers, we do not find this to be the case. Motivated workers are sensitive to pecuniary incentives to undertake the less difficult task that has fewer non-pecuniary benefits. These incentives dramatically reduce attention to the poor precisely because of their significant impact on the motivated. We conjecture that the difference between our experiment and previous work is the presence of multiple tasks with varying non-pecuniary benefits, and that this (more complex) structure yields different outcomes than what is predicted by single task settings.

These findings point to important questions for future research. First, how might the ability of workers to choose among tasks with different levels of difficulty and intrinsic motivation affect pay and recruitment strategies in other contexts? For example, police have ample discretion to undertake such distinct tasks as engaging with the community (community policing) and apprehending potentially dangerous individuals suspected of committing violent crimes. Second, what are reliable methods for quantifying differences among tasks in their difficulty and intrinsic benefits? Both variables are key to the design of appropriate compensation schemes. Third, what are the differences in pecuniary compensation for services delivered to the non-poor and poor? Again, these are understood to exist, but not quantified with enough precision to fix provider compensation. Fourth, we find that intrinsic motivation to serve the poor and health worker ability play a substantial role in willingness to serve the poor and in their responses to pecuniary incentives. This underlines the need to understand the effort consequences of selection effects: how do the selection effects of contract type affect the services delivered to the poor?

References

Ashraf, N., Bandiera, O. and Jack, B.K., (2014). "No margin, no mission? A field experiment on incentives for public service delivery." *Journal of Public Economics* 120: 1-17.

Banuri, Sheheryar and Philip Keefer (2016). "Pro-social motivation, effort and the call to public service." *European Economic Review* 83: 139-164.

Banuri, Sheheryar, Damien de Walque and Philip Keefer (2018). " Love the job... or the patient? Task vs. mission-based motivations in healthcare." World Bank Policy Research Working Paper 8338, The World Bank, Washington, DC..

Banuri, Sheheryar, Haidara Ousmane Diadie, Paul Jacob Robyn, Maurice Ye (2018) "The use of video vignettes to measure health worker knowledge. Evidence from Burkina Faso." *Social Science and Medicine.* 213: 173-180.

Barfort, Sebastian, Nikolaj A. Harmon, Frederik Hjorth, and Asmus Leth Olsen. "Sustaining honesty in public service: The role of selection." *American Economic Journal: Economic Policy* 11, no. 4 (2019): 96-123.

Bindman, Andrew B., Kevin Grumbach, Dennis Osmond, Miriam Komaromy, Karen Vranizan, Nicole Lurie, John Billings, and Anita Stewart. "Preventable hospitalizations and access to health care." Jama 274, no. 4 (1995): 305-311.

Brañas-Garza, Pablo, Marisa Bucheli, María Paz Espinosa, and Teresa García-Muñoz. "Moral cleansing and moral licenses: experimental evidence." Economics & Philosophy 29, no. 2 (2013): 199-212.

Carpenter, Jeffrey, and Erick Gong. "Motivating Agents: How much does the mission matter?." Journal of Labor Economics 34, no. 1 (2016): 211-236.

Cox, James C., Ellen P. Green, and Heike Hennig-Schmidt. "Experimental and behavioral economics of healthcare." Journal of Economic Behavior and Organization 131 (2016): A1-A4.

Dal Bó, Ernesto, Frederico Finan, and Martin A. Rossi (2013). "Strengthening State Capabilities: The Role of Financial Incentives in the Call to Public Service." *The Quarterly Journal of Economics* 128(3): 169-1218.

Delfgaauw, Josse and Robert Dur (2008). "Incentives and Workers' Motivation in the Public Sector." *The Economic Journal* 118.525: 171-191.

Das, Jishnu and Jeffrey Hammer. 2007. "Location, Location, Location: Residence, Wealth, and the Quality of Medical Care in Delhi, India." *Health Affairs* (Millwood) 26(3): w338–51.

Das, Jishnu and Aakash Mohpal. 2016. Socioeconomic Status And Quality Of Care In Rural India: New Evidence From Provider And Household Surveys. *Health Affairs* (Millwood): 35(10) w1764–73

Das, Jishnu, and Thomas P. Sohnesen. 2007. "Variations in Doctor Effort: Evidence from Paraguay." *Health Affairs* (Millwood) 26(3): w324–37.

Deci, E. L., Koestner, R., and Ryan, R. M. (1999). "A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation." *Psychological Bulletin* (125): 627–668.

Ensor, Tim, and Pham Bich San (1996). "Access and payment for health care: the poor of Northern Vietnam." *The International journal of health planning and management* 11(1): 69-83.

Fink, Günther, Paul Jacob Robyn, Ali Sié, and Rainer Sauerborn. "Does health insurance improve health?: Evidence from a randomized community-based insurance rollout in rural Burkina Faso." Journal of health economics 32, no. 6 (2013): 1043-1056.

Frey, Bruno S., and Felix Oberholzer-Gee (1997). "The cost of price incentives: An empirical analysis of motivation crowding-out." *The American Economic Review* 87(4): 746-755.

Galizzi, Matteo M., and Daniel Wiesen (2018). "Behavioral experiments in health economics." In *Oxford Research Encyclopedia of Economics and Finance*. https://doi.org/10.1093/acrefore/9780190625979.013.244.

Green, Ellen P., 2014. Payment systems in the healthcare industry: an experimental study of physician incentives. *Journal of Economic Behavior and Organization.* 106 (0), 367e378. http://dx.doi.org/10.1016/j.jebo.2014.05.009.

Green, E., Peterson, K., Markiewicz, K., O'Brien, J. and Arring, N., 2017. The Impact of a Merit-Based Incentive Payment System on Quality of Healthcare: A Framed Field Experiment.

Gwatkin, Davidson R. 2005. How much would poor people gain from faster progress towards the Millennium Development Goals for health? *Lancet* 365: 813–17

Haelterman, Edwige, Rikke Qvist, Patricia Barlow, and Sophie Alexander (2003). "Social deprivation and poor access to care as risk factors for severe pre-eclampsia." *European Journal of Obstetrics & Gynecology and Reproductive Biology* 111(1): 25-32.

Hanna, Rema and Shing-Yi Wang (2017). "Dishonesty and Selection in Public Service: Evidence from India." *American Economic Journal: Economic Policy* 9(3): 262-90.

Heinig, M. Jane, et al. (2009). "Sources and acceptance of infant-feeding advice among low-income women." *Journal of Human Lactation* 25(2): 163-172.

Hennig-Schmidt, Heike, Reinhard Selten and Daniel Wiesen (2011). How payment systems affect physicians' provision behavior: an experimental investigation. *Journal of Health Economics* 30 (4):637e646. http://dx.doi.org/10.1016/j.jhealeco.2011.05.001.

Hennig-Schmidt, Heike and Daniel Wiesen (2014). "Other-regarding behavior and motivation in health care provision: an experiment with medical and non-medical students." *Social Science & Medicine*; 108(0) : 156e165. http://dx.doi.org/10.1016/j.socscimed.2014.03.001.

Ingersoll, Richard M (2004). *Why do high-poverty schools have difficulty staffing their classrooms with qualified teachers?* Washington, DC: Center for American Progress, Institute for America's Future.

Institut National de la Statistique et de la Démographie - INSD/Burkina Faso and ICF International. 2012. Enquête Démographique et de Santé et à Indicateurs Multiples du Burkina Faso 2010. Calverton, Maryland, USA: INSD and ICF International.

Jones, Daniel, Mirko Tonin and Michael Vlasspopoulos (2018). "Paying for what kind of performance? Performance pay and multitasking in mission-oriented jobs." Working paper.

Joseph, A.E., Phillips, D.R. (1984). *Accessibility and Utilization—Geographical Perspectives on Health Care Delivery*. New York: Harper & Row Publishers.

Ku, Leighton, and Sheetal Matani (2017). "Left out: immigrants' access to health care and insurance." *Health Affairs*.

Lagarde, Mylène and Diane Blaauw (2017). "Physicians' responses to financial and social incentives: A medically framed real effort experiment." *Social Science & Medicine* 147-59.

Lipsky, Michael (2010). *Street-level bureaucracy, 30th ann. Ed.: dilemmas of the individual in public service*. New York: Russell Sage Foundation.

Peters, David H., Anu Garg, Gerry Bloom, Damian G. Walker, William R. Brieger, and M. Hafizur Rahman (2008). "Poverty and access to health care in developing countries." *Annals of the New York Academy of Sciences* 1136(1): 161-171.

Raso, Giovanna, Jürg Utzinger, Kigbafori D. Silué, Mamadou Ouattara, Ahoua Yapi, Abale Toty, Barbara Matthys, Penelope Vounatsou, Marcel Tanner, and Eliézer K. N'goran (2005). "Disparities in parasitic infections, perceived ill health and access to health care among poorer and less poor schoolchildren of rural Côte d'Ivoire." *Tropical medicine & international health* 10(1): 42-57.

Ridde, Valéry, Gerald Leppert, Hervé Hien, Paul Jacob Robyn and Manuela De Allegri. 2018. International Journal for Equity in Health. 17:5. DOI 10.1186/s12939-017-0717-5

Robyn, Paul Jacob, Günther Fink, Ali Sié, and Rainer Sauerborn. "Health insurance and health-seeking behavior: evidence from a randomized community-based insurance rollout in rural Burkina Faso." Social science & medicine 75, no. 4 (2012): 595-603.

Sudore, Rebecca L., Kala M. Mehta, Eleanor M. Simonsick, Tamara B. Harris, Anne B. Newman, Suzanne Satterfield, Caterina Rosano et al. (2006). "Limited literacy in older people and disparities in health and healthcare access." *Journal of the American Geriatrics Society* 54(5): 770-776.

Wagstaff, Adam, Caryn Bredenkamp, and Leander R. Buisman (2014). "Progress on Global Health Goals: Are the Poor Being Left Behind?" *The World Bank Research Observer* 29:137-162.

Weissman, Joel S., Robert Stern, Stephen L. Fielding, and Arnold M. Epstein (1991). "Delayed access to health care: risk factors, reasons, and consequences." *Annals of internal medicine* 114 (4): 325-331.

Appendix A
Table A.1: Summary statistics for all treatments

	Salary		Non-poor bonus		Equal bonus		Poor bonus		Joint F-test
Observations	96		99		79		105		
Age (Years)	36.94	(8.22)	36.04	(5.82)	35.66	(5.37)	35.77	(5.51)	0.507
Female (=1)	0.63	(0.49)	0.63	(0.49)	0.71	(0.46)	0.56	(0.50)	0.247
Income (CFA)	145269.30	(121536.60)	188445.00	(360636.50)	128890.50	(89564.30)	134650.30	(56199.54)	0.161
Antipoverty preference (donation to poor school / total donation)	0.65	(0.19)	0.63	(0.22)	0.63	(0.19)	0.66	(0.21)	0.616
Number of non-poor cases observed	3.00	(2.33)	4.36	(2.32)	3.15	(2.37)	2.73	(2.19)	0.000
Number of poor cases observed	1.91	(1.97)	1.11	(1.82)	1.84	(2.03)	2.66	(2.37)	0.000
Proportion of poor cases observed (% of total)	0.42	(0.40)	0.21	(0.33)	0.39	(0.39)	0.49	(0.40)	0.000
Accuracy rate (%) in non-poor cases	0.40	(0.21)	0.44	(0.17)	0.41	(0.16)	0.45	(0.20)	0.256
Accuracy rate (%) in poor cases	0.45	(0.21)	0.43	(0.24)	0.42	(0.19)	0.43	(0.19)	0.833
Score in ability round (Max: 16)	8.07	(1.77)	7.57	(2.04)	7.49	(1.74)	7.48	(1.88)	0.088
Time taken in ability round (in seconds)	1275.93	(508.71)	1244.86	(470.69)	1280.39	(601.62)	1293.23	(649.01)	0.940
Effort cost ratio, poor/non-poor cases in ability round	0.56	(0.14)	0.57	(0.16)	0.54	(0.12)	0.55	(0.13)	0.658
Time cost ratio, poor/non-poor cases in ability round	0.49	(0.07)	0.50	(0.07)	0.50	(0.07)	0.50	(0.06)	0.615

Note: Standard deviations in parentheses.

Table A.1 (con't): Summary statistics

	Salary		Non-poor bonus		Equal bonus		Poor bonus		Joint F-test
Observations	96		99		79		105		
Current state of personal finances (4 = Excellent)	1.86	(0.61)	1.76	(0.59)	1.76	(0.56)	1.77	(0.67)	0.574
Interest in task (5 = Very interesting)	4.21	(0.48)	4.13	(0.62)	4.27	(0.59)	4.14	(0.75)	0.446
Confidence in payment to schools (5 = Strongly agree)	3.61	(1.28)	3.72	(1.25)	3.57	(1.32)	3.65	(1.16)	0.879
Schools are effective (5 = Strongly agree)	3.39	(1.28)	3.54	(1.21)	3.51	(1.39)	3.76	(1.10)	0.184
Clarity of instructions (5 = Always)	3.54	(1.18)	3.26	(1.18)	3.49	(1.10)	3.66	(1.21)	0.111
Questions about cases were fair? (5 = Very fair)	3.85	(0.62)	3.87	(0.55)	3.81	(0.60)	3.83	(0.60)	0.912
Are the cases common? (5 = Strongly agree)	3.48	(1.44)	3.58	(1.31)	3.46	(1.52)	3.70	(1.32)	0.590
Qualifications									
Nurses (=1)	0.54	(0.50)	0.56	(0.50)	0.43	(0.50)	0.53	(0.50)	0.350
Midwife (=1)	0.08	(0.28)	0.10	(0.30)	0.08	(0.27)	0.11	(0.32)	0.808
Doctor (=1)	0.00	0.00	0.03	(0.17)	0.04	(0.19)	0.02	(0.14)	0.312
Other (=1)	0.38	(0.49)	0.31	(0.47)	0.46	(0.50)	0.33	(0.47)	0.218

Note: Standard deviations in parentheses.