Financial Incentives are Counterproductive in Non-Profit Sectors: Evidence from a Health Experiment

Elise Huillery, Juliette Seban

Financial Incentives are Counterproductive in Non-Profit Sectors: Evidence from a Health Experiment *

Elise Huillery[†] and Juliette Seban[‡]

March 2, 2015

Abstract

Financial incentives for service providers are becoming a common strategy to improve service delivery. However, this strategy will only work if demand for the service responds as expected. Using a field experiment in the Democratic Republic of Congo, we show that introducing a performance-based financing mechanism in the health sector has counterproductive effects because demand is non-standard: despite reduced prices and eased access, demand for health decreased, child health deteriorated, workers' revenue dropped. Ironically, expected perverse effects of incentives on worker behavior were not realized: incentives led to more effort from health workers on rewarded activities without deterring effort on non-rewarded activities, nor inducing significant score manipulation or free-riding. We also find a decline in worker motivation following the removal of the incentives, below what it would have been in the absence of exposure to the incentives. Management tools used in for-profit sectors are thus inappropriate in non-profit sectors such as health where user and worker rationalities are specific.

JEL Codes: H51, I18, 012

1 Introduction

Long-standing concerns about the cost, accessibility and quality of health and education services have raised a growing interest in financial incentives for service providers, in particular health workers and teachers. It is a central idea in economics that incentives encourage effort and performance

*We thank Yann Algan, Elizabeth Beasley, Nicolas de Borman, Vera Chiodi, Damien De Walque, Gyuri Fritsche, Emeric Henry, Paul Jacob Robyn, Hadia Samaha, Mark Stabile and Christel Vermeesch for helpful comments and discussions, as well as Emmanuelle Auriol, Roland Benabou, Julia Cagé, Denis Cogneau, Marc Gurgand, Guy Laroque, Alice Mésnard, Pieter Serneels, Philippe de Vreyer and numerous seminar participants at the World Bank, EUDN 2013 scientific conference, Paris School of Economics, Toulouse School of Economics, Sciences Po, and the Blavatnik School of Government, Oxford University. We wish to acknowledge the support of the World Bank. Jean-Jacques Frère and Hadia Samaha led the implementation of this project and we are grateful to them for their leadership. We thank the Health Authorities of Haut-Katanga and the Health Sector Rehabilitation and Support project (Projet d'Appui à la Réhabilitation du Secteur de la Santé - PARSS) in Lubumbahi and Kinshasa for their collaboration. We are deeply indebted to Julie Berthet-Valdois for outstanding project management, and to the Ecole de Santé Publique de Lubumbashi for excellent data collection. We also thank Saidou Ouedraogo for excellent assistance with data entry. Finally, we gratefully acknowledge the staff and patients of the health facilities as well as the households for the time and information they gave. All errors remain our own, and the opinions expressed in this paper are ours alone and should not be attributed to the institutions with which we are associated, the World Bank or the Government of the Democratic Republic of Congo.

[†]Sciences Po, Department of Economics (elise.huillery@sciencespo.fr)

[‡]International Rescue Committee (seban.juliette@gmail.com)

in the context of a classic agency problem: to produce a desired outcome, a principal entity provides a reward conditional on the agent achieving a pre-determined performance. The reward should reinforce agents' willingness to achieve the rewarded actions (*motivational crowd-in effect*), and the broad empirical literature on incentives in for-profit organizations shows that it is often the case (Lazear, 2000; Bandiera et al. 2007; Bandiera et al. 2013). However, we lack empirical evidence that the success of financial incentives can extend to non-profit sectors where both workers and clients are specific in terms of motives and rationality. This paper shows that the general positive effects of financial incentives in for-profit sectors cannot be extended to non-profit sectors where demand can respond negatively to price reduction and be inelastic to worker effort to ease service access.

Performance-based financing is a mechanism by which health facilities are, at least partially, funded by the government on the basis of their production of a pre-determined output. It is a team incentive scheme where performance is measured at the facility level. Performance-based financing (PBF) models have been implemented in many developed and developing countries (see Figure 1 for the implementation of PBF in subsaharian Africa) using various performance criteria. Most models use volume of service provision (number of patients) for some pre-determined health services as their performance criteria, often combined with a measure of service quality. In this paper, we study a PBF model targeting an output -the number of patients, rather than an input (e.g. daily attendance or technical quality[1]). The system pushes health workers to develop appropriate strategies and invest in the appropriate inputs to increase the output. Many different obstacles can hinder the demand for health services: prices, information, service quality, or behavioral issues. Since local health workers should be in a better position than the central government to identify the relevant obstacles in a specific area, PBF is a contract that decentralizes the task of finding the appropriate strategies to increase health service uptake.

This paper makes several contributions to the literature on improving health service delivery. First, even though performance-based financing schemes have become very popular in the health sector in both developed and developing countries, the scientific evidence on their impact remains thin. This paper constitutes the first study using the random assignment of a large number of health areas to estimate the effects of a performance-based mechanism as a way to allocate governmental

[1]Two studies provide evidence that rewards contingent on a specific input (respectively attendance and service quality) do motivate health workers to provide more of this input (at least in the short run), but did not lead to any increase in health service utilization (the output) (Banerjee and Duflo, 2008; Peabody et al., 2011).

resources. Second, this paper provides a deeper understanding of the behavioral effects of financial incentives, using detailed data on worker responses, strategies and motivation, which is a novelty in the empirical literature on performance-based financing. We build on the psychological and theoretical literature to test the potential adverse behavioral effects of financial incentives: (i) that incentives may be negative motivational reinforcers (*motivational crowd-out effect*) (Lepper et al. (1973), Deci (1975), Deci and Ryan (1985), Benabou and Tirole 2003, Benabou and Tirole 2006, Gneezy et al. 2011); (ii) that agents may concentrate their effort on the actions attached to the reward at the expense of other actions that might be important in producing the ultimate output (*multitasking problem*) (Holmström and Milgrom 1991); (iii) that PBF may induce a reduction in effort due to free-riding problems since rewards are collective and not individual (*free-riding effect*) (Bandiera et al. (2013); (iv) and that incentivized agents may manipulate performance measures in order to obtain more of the reward. We show that in the context of the health sector in DRC, none of these adverse behavioral effects happen. Third, we provide evidence of a new mechanism that explains why financial incentives may be counterproductive in non-profit sectors: we show that demand for health services decreases with price and does not respond to eased access. Health worker strategies to increase demand, although sensible, failed or even backfired.

This paper uses data from an experiment conducted in the Haut-Katanga district of the Democratic Republic of Congo (DRC) between 2009 and 2013 to compare the effect of a PBF approach to that of a fixed payment approach. The 96 health areas of the Haut-Katanga district were randomly assigned to performance-based or fixed governemental payments, while ensuring that the same amount of resources was allocated to each group to neutralize any ressource effect. All of the 152 public, private or religious health facilities in these health areas participated in the experiment. In this study, the PBF mechanism was based on relative incentives: a number of points was attributed to each facility based on the number of patients for some pre-determined services at this facility relative to all other PBF-group facilities. Unannounced visits to the facilities were performed in July, August and September 2012 in order to measure worker attendance, and a final independent survey was administered a few months after the payments had been withdrawn to collect data on (i) the supply and price of health services, (ii) health workers' work-related stress and motivation, (iii) service utilization, and (iv) the population health status during and after the PBF implementation. The analysis distinguishes targeted and non-targeted services in order to test the potential disruptive effect of incentives on non-targeted services. It is important to note that

there are many different PBF models (see Miller and Barbiaz (2013) for a review) and that the performance criteria used in DRC were kept simple, based on the number of patients and not on service quality, so that it could be feasibly implemented in the difficult conditions of this country. The results will thus be discussed in light of this particular PBF approach, although we refer to it in the rest of the paper as "PBF" for simplicity.

This study finds that the introduction of the financial incentives in the PBF group led to concrete changes in health workers' behavior. Health workers made greater effort to attract patients: (1) they were more present in facilities; (2) they significantly reduced fees for targeted services; (3) they organized more preventive health sessions at facilities; (4) they conducted more community-based outreach activities to sensitize the population about the services offered by the facility. Overall, the financial incentives thus induced an intensification of effort to increase utilization of targeted health services. This result contrasts with the finding of a study in Zambia in which financial incentives did not induce more effort by hairdressers to sell condoms (Ashraf et al. 2014). Equally important, we find that the increased effort invested in the targeted services did not happen at the expense of the effort invested in the non-targeted services. Also, the reward did not induce significant score manipulation and did not have a negative effect on service quality. Finally, we did not find evidence that the collective nature of the incentive induced significant free-riding. Overall, none of the perverse behavioral effects that could be anticipated with this performance-based financing scheme were realized.

However, the increased effort by the health workers led to a decrease in utilization of health services by the population. We find evidence that the demand for health services responded negatively to price reduction, and did not respond to the effort to ease access to health services. As a consequence, there was less total revenue in these facilities (42% less), even though the two groups received the same subsidy payment level from the government. Consequently, the financial incentive payment mechanism resulted in a 34% reduction in staff revenues, as well as a 12% reduction in health worker job satisfaction. Even more critical, we find a small deterioration in newborn and child health outcomes. The negative response of demand for health to price reduction suggests that prices signal service quality and efficiency. In a context where people are not informed about the benefits of using health services, they might have interpreted lower prices as lower benefits. The lack of response to greater attendance and supply for health services indicates that the population needs more than logistical information, which is also consistent with the idea that a better under-

standing of health service benefits is required. These findings echoe the argument by Glucksberg (1962) and Ariely et al. (2009) that increased motivation may not lead to increased performance when the task is difficult and requires innovative and creative thinking and, although in a different context, the view of Loevinsohn and Harding (2005) suggesting that providers may not possess adequate ability to innovate and change health-seeking behavior if they lack human capital. It is crucial to think about the barriers that impede the increase in demand for health services and were not addressed in the strategies implemented by the incentivized health workers, such as the lack of awareness of the benefit of health care services, or perhaps the general uncertainty about the quality of care. Qualitative evidence from the field indicates that people might interpret prices as signals for service utility, which would explain why price elasticity of demand is positive. The fact that the health workers developed counterproductive actions to stimulate the demand may suggest that health workers are not good entrepreneurs, or that their task was too difficult in a context where demand is non-standard.

Finally, an important result is that staff attendance, which was found to be *higher* in the incentivized health facilities than in the fixed payment facilities when the incentives were in place, was found to be *lower* three months after the incentives were withdrawn. Also, the previously incentivized health workers were found to attach more importance to job material benefits relative to non-material benefits. We show that these effects are not attributable to the decrease in worker income, suggesting that incentive-based payments deterred some of staff intrinsic motivation.

Several empirical studies advocate that performance-based financing improves accountability, efficiency, quality and quantity of service delivery (see Loevinsohn and Harding (2005) and Eichler and Levine (2009) for an overview). However, the presence of confounding factors[2] and the fact that it is often not possible to isolate the effects of financial incentives from other elements[3] make the question of the impact of PBF largely unanswered (Christianson et al. 2008; Eldridge and Palmer 2009; Oxman and Fretheim 2009). Olken et al. (2014) report on a field experiment using a PBF mechanism to improve the efficacy of a grant program to village committees in Indonesia. The PBF mechanism was applied to village committees rather than directly to health workers, so

[2]Until 2011, the studies of the impact of PBF did not use credible comparison groups: they compare very small groups (generally 2-3 districts) which were not randomly assigned to the different treatments (Soeters, 2011; Rusa et al., 2009; Soeters et al., 2005; Eicher et al., 2007; Soeters and Griffths, 2003; Forsberg, 2001), or the situation before and after the introduction of PBF (Sondorp et al., 2008; Eicher et al., 2007; Meessen et al., 2007).

[3]PBF has commonly been a part of a package that may include increased funding, technical support, training, changes in management, and new information systems. In most studies, the level of resources allocated to the health facilities in different treatments is not similar, as well as the level of technical supervision and information system.

this program tested monitoring of health workers by incentivized community members but not the direct effect of incentives to health workers. The incentive led to an increase in health workers' attendance, better health outcomes, and an absence of negative spillovers on untargeted outcomes. Basinga et al. (2011) conducted a quasi-experimental study on the effect of PBF that is the closest to our study. The study took place in Rwanda using a difference-in-difference strategy in order to control for potential selection effects[4]. It finds that PBF is an efficient way to increase utilization of some of the targeted services as well as worker productivity, and to improve some targeted health outcomes (Basinga et al. 2011; De Walque et al. 2013; Gertler and Vermeesch 2013). However, the study does not examine the effects of the incentive on the provision of non-targeted services, nor does it provide evidence on health worker strategies to increase their performance. The literature on the effect of PBF using clean identification is thus very limited, and the lack of information on precise worker responses and strategies still needs to be addressed[5] (Miller and Babiarz, 2013).

There are key policy implications of our findings for governments considering performance-based mechanisms as a way to allocate public resources to the health sector. First, financial incentives may increase health worker motivation overall without reducing service quality or non-rewarded services, nor inducing score manipulation or free-riding. However, there are two considerations. First, financial incentives might reduce the intrinsic component of health worker motivation. Second, our results suggest that contingent rewards might not be appropriate when the task requires complex strategies because users do not respond as expected. The translation of motivation into performance may be better in contexts where demand is classic and the rewarded task is easy. To the best of our knowlegde, this paper is one of the first to show that incentives can decrease performances in service delivery, along with the equally new paper of Rasul and Rogger (2014). In our paper, the reason is not that incentives induce perverse effects on service supply, but that demand for health services is inelastic to eased access and decreases with price.

[4]166 facilities were grouped into 8 pairs and one side of each pair was randomly assigned to pay-for-performance funding, while the other side continued with the traditional input-based funding until 23 months after study baseline. The paper uses a difference-in-difference strategy in order to control for potential selection effects since the number of units of randomization was very small and some post-randomization reassignment of some districts happened because of administrative boundaries' reorganization.

[5]The literature is not very developed in the context of high income countries either, and identification issues also limit the scope of many studies. One recent study on the effect of pay-for-performance mechanism is Mullen, Franck and Rosenthal (2010), which uses a difference-in-difference strategy on US data and show that pay-for-performance targeted on service quality did not lead to any major improvement in quality of targeted services, nor notable effect on the quality of non-targeted services. Note that pay-for-performance in high income countries tends to reward quality measures instead of service volume. This might be because the policy concern is more about service quality than about service utilization in rich countries relative to poor countries. See Stabile and Thomson (2014) for a review.

The remainder of the paper is organized as follows. Section 2 discusses the theoretical background. Section 3 presents the context in which the experiment was set up and the experimental design. Section 4 examines the data and econometric approach. Section 5 presents the effects of PBF compared to a fixed payment approach, and Section 6 concludes.

2 Theoretical Background

In this section, we show that reasonable assumptions on workers' behavior, different from informational and reputational effects, can produce predictions consistent with the evidence found in the literature that extrinsic rewards may backfire (Glucksberg (1962), Deci (1971), Lepper et al. (1973), Deci and Ryan (1985), Kohn (1993), Ariely et al. (2009), among others).

Let's consider that an agent engages in a certain task if and only if his interest in the activity is larger than the cost of undertaking the activity c. If the agent succeeds in producing the output, his benefit in the activity is two-fold: the intrinsic value that the agent attributes to the output, V, and a contingent reward b. If he does not succeed the benefit is zero, and the probability that he succeeds if he undertakes the activity is his ability θ. His utility of engaging in the activity is thus given by $U = \theta(V + b) - c$. Holding everything else equal, the reward b therefore increases the probability that the agent engages in the activity.

Benabou and Tirole's (2003) framework provides an explanation of why rewards may be counterproductive in the presence of information assymmetries: rewards from a knowledgeable principal may inform a more ignorant agent about (i) the difficulty of the task c, (ii) his ability to succeed in the task θ, or (iii) the intrinsic value of the task V[6]. These informational effects of rewards have a negative impact on the agent's utility which might be large enough to offset the positive impact of θb. Moreover, the permanent nature of informational effect makes the effect on motivation last after incentives have been withdrawn. Benabou and Tirole (2006) adds another explanation of why extrinsic rewards might crowd out agent's effort during the period when incentives are in place: an agent concerned by social reputation or self-respect concerns might want to reduce his effort in response to the introduction of extrinsic rewards in order to signal his intrinsic motives. As Gneezy et al. (2011) summarizes, economists think about the effects of incentives on behavior in terms of information and signaling.

[6]Individuals may have imperfect memory why they engaged in the task. Note that in this case there is no need for the principal to have information that the agent does not have. It is just information that the agent had and lost.

However, the literature suggests that informational and reputational effects are unlikely to drive all situations where rewards backfire. For instance in Deci (1971), Zeevi (1971) or Lepper et al. (1973), students who are offered external rewards for performing tasks such as solving a series of puzzles or art activity are less likely to engage in these tasks after the external reward has been withdrawn. The authors do not find evidence that students feel less able to perform or that they find the task more difficult. The authors find that students express less interest in the task itelf. The explanation through information asymmetries does not fit well with these observations because it seems unlikely that the students learned much about the difficulty of the puzzles, or their ability to perform the puzzles, or the intrinsic value of solving puzzles in the process. In such contexts where information assymmetries seem small or non-existent, the reason for a decrease in motivation after the reward is withdrawn should be found elsewhere. In fact, the authors point to the fact that the introduction of extrinsic motives causes a shift in attention from the value that the agent attributes to the output V in favor of the external benefit b. Actions perceived as "ends" in the absence of external motives tend to become "means" when external rewards are introduced (Lepper et al. 1982). Deci and Ryan (1985) argues that rewards change the locus of control from internal to external and make agents bored, alienated and reactive rather than proactive. Kohn (1993) notes that rewards make people less enthusiastic about their behaviour. Overall, what psychologists call the "overjustification effect" is compatible with situations where principals and agents have similar information about agent's ability or the difficulty and intrinsic value of the task to start with, but agents pay less *attention* to the intrinsic value of the task after they were offered extrinsic rewards.

We thus propose (and test in our empirical analysis) an alternative habit-forming based theory of motivation by introducing: (i) a parameterα reflecting the attention paid to the *extrinsic* benefit from the task b (the reward), and (ii) a parameterβ reflecting the attention paid to its *intrinsic* value V. The total motivation is given by: $U = \theta(\beta V + \alpha b) - c$ where α and β reflect weights attached to extrinsic versus intrinsic motives in agent's utility. Our framework is based on two key assumptions: (1) First, $\beta = f(\alpha)$ with $f^{'}(.) \leq 0$, which represents the idea that people have limited attention in the sense that more attention paid to extrinsic motives can only decrease attention paid to intrinsic ones, or do nothing. (2) Second, attention is prone to habit-forming: as soon as attention is drawn to a motive, it remains permanently because people do not forget. For example, an awareness campaign that would draw agent's attention to the intrinsic value of the task would shift β from 0 to a positive value. Similarly, the introduction of a financial remuneration for succeeding at the task

would shift α from 0 to a positive value. Then the levels of intrinsic value$V$ and extrinsic benefit$b$ are likely to vary over time depending on new information, experience, principal's decisions etc., but agent's attention paid on intrinsic and extrinsic motives cannot disappear. Attention parameters α and β evolve over time only through the relationship $\beta = f(\alpha)$ in contexts where attention is limited (f non constant).

Under this framework, post-reward motivation is smaller than pre-reward if and only if $f^{'} < 0$. The mechanism is as follows:

Before Exposure to Extrinsic Rewards The agent has never received any extrinsic reward for engaging in the task ($b = 0$), and therefore he does not pay attention to extrinsic motives ($\alpha = 0$). The utility of engaging in the activity is $U_0 = \theta f(0)V - c$.

During Exposure to Extrinsic Rewards The principal introduces an extrinsic reward $b > 0$, and the agent pays a level of attention $\alpha > 0$ to the reward. The utility of engaging into the activity is now $U_1 = \theta(f(\alpha)V + \alpha b) - c$. Since $\alpha > 0$, $f(\alpha) < f(0)$ and U_1 can be both smaller or larger than U_0 depending on the size of the increase in motivation due to the new extrinsic reward αb relative to the loss in motivation due to the shift of attention paid to intrinsic motives $(f(0) - f(\alpha))V$. Typically small rewards in contexts where intrinsic motives are large would decrease the total motivation, while large rewards in contexts where intrinsic motives are small would increase the total motivation.

After Exposure to Extrinsic Rewards The principal stops providing the agent with rewards (b is back to 0). However, attention parameters remain the same because attention is prone to habit-forming, and the utility of engaging in the activity is now: $U_2 = \theta f(\alpha)V - c$. U_2 is smaller than U_1 if and only if $\alpha > 0$, meaning exposure to extrinsic reward in the previous period effectively drew agent's attention on extrinsic motives. Moreover, U_2 is smaller than U_0 if and only if $f^{'} < 0$, meaning agent's attention is limited.

In a context of unlimited attention ($f^{'} \equiv 0$), the introduction of an extrinsic reward would attract some new attention from the agent without decreasing pre-existing attention paid to the intrinsic value of the task V. After the extrinsic reward is withdrawn, pre-existing attention paid to the intrinsic value$V$ would remain intact and motivation would be back at its pre-reward level, without any detrimental effect on later motivation.

In the empirical part of the paper, we test this theoretical framework by eliciting the relative size of α and β and comparing agents who have been exposed to extrinsic reward (the performance-based financing system) and people who have not. A testable prediction is that α should be larger and β smaller among the former than the latter. We also provide evidence on the comparison between the different utilities of engaging into the activityU_0, U_1 and U_2.

Finally, rewards may backfire not because of reduced motivation, but because increased motivation is accompanied by reduced performance. There are two main explanations for the reduced performance. One is the multitask interpretation proposed in Hölmstrom and Milgrom (1991) where agents focus on the rewarded actions at the expense of other actions that might be necessary to perform well, which we also refer to as a substitution effect[7]. Second, Glucksberg (1962) and Ariely et al. (2009) observe that people who are offered a reward for performing at some tasks perform better at simple tasks but worse at tasks calling for cognitive skills, indicating that rewards generate negative stress limiting one's creative thinking (Baumeister, 1984). In this case, lower ability is associated with larger rewards ($\theta'(b) < 0$) not because the agent infers that his ability is lower than expected, but because larger rewards constitute larger stakes and larger stress which narrows the mind and reduces the ability to perform (Kamenica, 2012). McGraw and McCullers (1979) explains that reward leads to underachievement when the task requires open-minded thinking because the focus of attention limits one's capacity to draw unusual connections between elements. Under negative stress, the agent initially provides *more* effort and yet produces a smaller output[8]. In other words, the utility$U$ is larger but a change in the content of the effort makes it less productive. One example would be a student whose motivation in passing her exam gets larger due to the promise of a gift conditional on passing and who would spend hours reading her lessons without being able to learn because her mind would be distracted. In the context of health workers exposed to financial rewards, the phenomena may show itself through motivated but stressed out workers who would be more tense, less staid with patients, or who would decrease time spent with the patients in order to consult more of them.

All in all, the theory thus leaves room for both positive and negative effects of financial incentives on both workers' effort and performance, making the question empirical in essence.

[7]For instance, increased use of prenatal visits might not lead to a reduction in child mortality if mothers do fewer postnatal visits

[8]After some periods, the agent may observe that her effort is *de facto* not as much productive as expected as adjust her effort accordingly.

3 Experimental Set-Up

3.1 Background on Health in DRC and Haut-Katanga

The Democratic Republic of Congo (DRC) is the second largest country in Africa by area, with the fourth largest population at 66 million (World Bank, 2012). It is also among the poorest countries in the world: the country is ranked second from the bottom of the Human Development Index (186 out of 187 in 2012) (UNDP, 2012), with an estimated per capita income of US$ 220 (current) in 2012 (World Bank, 2012). Impoverished by decades of war, instability and bad governance, DRC is not on track to reach the health-related Millenium Development Goals. Since the democratic elections in 2006, the country has started a slow reconstruction phase and a decentralization process, with the election of provincial governments, including provincial ministers of health. Performance-based Financing (PBF) is a strategy for improving health outcomes among the population which has been developed and implemented to promote effective service delivery.

The district of Haut-Katanga entails 1.26 million people in the province of Katanga in the south-eastern corner of the DRC. From September to November 2009, a survey was conducted to better understand the health situation in Haut-Katanga by providing a description of the functioning of the health facilities as well as the characteristics and behavior of the health workers, patients and households in the district. The survey sample entailed 152 health facilities (5% referral centers, 71% health centers and 24% health posts)[9]. This survey indicated that the initial situation of the health facilities in Haut-Katanga was worrying primarily because of the poor quality of health services rather than the coverage for basic health services. In regards to health services coverage, 87% of patients lived 10km or less from facilities, 70% spent less than one hour to travel to the facility, and there was one health worker for every 1860 individuals[10].

However, the poor quality of infrastructure was striking: only one out of four facilities had access to a water tap or electricity. The majority of facilities had only low-cost basic equipment. Most health workers were not public agents: one worker out of four did not receive any fixed wage from the government. Worker payment thus came from facility revenue, mainly user fees and drug sales, but also public grants and -sometimes- funds from NGOs and private donors. Relatedly, health workers were often engaged in multiple activities: the typical health worker earned 61% of his income from the health facility, while 39% from other jobs and/or agricultural production.

[9] 161 health facilities were recognized as part of the government health system in the district, among which 5 hospitals were excluded from the study and 4 health centers could not be reached.

[10] The ministry of health considers that there should be at least one health worker for every 1500 individuals.

However, health workers spent 52 hours per week working in the health facility. They received 35 patients the week before the survey equating approximately 7 patients per working day each which means thathealth workers were not overworked and likely spent some time waiting for patients to come. Patients reported quite short consultation time (16 minutes on average), and twice as much waiting time before the consultation (30 minutes on average)[11]. 56% of patients had to pay a fee for the service, although the median fee for a visit was quite low 800FC (0.88$).

In this survey, the health status of the population was found to be poor: 25% of the sample had been sick in the last four weeks, with malaria and diarrhea being the most prevalent diseases. Concerning maternal health, 31% of births in the last 12 months were not attended in a formal health facility. Mothers used more prenatal than postnatal health services: 76% of women pregnant in the last 12 months had at least one prenatal visit while only 10% attended a postnatal visit. However, according to women's recall, only a third of prenatal visits included the minimum tests. Despite frequent immunization campaigns, only 13% of children under 5 years-old were able to present an immunization card (although based on mothers' declaration a majority of children got immunized at least once). Finally, we found low exposure to prevention campaigns other than immunization, with around two thirds of the households never exposed to any HIV prevention, child nutrition, or maternal health campaign.

3.2 Experimental Design

Payment Calculation

In the Haut-Katanga district, the 96 health areas (totalizing 152 health facilities) were randomly assigned to one of two payment systems. *In the fixed payment group*, the amount allocated to each facility was calculated based on the staff in the facility: a list of eligible workers was established at the beginning of the pilot by the Ministry of Health. Each worker was entitled to a given amount of governmental payment depending on his grade and experience. *In the performance-based payment group*, payments were made based on declaration of service volumes by facilities. The targeted services included seven services at the primary care level (outpatient first curative consultations, prenatal consultations, deliveries, obstetric referral, children completely vaccinated, tetanus toxoid vaccination, and family planning consultations) and three additional services at the secondary care level (C-section, blood transfusion, and obstetric referrals to hospitals). Relative prices for each

[11]This survey did not allow for assessing the technical quality of medical procedures.

service are presented in Appendix Table 1.

Formally, payments to health facilities can be written as:

$$P_{i,m} = \alpha_i + \beta_m Q_{i,m}$$

where $P_{i,m}$ is the payment to facility i in month m, α_i represents a fixed component, $Q_{i,m}$ is the vector of targeted service quantities provided by facility i in month m, and β_m is the vector of prices that the government attach to each targeted service in month m. The PBF group was characterized by a pure performance-based mechanism ($\alpha_i = 0$ and $\beta_m > 0$), whereas the comparison payment group had a pure fixed payment ($\alpha_i > 0$ and $\beta_m = 0$). In order to ensure neutrality in the level of funds received by both groups and to isolate the incentive effect from the resource effect, the total budget allocated to health facilities in the PBF group was the same as the total budget allocated to health facilities in the fixed payment group:

$$\forall m, \sum_{i=1}^{n} \alpha_i = \sum_{i=1}^{n} \beta_m Q_{i,m}$$

where n is the number of health facilities in the PBF and in the fixed-payment group. Hence, noting $\overline{Q_m}$ the average service provision in the PBF group in month m and $\overline{\alpha}$ the average payment in the fixed payment group:

$$\overline{\alpha} = \beta_m \overline{Q_m}$$

Since the budget of the health provincial authorities was fixed and could not vary every month depending on the average service provision in the PBF group, $\overline{\alpha}$ was fixed and β_m was adjusted accordingly at $\frac{\overline{\alpha}}{\overline{Q_m}}$[12]. Although relative prices attached to the targeted services were constant (see Appendix Table 1), absolute prices and facility payments were thus determined by the quantity of services provided by the facility *relative to* the quantity of services provided by the other incentivized health facilities. As discussed in Bandiera et al. (2005), relative incentives might yield lower effort from the health workers than piece rates because effort imposes a negative externality on others, in particular when others are friends[13]. The budget used in this experiment estimated at $0.43 per capita per year (average monthly facility payments were $550 and the average catchment area

[12]The other way to equalize the two total budgets is to fix $\beta_m = \beta$ and adjust $\overline{\alpha}$ accordingly at $\beta\overline{Q_m}$. This technique was used in the Rwanda experiment where the governmental budget could increase according to the average service provision in the incentivized group.

[13]In the context of this PBF program, we do not have measures of interpersonnal connections between workers of different health facilities. However, health facilities are generally distant one from another and it seems unlikely that health workers from different health facilities live in the same neighborhood and are close friends.

population was 12,900)[14]. The average monthly payment by facility from June 2010 and September 2012 did not differ in the fixed payment and in the PBF group. This confirms that the experimental design was respected and that the study isolates the incentive effect from any resource effect. Figure 2 shows the distribution of the average monthly facility payment over the study period by treatment status. Payments proved more disperse under PBF than uner fixed-payment, suggesting heterogenous responses to the incentive with some health facilities getting less than under a fixed payment mechanism and others getting more.

Performance Verification

Service volumes were measured using monthly reports submitted by facilities, in which the number of patients for each targeted service was reported. These numbers were verified by public agents at the beginning of the following month by comparing reported volumes with those found in health facility registers[15]. Payments were calculated and paid as soon as the register verification was done, generally during the following month.The same payment lag applied to the fixed payment group since all payments happened at the same time. Subsequent verification of the information noted in the registers was also conducted: a random sample of 30 patients[16] from the registers were selected and visited by independent associations to check the accuracy of the information reported[17]. A system of retroactive financial sanctions was integrated in order to reduce providers' incentives to submit fraudulent reports and register phantom patients.

In reality, the community verification system proved weak: PBF facilities only received 3 community verifications on average throughout the experiment and there was no effective financial sanction associated with being caught for fraudulent over-reporting. Specifically, the reductions in payments were proportionally equal to the percentage of patients not being identified through community verification. For example, if 18% of patients were not found through community verification, the facility would only receive a reduction of 18% in their corresponding payment and

[14]This is lower than in other contexts where output budgets range between $2 and $3 per capita per year.

[15]Register verification was also meant to take place in health facilities under the fixed payment mechanism since the government wanted to improve the accountability of health facilities in general, not only as an element of PBF. At endline, the average number of register verifications in the last 12 months is 7 in both in the PBF and in the fixed-payment group (p-value of the test of equality of means in the two group = 0.48).

[16]The 30 patients were chosen such that each targeted service is present in the sample, but none of the non-targeted services.

[17]Community verifications were meant to take place only in the PBF group as part of the financing mechanism. However, we conducted community verifications in the fixed payment health facilities for impact evaluation purposes (1 community verification by facility in the comparison group). The fixed payment health facilities had no incentive to cheat on service volumes so the comparison of discrepancy rates between the PBF and the fixed payment groups allow for differentiating cheating from natural -unavoidable- discrepancies due to the fact that some patients moved or were absent at the time of the verification.

no additional sanctions were enforced. Despite the weak verification process, we did not find any significant difference in the propensity to report phantom patients in the registers: the average proportion of missing patients was found to be 17% in the fixed payment group and 21% in the PBF group. The difference was not statistically significant. However, the health workers in the PBF group were significantly more likely to fill out consultation reports for their patients than in the fixed payment group: in the endline survey, 94% of health workers declared that they fill out a consultation report for each patient in the PBF group, whereas only 78% of health workers in the fixed payment group. Since patients were reported in the register based on consultation reports, service utilization was under-reported in the fixed-payment group. It is interesting to note that the issue with administrative data is not that incentivized health workers inflate artifically the number of patients, but rather that the non-incentivized workers under-report the number of patients as they don't have any financial benefit of reporting accurate service volumes. Overall, it is crucial to rely on an independent source of information about service utilization for the impact evaluation since administrative data does not give an accurate idea of service utilization in the fixed-payment group.

Autonomy of Payment Allocation

The autonomy of payment allocation among facility staff in the PBF group led to a significantly more egalitarian distribution of payments among workers. In the fixed payment group, 77% of health workers received a share of the payment, whereas 93% of workers in the PBF group[18]. PBF actually benefited non-technical workers (pharmacists, managers, secretaries, receptionists and maintenance workers) who were not in the governmental payroll and therefore did not receive a share of the fixed payment[19]. Consistently, the average last payment to health workers showed less dispersion in the PBF group: the standard deviation was 36% lower in the PBF group than in the fixed payment group and the difference is significant. This confirms the finding of a more egalitarian distribution of the payment among workers in the PBF group.

[18]This information was collected at endline from the facility heads. The facility heads listed the workers in the facility, indicated whether each of them received a share of the last payment, and the corresponding amount.

[19]Since the fixed payment is calculated based on the number and qualification of workers on the governmental payroll, the workers typically receive the amount of money corresponding to their contribution to the fixed payment.

Resulting Structure of Worker Motivation

We now link our theoretical framework and the actual context of this specific experiment. The task workers engage into is attracting patients, and the output is the number of patients. The experiment manipulates the structure of motivation by adding a contingent reward in worker utility of attracting patients.

In the fixed payment group, worker utility of attracting patients is driven by the intrinsic value they attribute to this task, as well as a contingent benefit coming from user fees. In fact, the more health workers consult patients, the higher facility's revenue from user fees, which constitutes a large part of workers' remuneration. Workers in this group thus already pay attention to extrinsic motives through this benefit from user fees. Let's denote $(\alpha_0, f(\alpha_0))$ the levels of attention paid respectively on extrinsic and intrinsic motives, and F the level of worker remuneration from user fees, in the fixed payment group. Using the parameters of our theoretical framework, workers' utility of attracting patients in the fixed payment group is thus $U_0 = \theta(f(\alpha_0)V + \alpha_0 F) - c$

In the PBF group, the difference is the introduction of a new contingent reward from the government. Governmental payments play a role in the PBF group by increasing workers' benefit of engaging into attracting patients. Let's denote $(\alpha_1, f(\alpha_1))$ the new levels of attention paid on extrinsic and intrinsic motives resulting from the introduction of this new category of contingent reward, and P the governmental payment. Workers' utility of attracting patients in the PBF group is thus $U_1 = \theta(f(\alpha_1)V + \alpha_1(F + P)) - c$. After government payments are withdrawn, worker utility of attracting patients is unchanged in the fixed payment group, while it becomes $U_2 = \theta((f(\alpha_1)V + \alpha_1 F) - c = U_1 - \theta\alpha_1 P$ in the PBF group.

In the rest of the paper, we will present evidence on the relative size of U_0 and U_1, U_0 and U_2, as well as α_0 and α_1.

4 Data and Empirical Strategy

4.1 Data Sources

Five sources of data are used for the impact evaluation.

Baseline Survey A survey was administered between September and November 2009. Only 85% of health facilities involved in the experiment (129 out of 152) were interviewed in this survey. As a result, we perform the balance checks on this subsample of our experimental sample.

Administrative Data Administrative data are available every month from January 2010 to December 2012 for all 152 health facilities. This data includes the number of targeted services provided, the payment due to the health facility, the actual payment made to the health facility, whether a performance verification occurred and related indicators (e.g., % missing patients and consequent financial sanctions). We use this data to examine payments received by the facilities but we do not rely on it to measure service provision and utilization since it can be both manipulated and not evenly reported in the PBF and fixed payment groups as a consequence of the incentive.

Qualitative Data In April and June 2012, qualitative interviews were conducted in 31 health facilities randomly selected in 4 out of the 8 health zones (Kafubu, Kipushi, Kasenga and Lukafu). In each facility, one interview was done with the facility head and another one with a health worker (on a voluntary basis). In total, 29 facility heads and 31 health workers were interviewed, all by the same person. They were equally distributed between the PBF group and the fixed payment group. Questions were all open and dealt with the perception of the payment (transparency, fairness, understanding of the calculation), the general functioning of the health facility, recent changes that might have occurred in the facility, and obstacles to improve the number of patients and the quality of services.

Attendance Spot Checks Unannounced spotchecks were performed in July, August and September 2012 to collect data on worker attendance in the health facilities that is impervious to gaming.

Endline Survey A final survey was administered between December 2012 and February 2013, four months after the PBF mechanism was withdrawn. The endline survey was administered in 87 out of the 96 health areas involved in the experiment. The rainy season and the insecurity created by the Maï Maï insurgency made it impossible to reach the other 9 health areas. Attrition occurred at the same rate in both groups, with 44 health areas in the PBF group and 43 in the fixed payment group included.

The endline survey included four different questionnaires for facility heads, health workers, patients straight out of consultation, and households living in the catchment area. Appendix Table 2 reports the endline sample size by questionnaire and treatment status. All facilities in the 87 health areas that could be reached were interviewed, totalizing 123 health facilities. All the technical staff

in each health facility was interviewed up to 10 persons[20], totalizing 332 health workers. A sample of 10 patients per facility was randomly selected for exit interviews, or the maximum available if fewer are present, totalizing 1,014 patients. Finally, the household questionnaire was administered to 1,708 households: 20 households were interviewed in each of the 87 health areas, among which 10 households randomly chosen in the population and 10 randomly chosen among the households with a pregnancy in the last 12 months[21]. Appendix Table 3 shows basic descriptive statistics of the endline sample.

4.2 Outcomes of Interest

Cost of Health Services Changes in user fees in the incentivized group could be a strategy to increase utilization of targeted services and, therefore, increase payments by reducing the cost to patients. The reverse effect may happen on non-targeted services, as a way to compensate for the loss in revenue from targeted services, or discourage demand for non-targeted services.

User fees were collected from the facility heads at endline and from users in the last 12 months. In order to compare fees declared by facility heads across the largest number of health facilities, we used the fees of the most commonly offered services: curative consultations, birth delivery, prenatal visits, postnatal visits, and preschool consultations. To improve statistical power to detect effects that go in the same direction within a domain, we also present findings for a Fee Summary Index that aggregates information over all these user fees (following Kling et al, 2007), as well as a Fee Summary Index for targeted services (curative and prenatal consultations, and birth delivery) and a Fee Summary Index for non-targeted services (postnatal and preschool consultations). We also collected user fees from users in the last 12 months to examine price levels when payments were in place. For preventive services, we present user fees reported by users in the last 12 months on the one hand and users before September 2012 -when PBF was implemented- on the other hand. For curative services, we were not able to apply the same strategy as we only asked about the last visit

[20]In the facilities staffing more than 10 health workers, 10 were randomly chosen from the list of all health workers during the facility head interview. The health workers who were present the day when the interviewer visited the health facility were interviewed on-site, whereas the others were visited at home. Only those health workers who were out of the neighborhood at the time of the survey (because they were on vacation or because they temporarily migrated) could not be interviewed.

[21]The selection of the 20 households was done as follows: four axes in the locality were randomly drawn from a central point, then one household was visited every five houses on each axis. - On two axes, all households were eligible and took the survey if it consented to (otherwise the next household was visited). After each interview, the interviewer went five houses further and continued the selection until he could interview 5 household on each axis. - On the two other axes, only households where a woman had been pregnant in the last 12 months were eligible. If the household did not meet the criteria, then the next household was visited etc. until an eligible household was found. After each interview, the interviewer went five houses further and continued the selection until he could interview 5 household on each axis.

which mostly happened after September 2012.

Accessibility of Health Services The facility's opening hours, the number of service varieties offered, and the number and qualification of workers were collected from facility heads. To examine access, patients and household members were also asked whether they could consult every time they visited. Worker attendance (number of health workers present at the facility) and on-the-job effort (number of health workers actually working) were collected from the unannounced spotchecks done by independent research assistants[22]. Regular preventive sessions at the facility help service utilization by giving greater opportunity to users to access preventive services. The number of preventive sessions organized at the facility in the last 12 months was collected from facility heads. Also, outreach activities in communities are made to inform the population about the preventive sessions (topic, day and hour). The number of outreach activities in the community in the last 12 months was collected from health workers. Using the service related to each preventive session and outreach activity, we can separate the number of activities related to targeted services (prenatal care, immunization and family planing) from the number of activites related to non-targeted services (postnatal care and HIV prevention).

Service Quality Service quality is primarily measured by technical quality. Consultation time is considered as a component of service technical quality, although we consider compliance with standard medical procedures as the main indicator. Compliance was assessed on patients immediately following the consultation who consulted for illness: they were asked whether three basic procedures were followed during the consultation (being weighted, examined and having his tension checked). Compliance was also assessed on women who gave birth in the last 12 months who were asked about standard procedures applied during prenatal visits (weighing, stomach palpation, tension check, stomach measure, HIV test, tetanus shot, blood test, urine analysis and information on immunization schedule) and postnatal visits (stomach palpation, child weighing, child examination, child immunization and child immunization card). We also measure the proportion of patients straight out of consultation who visited for illness whether they were prescribed drugs without being examined, as well as the number of days women attended the facility after giving birth.

[22]Note that the interviewer reported the number of workers present and working without telling the facility heads and the workers. The purpose of the visit was officially related to administrative matters and not attendance checks in order to avoid any interference with worker behavior at a later point. Observational data on workers' attendance and on-the-job effort was anonymous and aggregated at the facility level.

Finally, as complementary measures of service quality, we use the proportion of patients who understood the diagnosis and prescriptions, as well as the proportion of patients and household members who were satisfied with the visit.

Service Utilization First, we measure overall health service utilization by asking each household member whether s/he visited a health facility in the last 12 months. Second, we disentangle utilization of different services: curative services, child immunization, maternal health services and family planning

For curative services we examine whether each household member has been sick and visited a health facility in the last 12 months and, if so, the number of days with symptoms before s/he visited.

For child immunization we look at whether each child aged 0-5 had at least one immunization shot, the number of immunization shots reported in the immunization card (if any immunization card), and whether a scar from TB immunization could be observed on the child's shoulder. To focus on immunization when payments were in place, we restrict the sample to children aged at least 15 months at endline who were thus at least 1 year-old when payments were withdrawn.

For maternal health services we look at the number of prenatal visits to a health facility by women who have been pregnant in the last 12 months, whether delivery (if any) was attended, whether delivery (if any) was done with a c-section, and at the number postnatal visits to a health facility. We focus on utilization when payments were in place by restricting the sample to women who gave birth before September 2012. We also examine the total number of pre- and postnatal visits including healers and den mothers out of health facilities in order to take into account potential substitution effects between modern and traditional maternal health services.

Finally, for family planning we asked each woman aged 15-49 whether she was using a modern contraceptive method: IUD, daily pill or implant. We also use whether each woman aged 15-49 has had a pregnancy in the last 12 months as a direct measure of utilization of family planning. Pregnancy rate was calculated on the representative sample (randomly selected households) only since, by construction, all women in the other sample have been pregnant in the last 12 months.

Population Health Status We use mortality rates as well as standard under-5 weight-for-age and height-for-age z-scores to assess health status. Mortality rates are measured using the number of persons who died in the last 12 months in the household, in particular the number of women

who died for perinatal reasons, and the number of children under 5. We also use the proportion of new-born in the last 12 months that are still alive. To focus on impact when payments were in place, we show results restricting the sample to children born before September 2012.

Health Facility Revenue Depending on the price-elasticity and access-elasticity of the demand for health services in the population, it is unclear what the effect of PBF on total resources in health facilities is. We thus examine all sources of revenue at the facility level the month before the endline survey as reported by facility heads, as well as workers' payment the month before the survey as reported by facility heads and health workers themselves.

The enumerators also observed the quantity and quality of equipment and infrastructure during their visit, which reflect both total revenue and management decisions made at the facility level. We constructed three indices, each index being the first component of a principal component analysis. The *quality index* is based on direct observation by the enumerator when s/he arrived at the facility for the endline survey of twelve items: building quality, waiting room, consultation room, lavabo, soap, clean towels, bathrooms, sterilization material, permanent display of user fees and drugs' costs, use of an examination table and ordinogram. The *infrastructure index* includes six items: phone ownership, motorized transportation mean ownership, access to clean water, toilet and electricity, and hard roof. Finally, the *equipment index* includes the quantity of fifteen types of medical equipment owned by the health facility: generator, sterilizer, tensiometer, stethoscope, baby-scales, weighing scale, height gauge, microscope, gynecological examination table, fridge, delivery boxes, fuel, kerosene, bed and solar panel.

Worker well-being and motivation To measure workers' well-being and motivation we measured their job satisfaction, stress, anxiety and conflicts within the facility. We also looked at free-riding behavior and at the nature of worker's motivation.

We measure job satisfaction by asking at which level the worker would place his job satisfaction on a scale from 0 to 10. We measure stress by asking the workers whether they find their workload heavy, report too much work, or felt tired in the last 7 days (a summary index of these three components reflecting stress, the "subjective workload index", is also presented). We measure anxiety by asking the workers whether they worry about job remuneration because of (i) its volatility, or (ii) its level; we also ask whether workers feel in competition with other health facilities. Finally, we measure conflicts by asking the workers whether the payment allocation was a source of conflict

within the facility, and at which level the worker would place conflicts among workers on a scale from 0 to 10.

Since free-riding is a concern when incentives are collective, we present some statistics on the distribution of effort within the facility using the number of outreach activities for targeted services in the last 12 months per agent. First, we show the proportion of agents who did not do any outreach activities in order to assess whether some workers changed their effort on the extensive margin. Second, we present the 25th, 50th and 75th percentiles among agents who did some outreach activities to assess whether workers changed their effort on the intensive margin, and where. Finally, for facilities with at least two agents, we present the standard deviation of the number of outreach activities per agent at the facility level to test whether the incentives changed the dispersion of effort among workers, and not simply induced a homogenous translation.

The effect of financial incentives on the nature of worker motivation is measured using worker attendance after the payments were withdrawn on the one hand, and worker motive elicitation on the other hand. At the time of the endline survey, workers are no longer incentivized in the PBF group so the incentive structure does no longer differ between the two groups: workers' behavior is driven by intrinsic motivation (perceived value of the job) and extrinsic motives (job remuneration). Any difference in worker behavior therefore reflects persistent effects of PBF on either intrinsic motivation, or job remuneration. Staff attendance provides a measure of workers' total motivation (intrinsic plus extrinsic). We also elicit workers' motives: workers were asked about the main advantage of their occupation, then about the main disadvantage. These questions were open in order not to induce any type of response and capture the most salient motives, which are those that come at the top of their mind. We classified the responses into seven categories of advantages (social recognition, remuneration, material comfort, care about others' health and life, power, interest in the activity) and six categories of disadvantages (lack of social recognition, low remuneration, low material comfort, responsibility over others' life, too much pressure and responsibility, risk of being sick due to the contact with patients). We calculate the proportion of workers who mention either remuneration or material comfort as the main advantage, or low remuneration or low material comfort as the main disadvantage. We use this proportion as a measure of the relative importance of extrinsic versus intrinsic motives in workers' total motivation.

4.3 Empirical Strategy

Validation of the Experimental Protocol The internal validity of the impact evaluation relies on the comparability of the fixed payment and the PBF groups as observed at endline. With a large number of units of randomization, the law of large numbers insures that the characteristics in both groups are balanced. Here randomization was done on 96 health areas and it is preferable to check out whether the pre-program characteristics of the fixed payment and the PBF groups are similar.

This comparison was done using the 2009 survey administered to health facilities, health workers, and randomly chosen households in the catchment area. As explained earlier, only 85% of health facilities involved in the experiment took the 2009 survey. As a result, 129 out of the 152 pilot health facilities can be observed to check how characteristics were initially balanced between the fixed payment and the PBF groups. Most initial characteristics are balanced, although the urban health facilities (17% of the sample) were not equally distributed in the PBF and fixed payment groups: they represent 12% of the PBF health facilities while 23% of the fixed payments ones. Since the urban health facilities, staff, patients and households are likely to differ from the rural ones, Appendix Table 4 presents the means of observables collected in 2009 in the PBF and fixed payment groups as well as t-tests for the following null hypothesis: the difference is zero controlling for a dummy indicating whether the unit of observation is located in a urban area. 2 differences in means are significant at the 10% level out of 57 tests, which is consistent with what would be expected with random sampling variations.We are therefore confident that differences in outcomes at endline between the two groups are not driven by initial conditions as long as we control for urban location.

Estimation Strategy For each outcome of interest, we show the estimation results of an equation of the form:

$$Y_i = \alpha + \beta PBF_i + X_i^{'}\gamma + \varepsilon_i$$

Where PBF is a dummy for being in the PBF group. Because the treatment was randomly assigned, it is in expectation uncorrelated with the error term and can therefore be estimated through OLS. Coefficient β estimates the average local effect of PBF and is presented in the third column of our result tables after the unit and number of observations We show the p-value for a

test that this coefficient is equal to zero in the fourth column of the result tables.

The unit of observation i varies: it stands either for a health area, a health facility, a health worker, a patient straight out of consultation, a household, or a household member. Following the results of the balance checks discussed above, we control for a dummy indicating whether the facility is urban. To improve the precision of the estimation of the average treatment effect, we also use a small set of controls X_i which varies according to the unit of observation i: At the *health area level*, it includes a dummy variables for the health zone (the Haut-Katanga province entails eight health zones) and whether the majority of health facilities in a specific geographic area are religious. At the *health facility level*, it includes dummies indicating the health zone, and whether the health facility is religious. At the *health worker level* it also includes dummies indicating that the health worker is a female, a doctor, a nurse, as well as the age and number of years of experience of the health worker. At the *patient level* it includes a dummy indicating that the patient is a female, the age of the patient, and the reason for the visit. At the *household level*, it includes the sex and age of the household member, and for women a dummy indicating that the woman is literate. The results are robust whether or not these controls are included in the regressions. We favor the results controlling for these characteristics since it improves the precision of the estimates. Finally, we clustered error terms at the health area level to take into account potential correlation between units in the same assignment unit.

5 Results

5.1 Prices of Health Services

Table 1 presents the effects of PBF on user fees and drug prices. We find consistent evidence that PBF induced a reduction in user fees for some targeted services.

Based on the information collected from facility heads, the mean Summary Fee Index is not statistically different between both groups but looking at targeted services only it is significantly lower in the PBF group : the mean Summary Fee Index for *targeted services* in the PBF group is .81 standard deviations below the mean in the fixed payment group (significant at the 10% level). Figure 3 shows that PBF reduced the proportion of health facilities offering high fee levels, and we see a higher concentration at low fee levels, suggesting that the decrease in prices mostly results from the response of those health facilities which would offer high prices under a fixed payment mechanism. In contrast, the mean Summary Fee Index for *non-targeted services* of the PBF group

is 0.4 standard deviations below the mean of the fixed payment group, a difference which is not statistically significant[23]. The detailed analysis of fees by service shows that the average fee is lower in PBF facilities for all services, although the difference is significant only for prenatal visits: the first prenatal visit costs 442 FC in the PBF group instead of 850 FC in the fixed-payment group (a 48% decrease), and the second prenatal visit costs 51 FC in the PBF group and 132 FC in the fixed-payment group (a 61% decrease).

The reduction in user fees is confirmed by users. Patients straight out of consultation who visited for a curative consultation paid 26% less for the consultation and 49% less for the drugs in the PBF health facilities than in the fixed payment facilities (these differences are significant respectively at the 5% and 1% levels). Moreover, women who have been pregnant in the last 12 months and visited for a prenatal visit report a 21% lower fee in the PBF group than in the fixed-payment group (significant at the 5% level). This effect is even more pronounced on pregnant women who visited before September 2012 when PBF was in place: we find a 31% decrease in fees for prenatal visits (significant at the 1% level). Overall, the data consistently suggest that PBF encouraged health facilities to decrease the price of some targeted services, in this case curative care and prenatal care, probably a strategy to attract more patients. Prices are also found to be rigid since they remained lower in the PBF group even after PBF was withdrawn.

5.2 Accessibility of Health Services

Table 2 presents the effects of PBF on service accessibility: facility opening, service supply, staff composition and attendance, as well as supply for preventive sessions at the facility and outreach activities.

Health Facility Opening and Services Offered We find that PBF did not change the extent to which health facilities are open: according to the facility heads, facilities open on average 30 days per month and 139 hours per week. Ninety-four percent of patients and 86% of households report that they could consult every time they visited the facility. These results suggest that health facilities are generally open and that the margin of improvement in this domain is almost nonexistent.

Out of a list of 23 potentially offered health services, the typical health facility offers 14 services[24].

[23] Figure 4 shows that prices for non-targeted services in the PBF group are close to the prices in the fixed payment group at all parts of the distribution.

[24] Curative consultations, pre and postnatal visits, birth delivery and preschool consultations are offered by more

PBF health facilities offer the same number of targeted and non-targeted services as in the fixed payment group. PBF did thus not induce changes in the extensive margin of service supply.

Staff Composition and Attendance On average, facilities counted 7 workers among which two-thirds were technical workers (doctors –only 3% of staff – nurses and birth-assistants) and one third non-technical workers (pharmacists, managers, secretaries, receptionists and maintenance workers). Contrary to the result of Bandiera et al. (2013) that team incentives change team composition in for-profit organizations, PBF did not change the quantity of staff nor the type of workers The finding that staff composition remained stable suggests that worker mobility was low and that financial incentives were not able to spur health workers into assortative matching by ability (as observed in the context of Bandiera et al. 2013). Indeed, the number of workers who left the facility in the last 12 months was found to be equal in both groups (0.76 workers).

However, we find higher staff attendance under PBF than under fixed-payment in the unannounced visits in July, August and September 2012: 58% in the fixed payment group versus 65% in the PBF group, a 14% increase, significant at the 10% level. Figure 9 shows the distribution of staff attendance at facilities by treatment status. We see that staff attendance is higher in the PBF group than in the fixed payment group at any point of the distribution, suggesting that incentivized facilities responded quite similarly to the incentive in terms of staff attendance. The higher attendance due to PBF echoes workers' statements in the qualitative interviews: "If we work a lot, we will have more money and conversely"; "We need to work many days and hours in order to have more patients".

Preventive Sessions at Facilities Incentivized workers organized more preventive sessions at facilities in the last 12 months than non-incentivized workers (120 instead of 100, although the difference is not significant). The difference is driven by targeted services (immunization, prenatal care and family planning) for which 74 preventive sessions were offered in the fixed payment group and 106 in the PBF group (a 43% increase significant at the 5% level). For non-targeted services (postnatal care and VIH prevention), the number of preventive sessions is also higher in the PBF group but the difference is not significant[25]. Figures 5 and 6 show the distributions of the number of

than 90% of health facilities. Immunization is offered by 88% and family planning by 84%. A smaller proportion of health facilities offer the other services.

[25]The supply for preventive sessions is already higher for targeted services than for non-targeted services (out of 100 preventive sessions in the last 12 months, 74 were devoted targeted services and 26 to non-targeted ones). PBF thus widened this gap.

preventive sessions for targeted and non-targeted services by treatment status: we see a reduction in the proportion of facilities offering few preventive sessions (less than 80 a year), and an increase in the proportion of facilities offering many sessions (above 100 a year). As a result, access to targeted health services is easier in the PBF group since a larger number of preventive sessions gives more opportunities to use the service. In contrast, access to non-targeted services remained the same in both groups.

Number of Outreach Activities in the Community The number of outreach activities by health workers in the community is higher in the PBF group: health workers performed an average of 22 visits to the community in the last 12 months, versus 15 in the fixed payment group, but this difference is not significant. In fact, the difference in the number of outreach activities is driven by targeted services: health workers made 16 visits to communities for these services in the PBF group, versus 10 in the fixed payment group (a 60% significant increase significant at the 10% level). In contrast, the difference in the number of outreach activities for non-targeted services is small and not significant. Figures 7 and 8 show the distributions of the number of outreach activities for targeted and non-targeted services by treatment status. We see a reduction in the proportion of health workers offering fewer than 50 visits to community for targeted services. The population in the catchement area of a PBF facility was thus better informed about the preventive sessions organized at that facility than the population in the catchement area in a fixed payment facility.

Conclusion on Accessibility of Health Services Overall, access to health services was improved in PBF facilities through a decrease in user fees, an increase in staff attendance, a larger supply for preventive sessions organized at the facility, and a greater number of visits to the community.

5.3 Service Quality

Table 3 presents the effects of PBF on service quality: technical quality, patient understanding of diagnosis and prescriptions, and patient satisfaction. While analyzing the results on quality, it is important to keep in mind that this PBF approach did not tie payments to any quality measures.

Technical Quality of Health Services On average, patients straight out of consultation reported 16-minute consultations in the fixed payment group and PBF did not impact this con-

sultation time. Household members who visited in the last 12 months reported a slightly longer consultation time for their last consultation in the PBF group (19 minutes) than in the fixed payment group (16 minutes) (the difference is significant at the 1% level). This finding dispels the fear that incentives based on the quantity of health services would imply maximizing the number of patients at the expense of time spent with each of them.

The average compliance rates with standard medical procedures were low: 32% for curative visits, 67% for prenatal visits, and 62% of postnatal visits. Forty-nine percent of patients straight out of consultation also reported that drugs were prescribed without their being examined. On average, women stayed 2.3 days in the health facility after giving birth. These measures of technical quality were not affected by PBF. Our data thus show that technical quality is poor and PBF had no impact on it.

Users' Understanding and Satisfaction Users' understanding of the diagnosis and medication seems high and also unaffected by PBF: 82% of patients straight out of consultation and 94% of household members who visited in the last 12 months understood the diagnosis, and 90% of patients were aware of what drugs they were supposed to take. Users' satisfaction is also very high (94% among patients straight out of consultation and 91% among household members who visited in the last 12 months) and unaffected by PBF. The main reason for being satisfied or dissatisfied of the visit is care quality (Table 3 reports only reasons related to service quality[26]).

Conclusion on Service Quality Together, these results suggest that health workers did not reduce quality in response to the volume-based incentive, nor improved quality to attract more patients. The reason why service quality is not elastic to incentives might be that quality depends on worker background, education and skills that might be difficult to change. It might also reflect the fact that health workers are not sure how to improve service quality, and/or not sure what aspect of service quality users observe and value. Low technical quality suggests that there are margins for improvement in that domain, but it is unclear that health workers can make such improvements and that users will be able to observe and be responsive to it. This experiment shows that service quality does not respond to an incentive which is tied to patient volume.

[26]Other reasons for being satisfied or disatisfied by the visit were: price and distance. They both account for a very small proportion of satisfactions, and an even smaller proportion of dissatisfaction.

5.4 Utilization of Health Services

Table 4 presents the effects of PBF on service utilization. We find a small decrease in health service utilization, in particular for services whose price was reduced.

In the control group, 50% of people visited a health facility in the last 12 months. PBF reduced this proportion to 45%, a difference which is significant at the 1% level. One reason could be that people had less reason to visit for illness but data show that the reduction in visits occurred among people who have been sick in the last 12 months: the proportion of sick people who visited a facility was reduced from 62% in the fixed payment group to 55% in the PBF group (significant at the 10% level). This finding is particularly striking given that the price of curative consultation and drugs was reduced in incentivized facilities. We also find that sick people waited almost three days with symptoms before they visited and that PBF did not change utilization on this intensive margin.

We do not find any effect of PBF on utilization of immunization. . Eighty-five percent of children aged 0-5 received at least one immunization shot based on mother declaration and the enumerators could see the TB immunization scar on the shoulder of 60% of children. When the immunization card could be seen, the average number of immunization shots was 2.7[27].

Take-up for prenatal visits was lower in the PBF group. In the fixed payment group, women who have been pregnant in the last 12 months visited 3.3 times for a prenatal consultation on average. While this proportion is lower but not significantly different in the PBF group over the last 12 months, the difference is significant when we focus on the period when PBF was in place: for women who gave birth before September 2012, the number of prenatal visits is 11% lower in the PBF group than in the fixed payment group. We don't find any substitution of modern medicine for traditional medicine: the same reduction in the number of prenatal visits is found when we include visits to healers and den mothers[28]. Despite a reduction in the price of prenatal consultation, we thus find a clear reduction in take-up of prenatal care due to PBF.

Utilization of attended delivery, postnatal consultation, and family planning were not changed by PBF. Eighty-two percent of births were attended in a health facility over the last 12 months, and women visited on average only once for a postnatal consultation. Only 5%of women aged 15-49 were using a modern contraceptive method[29] and the fertility rate was very high: 35% of

[27]Full immunization requires 5 shots, ideally before the child turns 2.

[28]The negligible increase in the number of visits when we include healers and den mothers suggest a small utilization of traditional medicine for prenatal care.

[29]Modern contraceptive methods are pill, shot, condom, IUD, spermicidal, implant and sterilization.

women aged 15-49 had been pregnant in the last 12 months (here we consider only women from the representative sample since all women in the other sample have been pregnant in the last 12 months by construction).

Finally, women who gave birth before September 2012, during the pilot, were more likely to have a c-section. The effect size is large: while 1.1% of women had a c-section in the fixed payment group, 4% of women had a c-section in the PBF group. This result suggests that doctors performed strategically more c-sections in response to the incentive, which should be related to the fact that c-section brings a huge amount of points in this particular PBF system (see Appendix Table 1). The consequence in terms of public health is not clear since c-sections should be used to avoid difficult deliveries, although it might also be unjustified and cause negative outcomes for the mother and the child. Health outcomes are covered in the next section.

Conclusion on Utilization of Health Services Overall, we find that PBF decreased utilization of curative and prenatal services, and did not change utilization of other services. This result is surprising given the effort made by health workers to improve the accessibility of targeted services, and the price reduction of curative and prenatal services. Our data show that the average health worker spent 10 hours in the facility the day before the survey, and consulted six patients spending 16 minutes with each. That means much idle time during the day. The important result is that the incentivized health workers could not identify and address the barriers to health service uptake, which might be particularly severe in developing regions like Haut-Katanga. Classic and sensible strategies to improve service uptake proved counterproductive. Positive price-elasticity of demand for health suggests that prices of health services work as signals for the value of health services, and that people might be reluctant to use health services when they do not fully understand its benefits. Preventive sessions and outreach activities would help only if demand is discouraged by uncertainty about service availability. But in the context of Haut-Katanga, it seems that more than eased access is needed. The price signal of quality suggests that the population needs information about the benefits of health services. In contrast, Dupas (2014) shows that lower prices of bednets lead to higher adoption. However, households who were offered a subsidy to acquiere a bednet in this experiment were informed about the market value of the bednet, which was printed on the voucher. It means that households knew both the value of the bednet and the price reduction associated with the subsidy. Prices thus did not work as a signal for quality. Our findings add to this literature

by showing that prices work as signal for quality when the population is poor, low-educated and generally uninformed about modern medicine, suggesting that price reduction should be combined with information on the value of health products.

5.5 Population Health Status

Table 5 presents the effect of PBF on health outcomes. We find a deterioration in newborn and child health: height-for-age and weight-for-age z-scores were found to be lower in the in the catchment area of a PBF facility than in the catchment area of a fixed payment facility, as well as the proportion of children born in the last 12 months that were still alive. The effect size is substantial: the means of weight-for-age and height-for-age in the PBF group are 0.18 standard deviations below the mean in the fixed payment group, and the proportion of newborns who did not survive in the catchment area of a PBF facility is twice as big as in the in the catchment area of a fixed payment facility. The estimates are less precise when we focus on children born before September 2012, when PBF was in place, but the point estimates are of similar magnitudes. We don't find that PBF affected the overall mortality rate in the households, nor mortality of women who gave birth in the last 12 months and children aged 0-5. The negative impact of PBF on mortality thus concentrated on children born in the last 12 months. The negative effects on newborn mortality and child weight and height might be related to the negative impact PBF on prenatal care utilization.

5.6 Health Facility Revenue

Table 6 presents the effects of PBF on facility resources, worker's payment, and the overall quality of facility infrastructure and equipment.

Total Resources at the Facility Level We find 45% less total resources in the hands of PBF health facilities than fixed payment health facilities the month before the survey (the difference is significant at the 5% level). The reduction in facility resources comes from a 56% reduction in revenue from users (significant at the 10% level). In contrast, we don't observe any difference in revenue from the government or NGOs. This result is consistent with our previous findings that PBF led to lower user fees and price of drugs, and lower service utilization, than fixed payments. According to the qualitative interviews, incentivized health workers who reduced their fees to increase demand found themselves in a situation where they were not able to re-adjust their price

schedule and raise prices back to their initial values as the population had become accustomed to the reduced prices and they were fearful of reducing demand to even lower levels.

Workers' Payment As a consequence, salary to health workers was significantly lower in PBF health facilities than in fixed payment ones. We find a 35% reduction in workers' total payment in the last month as reported by the facility head, and a 28% decrease as reported by the health workers (significant at the 10% and 5% level respectively). Wages from the government are not lower in the PBF and the fixed payment group, but payment from the facility itself is lower. This result is consistent with the reduced user fees and drug prices observed in the PBF health facilities' revenue.

Quality of the Facility Infrastructure and Equipment We find a significant negative impact of PBF on the quantity and quality of equipment and infrastructure. The mean *quality index* in the PBF group is 0.35 standard deviations below the mean in the fixed payment group. Most of the twelve items included in this index indicate a lower quality of equipment in the PBF facilities - negative differences are significant for four items: sink, clean towels, sterilization material and the availability of an examination table[30]. Furthermore, the mean *equipment index* in the PBF group is 0.29 standard deviations below the mean in the fixed payment group. The components of this index show that PBF facilities have consistently less equipment than the comparison ones. The differences are significant for four medical equipments: microscope, gynecological examination table, fridge and fuel. The day of the survey, the enumerator also checked the availability of five common vaccines[31] and nine common drugs[32]. We find less-than perfect –although not so bad- availability of these products: four out of five vaccines and seven out of nine drugs were available in the health facility the day of the survey. The PBF had a negative impact on the availability of vaccines the day of the survey, with fewer than 3.5 out of five vaccines available in the PBF group, but no impact on the availability of drugs, nor on the *infrastructure index*. The negative effects of PBF on the quality index, the equipment index and the availability of vaccines are likely to be related to the reduced revenue in the PBF group. Because of the lack of resources, PBF health facilities had difficulties in investing in new equipment and repairing existing equipment.

[30]However, it is worth noting that PBF facilities are more likely to permanently display the user fees and drugs' costs in the facility.

[31]Vaccines: DTP, Poliomyelitis, BCG, Measles and Yellow Fever.

[32]Drugs: oral rehydration salts, paracetamol, co-trimoxazole, ampicillin, metronidazole, quinine sulfate, mebendazole, tetracycline and Ringer's solution.

Conclusion on Health Facility Revenue Overall, PBF had a significant and substantial negative impact on health facilities' resources, health workers salaries and health facilities' infrastructure and equipment.

5.7 Staff Well-Being and Motivation

5.7.1 Staff Well-Being

Table 7 presents the effect of PBF on staff well-being, measured by staff satisfaction, conflicts, stress, and anxiety. All these outcomes are based on self-reported information so it is clearly subjective. Since we do not see any reason why social desirability bias would be different in the PBF and in the fixed payment group, the comparison between the two groups gives evidence on how PBF affected staff subjective well-being.

Job Satisfaction PBF induced a 14% significant decrease in the job satisfaction of facility staff – going from 5.7 to 4.9 on a scale from 0 to 10. In the qualitative interviews, many health workers from the PBF group complained about the PBF system and the frustration they had from the inefficiency of their strong efforts to increase the demand: "If there is no patient, we can't do more than working 26 days". The lower job satisfaction is also plausibly related to the unsuccessful effort and reduced worker salary.

Conflicts, Stress and Anxiety Overall, more than half of the facility staff found their workload heavy (53%), felt that they had too much work last week (61%) and that they have been tired (56%) in the last seven days. The PBF decreased significantly the perceived workload and fatigue: these three indicators decreased by respectively 11%, 28% and 16% (these last two decreases being significant at the 5 and 10% level). As shown in the previous section, the *effective* workload was actually reduced in the PBF group. The change in perception could also be due to the disappointing impact of the effort facility staff made to increase the number of patients: the increased effort to attract patients made the lack of demand for health services more salient, which could also have contributed to the lower subjective workload. Overall, the mean Subjective Workload index in the PBF group is 0.19 standard deviations below the mean in the fixed payment group.

As PBF led to lower and more uncertain payments than fixed payments, we tested if it caused more stress for the workers. In the fixed payment group, 34% of the workers declared that they worry about the volatility of their remuneration, and PBF did not significantly affect this proportion.

Importantly, workers worried about the level of their remuneration as much in the PBF group as in the fixed payment group (41%). It is surprising that the reduction in worker payment did not translate to an increase in the proportion of workers who were worried about the level of their remuneration.

As for conflicts, PBF did not change workers' perception of competition between facilities, nor the level of conflicts within facilities.Thirty-six percent of facility staff reported that the facility was in competition with other health facilities in both the fixed payment group and the PBF group. Workers reported a low level of conflict within the facility (1.72 on a scale from 0 to 10), with again no impact of PBF. Only 8% of workers reported that the allocation of government payment was a source of conflict in the facility in the fixed payment group, and 12% in the PBF group, a difference which is not statistically significant.

Overall we observe a lower perceived workload and a lower satisfaction in the PBF group, both likely due to the disappointing impact of increased effort on the number of patients and worker salary. However, PBF did not create more stress, anxiety or conflicts in the facilities.

5.7.2 Staff Motivation

Table 7 also presents the effect of PBF on staff motivation, as measured by free-riding, attendance after incentives were withdrawn and importance attached to job remuneration. As shown in Table 2, workers' effort to attract patients is larger under performance-based payment than fixed-payment, which we interpret as an increase in worker motivation due the incentive. In this section, we look more precisely at the effect of PBF on the nature of motivation.

Free-riding We don't find evidence that the collective nature of the incentive led to free-riding. First, the proportion of agents who did not do any outreach activities remained equal in both groups (48%). Second, the 25th, 50th and 75th percentiles of the number of outreach activities among agents who did some activities all increased due to PBF, which means that effort increased all over the counterfactual distribution. Finally, the standard deviation of the number of outreach activities per agent at the facility level is larger in the PBF group, but not statistically different. We cannot reject the hypothesis that the effect of PBF on worker effort is a pure upward translation for all workers. Altogether, these findings suggest that workers did not free-ride on others' effort.

Staff Attendance After the Incentives were Removed The positive effect of the incentive on staff attendance when incentives were in place reversed after incentives were withdrawn. The interviewers did not announce the day they would arrive in the facility for the endline survey to avoid manipulation of staff attendance. The attendance rate in the fixed payment group was 57%, similar to before the payment was withdrawn. This finding suggests that the termination of fixed payments did not affect staff effort, leaving intrinsic motivation intact. In contrast, a striking reversal happened in the PBF group: the attendance rate was at 65% before the incentive was withdrawn while only 45% after. This represents a substantial and statistically significant (at the 5% level) difference in the number of workers observed by the interviewer when s/he arrived: 3.8 in the comparison group while only 2.5 in the PBF group. Figure 10 shows the distribution of staff attendance at facilities after the pilot by treatment status. Staff attendance is lower in the PBF group than in the fixed payment group at any point of the distribution, suggesting again that workers responded quite similarly to the termination of the incentive[33].

The financial incentive thus induced higher worker motivation compared to fixed payments as long as the incentives were in place, but lower motivation after the incentive was withdrawn. It is important to keep in mind that payments from the government stopped in both the PBF and the fixed payment groups at the same time, which represents the same average reduction in health facilities' revenue by design. The reversal in staff attendance difference between PBF and fixed payment facilities thus happened in a context where facilities' revenue decreased by the same amount. In addition, the lower staff attendance after the incentives were withdrawn is unlikely to be due to the lower worker salary induced by PBF: salaries were already lower before the incentives stopped[34], whereas staff attendance was found higher at that time. Therefore, the reason for the reversal is likely to be driven by a change in levels of worker intrinsic motivation.

Attention Paid to Material Benefits In the fixed payment group, 38% of workers mention spontaneously remuneration or material comfort as the main advantage or disadvantage of their position, as opposed to non material benefits (like social recognition or health benefits to the population). This proportion increases dramatically to 51% in the PBF group (a 34% increase significant at the 5% level). This finding suggests that exposure to PBF changed the salience of financial motives in health workers' mind. Importantly, this change is also unlikely to be driven by

[33]This result is consistent with declarative data from the workers: worker attendance rate in the last seven days is found 78% in the fixed payment group while 71% in the PBF group (p-value of the test of equality 0.04).

[34]Tables 1 and 4 show that prices and utilization went down before the end of the pilot.

the decrease in worker salary since we observe an increase from 11% to 17% (significant at the 10% level) in the proportion of workers who mention financial benefits as the main *advantage*, while a smaller and insignificant increase in the proportion of workers who mention financial benefits as the main *disadvantage* (from 29% to 35%, p-value 0.15). This finding gives evidence of a shift in attention from the intrinsic value that the worker attributes to her job to its material benefits. We interpret this effect as evidence that incentives change the locus of control from internal to external by increasing the weight of external motives relative to intrinsic motives in worker utility.

Conclusions on Staff Motivation To summarize the effects of PBF on worker motivation, we find that: 1) When PBF is in place, worker's motivation was higher under PBF ($U_0 < U_1$) despite the fact that worker salaries from user fees were lower than under fixed payments ($F_1 < F_0$); 2) After PBF is withdrawn, worker motivation is lower in the group that was exposed to PBF ($U_2 < U_0$); and 3) Attention paid to financial motives relative to intrinsic motives is larger among workers who have been exposed to incentives ($\alpha_0 < \alpha_1$). Together, these effects give evidence that incentives reduce worker intrinsic motivation.

6 Conclusion

This study compares a performance-based payment mechanism to a fixed payment mechanism for health care providers in the district of Haut-Katanga, DRC. The performance-based payment studied was conditional on the number of patients for some pre-determined services, which is one specific approach of PBF. The findings show that the performance-based mechanism led to more effort by health workers to attract patients for the services included in the performance measure, without crowding out non-targeted services and service quality, nor generating conflicts, score manipulation or free-riding within the facilities. However, the increased effort from the health workers was associated with a smaller utilization of health services by the population, leading to a very disappointing reduction in facility revenue and worker income, as well as to a deterioration in newborn and child health outcomes. These findings suggest that existing health workers cannot be treated as entrepreneurs as they were not able to identify the successful strategies to increase demand for health services[35]. Importantly, we also find that PBF created a shift in workers' attention from

[35]New health workers informed about the payment system who would have self-selected in the health sector might be more able to develop appropriate strategies, as suggested by the results of Ashraf, Bandiera and Lee (2014) on the effects of career incentives on selection in the public health sector.

non-financial to financial motives apart from the reduction in worker income, and that workers decreased their effort after the removal of the incentives below its non-incentivized level.

In terms of policy lessons, these findings suggest that financial incentives should be used as a permanent policy rather than a temporary policy in order to limit the adverse effects of the motivational shift, and only in situations where the task is easy so that workers have the capacity to carry out the rewarded output, which might be a challenge in non-profit sectors where users' rationality is non standard. Indeed, the lack of response of the population challenges the idea that the demand for health services is normal: substantial decreases in prices deterred demand for health services, and improved accessibility was not able to encourage more demand. Specific interventions to stimulate demand for health may be combined with supply-side interventions like PBF. One possibility would be thus to include service quality in the set of purchased performances as it was done in Rwanda (Basinga et al. 2011), with the hope that health providers would have the capacity to make quality improvements that would be observed by the population and would attract more patients. Alternatively, interventions to improve awareness about the benefits of health products or to help people overcome behavioral issues like procrastination could supplement a PBF mechanism.

References

[1] Ariely D., U. Gneezy, G. Loewenstein and N. Mazar. 2009. "Large Stakes and Big Mistakes", Review of Economic Studies, 76: 451–469.

[2] Ashraf, Nava, Oriana Bandiera and Kesley Jack. 2014. "No margin, no mission? A field experiment on incentives for public service delivery", *forthcoming* Journal of Public Economics.

[3] Ashraf, Nava, Oriana Bandiera, Scott Lee. "Do-gooders and Go-getters: carreer incentives, selection and performance in public service delivery", Working Paper, July 2014.

[4] Bandiera, Oriana, Iwan Barankay and Imran Rasul. 2005. "Social Preferences and the Response to Incentives: Evidence from Personnel Data", The Quarterly Journal of Economics, 120(3): 917-962.

[5] Bandiera, Oriana, Iwan Barankay and Imran Rasul. 2007. "Incentives for Managers and Inequality among Workers: Evidence from a Firm-Level Experiment", The Quarterly Journal of Economics, 122(2): 729-773.

[6] Bandiera, Oriana, Iwan Barankay and Imran Rasul. 2013. "Team Incentives: Evidence from a Firm Level Experiment", The Journal of the European Economic Association, 11(5): 1079-1114.

[7] Banerjee A., E. Duflo and R. Glennester, 2008, "Putting a Band-Aid on a Corpse: Incentives for Nurses in the Indian Public Health Care System", Journal of European Economic Association, 6(2-3), 487-500.

[8] Basinga P., P. Gertler, A. Binagwaho, A.s Soucat, J. Sturdy, C. Vermeersch, 2011, "Effect on Maternal and Child Health Services in Rwanda of Payment to Primary Health-Care Providers for Performance: an Impact Evaluation", The Lancet, 377(9775): 1421 - 1428.

[9] Baumeister, R. F. 1984. "Choking Under Pressure: Self-Consciousness and Paradoxical Effects of Incentives on Skillful Performance", Journal of Personality and Social Psychology, 46 (3), 610–620.

[10] Benabou, R. and J. Tirole. 2003. "Intrinsic and Extrinsic Motivation", The Review of Economic Studies, 70(3): 489-520.

[11] Benabou, R. and J. Tirole. 2006. Incentives and Prosocial Behavior", The American Economic Review, 96(5): 1652-1678.

[12] Bertone MP et al. 2011. "Revue des expériences de financement base sur les résultats in République Démocratique du Congo", USAID-HS2020.

[13] Chee G, N. Hsi, K. Carlson, S Chankova, P.Taylor, 2007, "Evaluation of the first five years of GAVI immunization services support funding", Gavi alliance report.

[14] Christianson, J.B., S. Leatherman and K. Sutherland. 2008. "Lessons from Evaluations of Purchaser Pay-for-Performance Programs: A Review of the Evidence", Medical Care Research and Review, 65(6supp): 5S-35S.

[15] Deci ,E. 1975. Intrinsic Motivation. New York: Plenum Press.

[16] Deci E, Koestner R, Ryan RM. 1999. "A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation", Psychologycal Bulletin, 125(6):627-668.

[17] Deci, E. and R. Ryan. 1985. Intrinsic Motivation and Self-Determination in Human Behavior, New York; Plenum Press.

[18] De Walque, D., Gertler P., Baitista-Arredondo S., Kwan A., Vermeesch C., de Dieu Bizimina J., Binagwaho A., Condo J. 2013. "Using Provider Performance Rewards to Increase HIV testing and Counseling Services in Rwanda", World Bank Policy Research Working Paper n° 6364, Impact Evaluation Series 84, Washington DC: The World Bank.

[19] Dupas, Pascaline. 2014. "Short-Run Subsidies and Long-Run Adoption of New Health Products: Evidence from a Field Experiment", Econometrica, 82(1): 197-228.

[20] Eichler, R. and R. Levine. 2009. Performance Incentives for Global Health: Potential and Pitfalls. Washington DC: Center for Global Development.

[21] Eldridge C. and N. Palmer, 2009, "Performance-based payment:some reflections on the discourse evidence and unanswered questions", Health Policy and Planning, 24:160–166.

[22] Forsberg E, R Axelsson, B. Arnetz, 2001, "Financial incentives in health care. The impact of performance-based reimbursement", Health Policy, 58.

[23] Furth, R., 2006, "Zambia Pilot Study of Performance-Based Incentives", USAID http://www.qaproject.org/news/PDFs/ZambiaPerformancePilotStudyInitiatives.pdf

[24] Gertler P. and C. Vermeesch, 2013, "Using Performance Incentives to Improve Medical Care Productivity and Health Outcomes", unpublished manuscript.

[25] Glucksberg, S. 1962. "The influence of strength of drive on functional fixedness and perceptual recognition", Journal of Experimental Psychology, 63: 36–41

[26] Gneezy, U., S. Meier and P. Rey-Biel. 2011. "When and Why Incentives (Don't) Work to Modify Behavior". The Journal of Economic Perspectives, 25(4): 191-209.

[27] Holmström B. and P. Milgrom. 1991. "Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design", Journal of Law, Economics, & Organization, 7: 24-52.

[28] Kamenica, E. 2012. "Behavioral Economics and Psychology of Incentives", Annual Review of Economics, 4: 427-452.

[29] Kohn, A. 1993. Punished by Rewards, New York: Plenum Press.

[30] Lazear, E. 2000. "Performance, Pay and Productivity", American Economic Review, 90(5): 1346-1361.

[31] Lepper, M, D. Greene and R. Nisbett. 1973. "Undermining Children's Interest with Extrinsic Rewards: A Test of the "Overjustification Hypothesis"", Journal of Personality and social Psychology, 28: 129-137.

[32] Loevinsohn B, A Harding , 2005, "Buying results? Contracting for health service delivery in developing countries", The Lancet, 366(9486): 676-681.

[33] Miller, G. and K. Singer Babiarz. 2013. "Pay-For-Performance Incentives in Low and Middle-Income Country Health Programs", NBER Working Paper 18932.

[34] Meessen B, Kashala JP, Musango L, 2007, "Output-based payment to boost staff productivity in public health centres: contracting in Kabutare district, Rwanda", 85(2):108-15.

[35] Mullen, K., R. Franck and M. Rosenthal. 2010. "Can You Get What You Paid for? Pay-For-Performance and the Quality of Healthcare Providers", The RAND Journal of Economics, 41(1): 64-91.

[36] Olken Ben, Junko Onishi and Susan Wong, 2014. "Should Aid Reward Performance? Evidence From a Field Experiment on Health and Education in Indonesia", American Economic Journal: Applied Economics, 6(4): 1-34.

[37] Oxman, A.D. and A. Fretheim, 2009, "Can paying for results help to achieve the Millennium Development Goals? A critical review of selected evaluations of results-based financing", Journal of Evidence Based Medicine.

[38] Peabody JW, Shimkhada R, Quimbo S, Florentino J, Bacate MF, McCulloch C, Solon O., 2011, "Financial Incentives and Measurement Improved Physicians' Quality of Care in the Philippines." , Health Affairs 10 (4) 773-81.

[39] Rasul, Imran and Daniel Rogger. 2014. "Management of Bureaucrats and Public Service Delivery: Evidence from the Nigerien Civil Service", Working Paper.

[40] Rusa L, JD Ngirabega, W Janssen, 2009, Performance-based financing for better quality of services in Rwandan health centres: 3-year experience, Tropical Medicine & International Health, Volume 14, Issue 7, pages 830–837.

[41] Rusa L, M Schneidman, G Fritsche, 2009, "Rwanda: Performance Based Financing in the Public Sector ", cgdev.org.

[42] Stabile, Mark and Sarah Thomson. 2014. "The Changing Role of Government in Financing Health Care: An International Perspective", Journal of Economic Literature, 52(2): 480-518.

[43] Soeters R, F Griffths, 2003, “Improving government health services through contract management: a case from Cambodia“, Health Policy and Planning, Oxford Univ Press.

[44] Soeters, R. Musango, L.Meessen, B., 2005, “Comparison of two output based schemes in Butare and Cyangugu provinces with two control provinces in Rwanda”, rapport pour le GPOBA, WB Ministry of health Rwanda.

[45] Soeters Robert, P.-B. Peerenboom, et al. 2011, “Performance Based Health Financing Experiment Improves Care in a Failed State.Performance-Based Financing Experiment Improved Health Care In The Democratic Republic Of Congo Health”, Health Affairs 30(:8): 1518-1527.

[46] Sondorp E, Palmer N, Strong L, Wali E. 2008. “Afghanistan: Paying NGOs for performance”. Washington, DC: Centre for Global Development.

[47] Souvorov A.. 2003. “Addiction to Rewards”, Mimeo, Gremaq, Toulouse School of Economics.

Figure 1: Performance-Based Financing in Africa

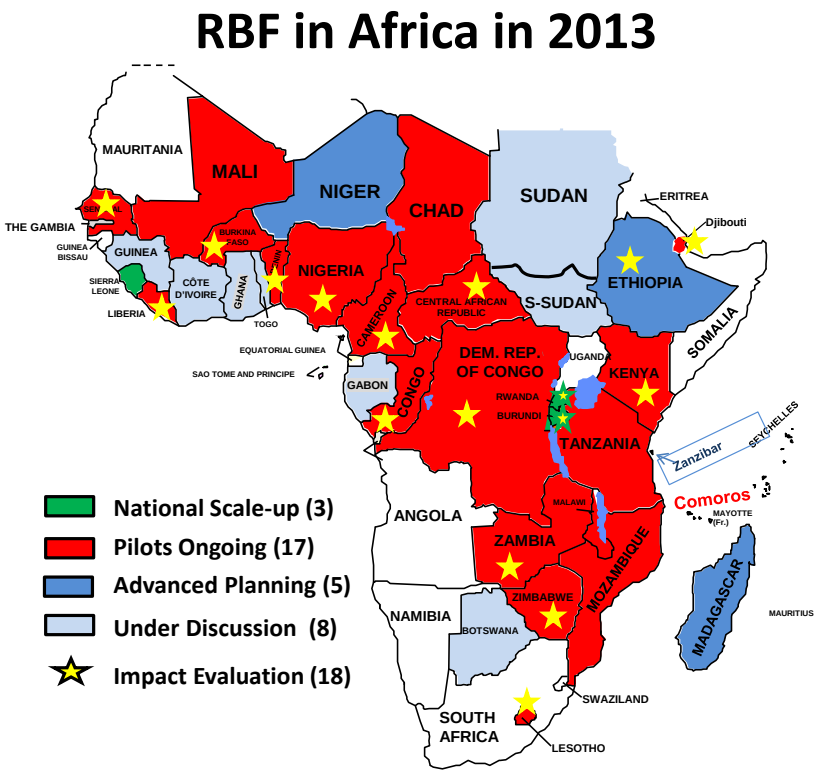

Figure 2: Monthly Payment Distribution, by Treatment Status

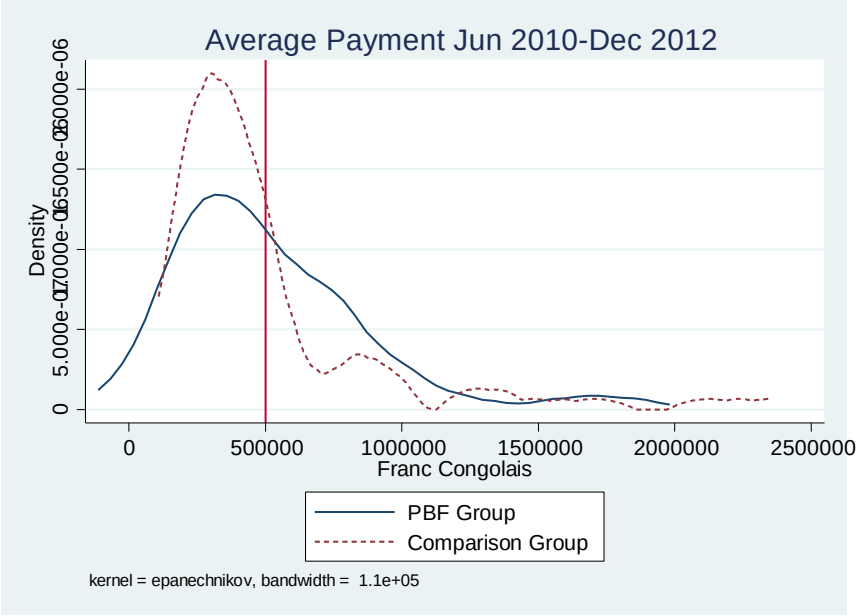

Figure 3: Distribution of the Fee Summary Index for Targeted Services, by Treatment Status

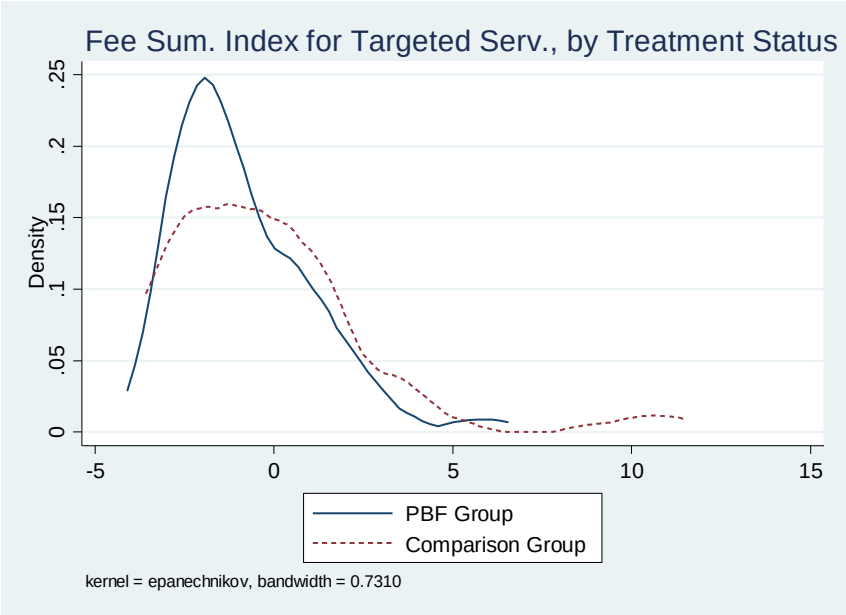

Figure 4: Distribution of the Fee Summary Index for Non-Targeted Services, by Treatment Status

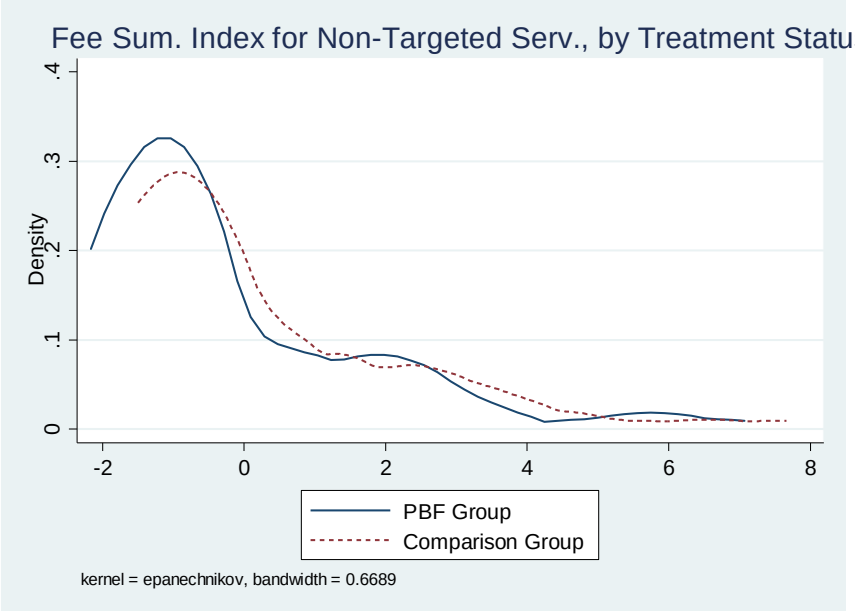

Figure 5: Distribution of Preventive Sessions Organized at Facilities for Targeted Services, by Treatment Status

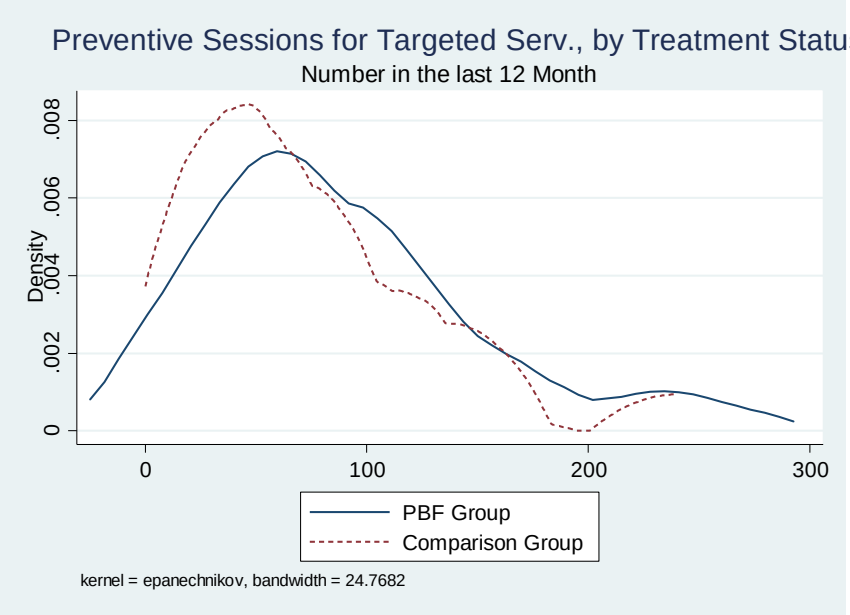

Figure 6: Distribution of Preventive Sessions Organized at Facilities for Non-Targeted Services, by Treatment Status

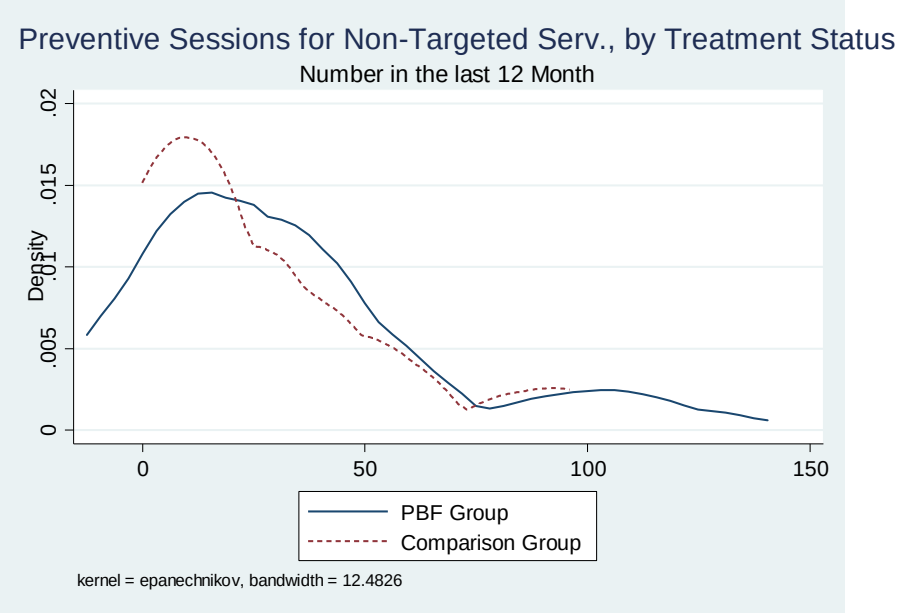

Figure 7: Distribution of Outreach Activities for Targeted Services, by Treatment Status

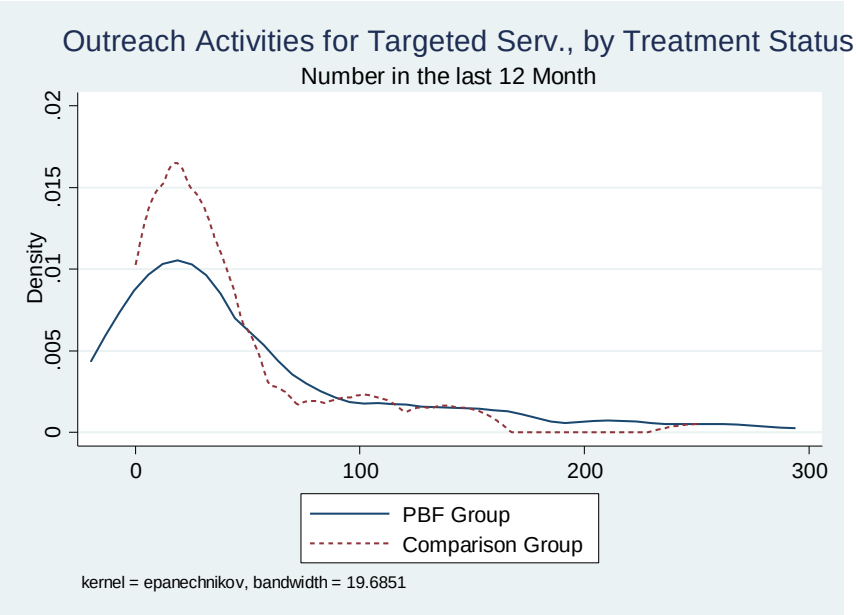

Figure 8: Distribution of Outreach Activities for Non-Targeted Services, by Treatment Status

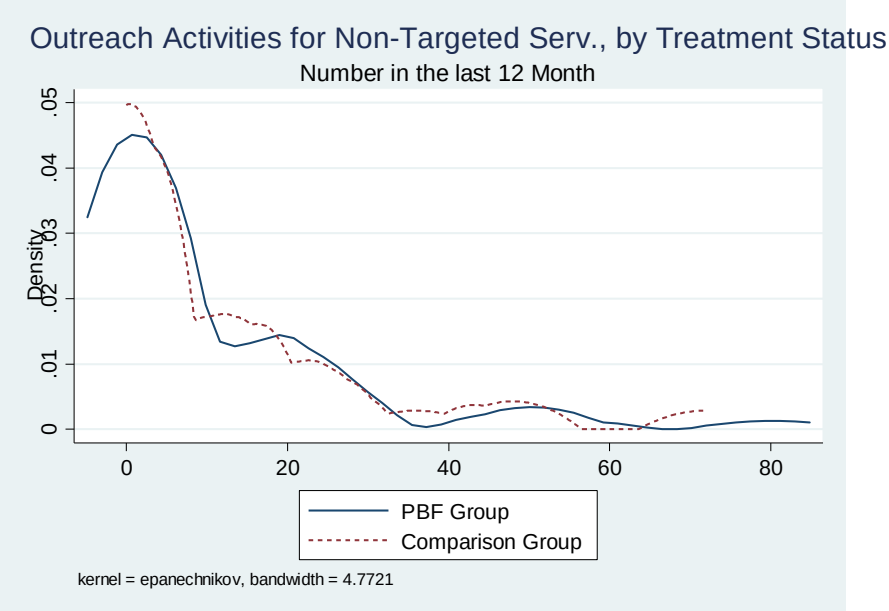

Figure 9: Distribution of Staff Attendance *during* the Pilot, by Treatment Status

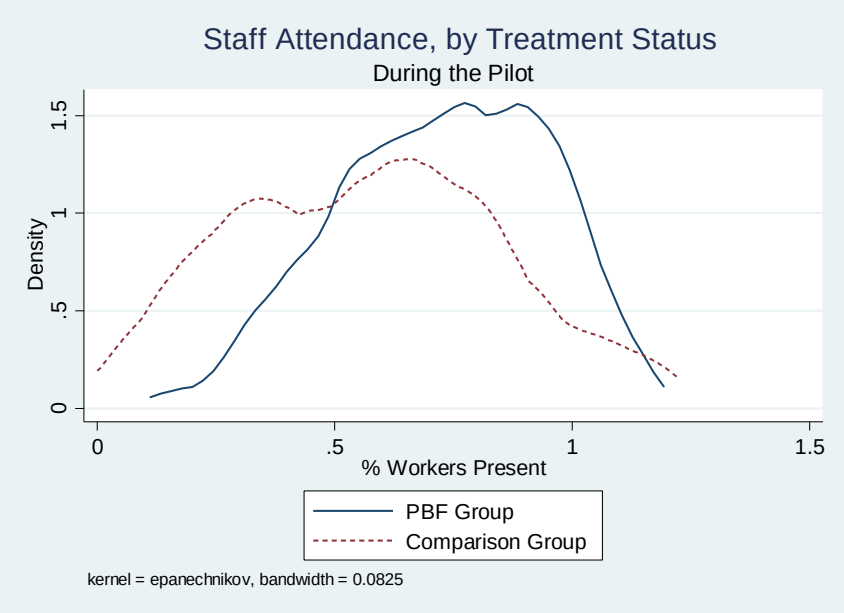

Figure 10: Distribution of Staff Attendance *after* the Pilot, by Treatment Status

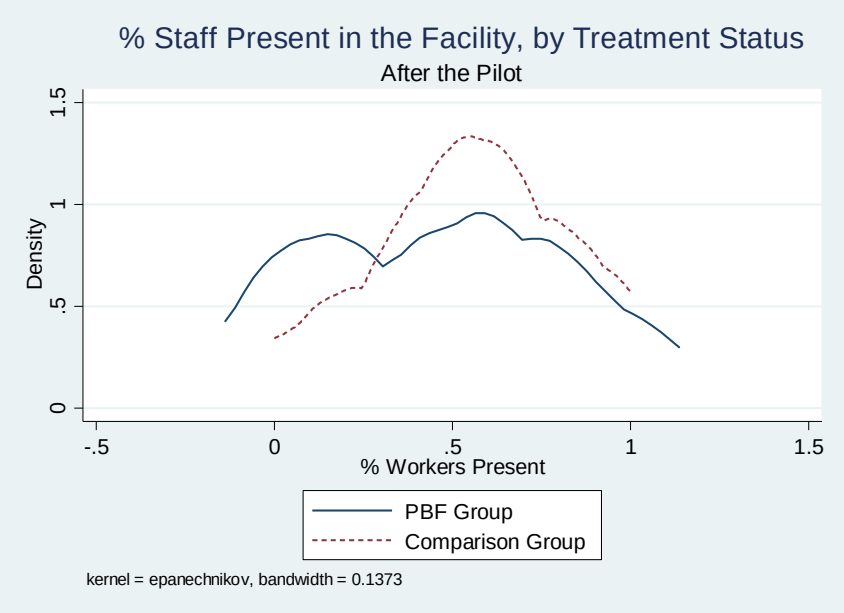

Table 1: Effects on User Fees

	Unit of Observation	Number of Observations	Average Treatment Effect (ATE)	p-value (ATE=0)	Mean of Dep. Var. (Control)	St.dev. of Dep. Var. (Control)
User fees as reported by facility heads at endline						
Fee Summary Index	Health Facility	93	-1.077	0.141	.166473	4.212105
Fee Summary Index, targeted services	Health Facility	109	-.807	0.061*	.0366889	2.866472
Fee Summary Index, non-targeted services	Health Facility	95	-.398	0.346	.1007338	2.064238
Targeted Services						
User fee for the first curative consultation (FC)	Health Facility	123	-692.45	0.281	1263.492	4557.316
User fee for delivery (FC)	Health Facility	113	-224.185	0.655	2747.414	2423.25
User fee for the first prenatal visit (FC)	Health Facility	118	-407.873	0.095*	850	1741.42
User fee for the second prenatal visit (FC)	Health Facility	115	-80.801	0.053*	132.2034	264.8622
Non-Targeted Services						
User fee for the second curative consultation (FC)	Health Facility	112	-178.082	0.18	459.4828	799.0377
User fee for postnatal visit (FC)	Health Facility	111	-57.43	0.386	105.3571	430.8215
User fee for preschool consultation (FC)	Health Facility	112	-6.718	0.838	86.66666	154.8281
User fees as reported by users in the last 12 months						
Fee the last postnatal visit (if any in the last 12 months) (FC)	Mother	388	-40.458	0.544	349.2684	585.9773
Fee for attended delivery (if any in the last 12 months) (FC)	Mother	762	-234.426	0.732	8768.171	6910.282
Fee the last prenatal visit (if any in the last 12 months) (FC)	Pregnant Women	918	-120.798	0.028**	583.4368	721.459
Fee the last immunization visit (if any in the last 12 months) (FC)	Children 0-5	2039	-22.096	0.237	87.71028	316.9161
Fee for the present curative visit (if the reason for visit was illness) (FC)	Patient	718	-1897.282	0.034**	7311.323	16030.53
Cost of drugs at the facility for the present curative visit (if the reason for visit was illness) (FC)	Patient	549	-1581.249	0.002***	3628.322	6160.657
User fees as reported by users *before September 2012*						
Fee the last postnatal visit (if any before September 2012) (FC)	Mother	227	-32.896	0.637	315.7232	539.3551
Fee for attended delivery (if any before September 2012) (FC)	Mother	493	463.057	0.546	8726.834	6926.556
Fee the last prenatal visit (if any before September 2012) (FC)	Pregnant Women	581	-187.611	0.001***	598.0456	706.0172
Fee the last immunization visit (if any before September 2012) (FC)	Children 0-5	508	18.014	0.489	73.16177	211.3834

Data Source: Endline survey. ***, **, * indicate significance at 1, 5, and 10%. Error terms are clustered at the health area level. We control for the urban/rural location of the unit of observation.
Fee Summary Index is the equally weighted average of z-scores of its components. The z-scores are calculated by subtracting the control group mean and dividing by the control group standard deviation. The components of the index are fees paid for first and second curative consultations, delivery, prenatal and postnatal visits, and preschool consultation.
Targeted services are: first curative consultation, delivery, and prenatal visits.
Non-targeted services are: second curative consultation, postnatal visit, and preschool consultation.
Unit of Observation: Pregnant Women = Women 15-49 who were pregnant in the last 12 months ; Mother = Women who gave birth in the last 12 months.

Table 2: Effects on Service Accessibility

Table 2: Effects on Service Accessibility

	Unit of Observation	Number of Observations	Average Treatment Effect (ATE)	p-value (ATE=0)	Mean of Dep. Var. (Control)	St.dev. of Dep. Var. (Control)
Health Facility Opening and Services Offered						
Opening						
Number of opening hours per week (as reported by the facility head)	Health Facility	116	-6.522	0.524	138.9262	47.86586
Number of opening days in the last month (as reported by the facility head)	Health Facility	119	-.139	0.816	29.73016	1.885482
The patient could consult each time s/he visited	Patient	993	-.019	0.322	.9375	.2422843
The household member could consult each time s/he visited	Household Member	4323	.016	0.351	.857081	.3500661
Services Offered at the Facility						
Number of services offered by the facility (between 0 and 23)	Health Facility	123	-.492	0.35	13.55556	3.644606
Number of targeted services offered by the facility (between 0 and 10)	Health Facility	123	-.141	0.606	7.730159	1.715267
Number of non-targeted health services offered by the facility (between 0 and 13)	Health Facility	123	-.351	0.329	5.825397	2.393133
Staff Composition and Effort						
Number of workers in the facility	Health Facility	121	-.772	0.215	6.725806	5.128415
% health workers in the facility	Health Facility	123	.027	0.425	.683401	.1826084
% doctors in the facility	Health Facility	123	-.001	0.933	.0271569	.0575394
Number of workers who left the facility in the last 12 months	Health Facility	123	-.009	0.972	.7619048	1.011455
Av. % workers present in the facility on unannounced visits 1, 2 and 3	Health Facility	138	.074	0.067*	.5807223	.2924829
Preventive Sessions Organized at the Facility in the Last 12 Months						
Number of preventive sessions at facility provided in the last 12 months	Health Facility	118	20.084	0.291	100.4426	82.87933
Number of preventive sessions at facility for targeted services provided in the last 12 months	Health Facility	119	31.542	0.044**	73.91803	57.09679
Number of preventive sessions at facility for non-targeted services provided in the last 12 months	Health Facility	120	10.808	0.107	26.87097	31.89197
Outreach Activities by Health Workers in the Last 12 Months						
Number of outreach activities in the last 12 months	Health Worker	326	7.184	0.171	15.23295	44.47532
Number of outreach activities for targeted services in the last 12 months	Health Worker	326	5.976	0.096*	9.829545	26.42281
Number of outreach activities for non-targeted services in the last 12 months	Health Worker	326	1.208	0.523	5.403409	19.53698

Data Source: Endline survey. ***, **, * indicate significance at 1, 5, and 10%. Error terms are clustered at the health area level. We control for the urban/rural location of the unit of observation.
Preventive sessions include: immunization, prenatal care and family planning (targeted services), postnatal care and HIV prevention (non-targeted services).
Outreach activities include: immunization, prenatal care and family planning (targeted services), postnatal care and HIV prevention (non-targeted services).

Table 3: Effects on Service Quality

	Unit of Observation	Number of Observations	Average Treatment Effect (ATE)	p-value (ATE=0)	Mean of Dep. Var. (Control)	St.dev. of Dep. Var. (Control)
Technical Quality						
Consultation time (minutes)	Patient	974	1.028	0.422	16.09263	15.51822
If visited a health facility, consultation time (minutes)	Household Member	4265	2.506	0***	16.125	11.96238
If visited for illness, compliance rate with medical procedure for the present curative consultation	Patient	713	-.008	0.805	.3239075	.2992397
If visited for illness, drugs were prescribed to the patient and the patient was not examined	Patient	719	-.036	0.495	.4936387	.5005968
If delivery attended, number of days in the health facility after the delivery	Mother	767	-.077	0.689	2.313283	1.702673
If any postnatal visit, compliance rate with medical procedures for the last postnatal consultation	Mother	389	.048	0.123	.6166667	.258334
If any prenatal visit, compliance rate with medical procedures for the last prenatal consultation	Pregnant Women	923	.004	0.818	.6657578	.1680248
Users' Understanding						
If visited for illness, patient understands diagnosis and next steps	Patient	720	-.001	0.971	.822335	.3827164
If visited for illness, patient knows what drugs to be taken	Patient	718	-.046	0.187	.8982188	.3027457
If visited a facility, household member understands diagnosis	Household Member	4258	.017	0.241	.9372237	.2426138
Users' Satisfaction						
The Patient reports that s/he was…						
satisfied	Patient	994	.013	0.359	.9430147	.2320279
satisfied thanks to care quality	Patient	990	.003	0.937	.5722222	.4952152
satisfied thanks to welcome quality	Patient	990	-.027	0.442	.2796296	.4492334
satisfied thanks to equipment quality	Patient	990	0	0.997	.0333333	.1796719
dissatisfied because of care quality	Patient	993	-.005	0.671	.0349265	.1837626
dissatisfied because of welcome quality	Patient	993	0	0.946	.0073529	.0855121
dissatisfied because of equipment quality	Patient	993	-.006	0.359	.0110294	.1045364
If visited a health facility, the Household Member reports that s/he was…						
satisfied	Household Member	4326	.004	0.778	.9142857	.2800023
satisfied thanks to care quality	Household Member	4318	-.005	0.857	.7417678	.4377572
satisfied thanks to welcome quality	Household Member	4318	-.008	0.547	.0836222	.2768804
satisfied thanks to equipment quality	Household Member	4318	.001	0.855	.0186308	.1352467
dissatisfied because of care quality	Household Member	4312	-.002	0.853	.0487593	.2154112
dissatisfied because of welcome quality	Household Member	4312	-.001	0.844	.0104484	.1017042
dissatisfied because of equipment quality	Household Member	4312	.001	0.76	.008707	.0929245

Data Source: Endline survey. ***, **, * indicate significance at 1, 5, and 10%. Error terms are clustered at the health area level. We control for the urban/rural location of the health facility.
months.

Table 4: Effects on Service Utilization

	Unit of Observation	Number of Observations	Average Treatment Effect (ATE)	p-value (ATE=0)	Mean of Dep. Var. (Control)	St.dev. of Dep. Var. (Control)
Overall Utilization						
The household member visited a health facility in the last 12 months	Household Member	9113	-.051	0.006***	.4961274	.5000388
Curative Services						
If was sick in the last 12 months, the individual visited a health facility in the last 12 months	Household Member	5926	-.043	0.074*	.6225681	.4848229
If was sick and visited in the last 12 months, number of days before s/he visited a health facility	Household Member	3477	.014	0.916	2.70134	2.8628
Child Immunization						
Ever had an immunization shot	Children 0-5	2448	-.002	0.94	.8486739	.3585063
Has a scar from tuberculosis immunization	Children 0-5	2441	.016	0.677	.6	.4900902
If the immunization card seen by the interviewer, number of immunization shots	Children 0-5	833	-.023	0.961	2.706977	3.186173
If aged 15 months or older , ever had an immunization shot	Children 0-5	1415	.016	0.359	.9282759	.2582087
If aged 15 months or older , has a scar from tuberculosis immunization	Children 0-5	1411	.041	0.322	.6546463	.4758135
Prenatal Services						
Number of prenatal visits at a health facility	Pregnant Women	1120	-.281	0.14	3.357782	2.122774
Number of prenatal visits, including healers and den mothers	Pregnant Women	1117	-.292	0.13	3.482944	2.243731
If gave birth before Sept. 2012 , number of prenatal visits at a health facility	Pregnant Women	603	-.373	0.054*	3.467105	1.751634
If gave birth before Sept. 2012 , number of prenatal visits including healers and den mothers	Pregnant Women	607	-.422	0.044**	3.674267	1.919479
Delivery						
The mother delivered in a health facility	Mother	961	-.015	0.684	.8241309	.3810987
If delivery attended, had a C-section	Mother	773	.018	0.121	.0173697	.130807
If gave birth before Sept. 2012 , delivered at a health facility	Mother	624	-.021	0.623	.8285714	.3774827
If gave birth before Sept. 2012 and delivery attended, had a c-section	Mother	500	.029	0.071*	.0114943	.1067981
Postnatal Services						
Number of postnatal visits at a health facility	Mother	959	.055	0.622	.8650306	1.426543
Number of postnatal visits, including healers and den mothers	Mother	957	.058	0.655	1.10041	1.778309
If gave birth before Aug. 2012 , number of postnatal visits at a health facility	Mother	528	-.059	0.647	.8769231	1.282585
If gave birth before Aug. 2012 , number of postnatal visits, including healers and den mothers	Mother	523	.011	0.943	1.011719	1.387571
Family Planning						
Uses a modern contraceptive method	Women 15-49	1873	.005	0.69	.0505263	.2191437
Has been pregnant in the last 12 months (representative sample only)	Women 15-49	902	-.005	0.882	.3522976	.4782096

Data Source: Endline survey. ***, **, * indicate significance at 1, 5, and 10%. Error terms are clustered at the health area level. We control for the urban/rural location of the health facility.
Unit of Observation: Pregnant Women = Women 15-49 who were pregnant in the last 12 months ; Mother = Women who gave birth in the last 12 months.

Table 5: Effects on Health Outcomes

	Unit of Observation	Number of Observations	Average Treatment Effect (ATE)	p-value (ATE=0)	Mean of Dep. Var. (Control)	St.dev. of Dep. Var. (Control)
Newborn and Child Health						
Weight-for-age z-score	Children 0-5	2428	-.181	0.048**	-1.061262	1.741033
Height-for-age z-score	Children 0-5	2402	-.184	0.044**	-1.654884	1.864021
If gave birth in the last 12 months, her child is still alive	Mother	961	-.01	0.093*	.9897751	.1007032
If born before September 2012 , weight-for-age z-score	Children 0-5	2109	-.126	0.184	-1.305793	1.577741
If born before September 2012 , height-for-age z-score	Children 0-5	2087	-.177	0.073*	-1.842894	1.782993
If gave birth before Sept. 2012 , her child is still alive	Mother	624	-.012	0.203	.9936508	.0795549
Mortality						
Number of persons in the household who died in the last 12 months	Household	1708	.007	0.732	.1366313	.4006933
Number of women in the household who died for perinatal reasons in the last 12 months	Household	1707	-.004	0.427	.009434	.0967264
Number of children under 5 in the household who died in the last 12 month	Household	1707	.012	0.55	.0896226	.3171387

Data Source: Endline survey. ***, **, * indicate significance at 1, 5, and 10%. Error terms are clustered at the health area level. We control for the urban/rural location of the health facility.
Unit of Observation: Mother = Women who gave birth in the last 12 months.

Table 6: Effects on Facility Total Resources

	Unit of Observation	Number of Observations	Average Treatment Effect (ATE)	p-value (ATE=0)	Mean of Dep. Var. (Control)	St.dev. of Dep. Var. (Control)
Total Resources at the Facility Level						
Revenue from users	Health Facility	116	-258312.7	0.051*	463328.7	945376.8
Revenue from the government	Health Facility	121	6917.788	0.903	121674.5	250612.5
Revenue from NGOs and private donors	Health Facility	121	-121.757	0.555	435.4839	3429.003
Total revenue	Health Facility	116	-290462.7	0.023**	642700.7	1028920
Workers' Payment						
Payment to the Workers (reported by the Facility Head)						
Average total payment per worker in the last month (FC)	Health Facility	116	-17553.24	0.084*	50679.49	57950.96
Average wage from the government per worker in the last month (FC)	Health Facility	118	4660.341	0.083*	1731.591	5442.607
Average payment from the facility per worker in the last month (FC)	Health Facility	119	-12444.67	0.154	42580.55	47536.47
Payment to the Health Workers (reported by the Health Workers)						
Total payment in the last month (FC)	Health Worker	282	-35885.75	0.031**	127139.5	174494.9
Wage received from the government in the last month (FC)	Health Worker	326	-4999.407	0.5	23654.04	88004.44
Payment received from the facility in the last month (FC)	Health Worker	285	-28682.54	0.061*	102552.8	153866.8
Quality of the Facility Infrastructure and Equipment						
Quality index^ based on interviewers' observation (Principal Component Analysis)	Health Facility	116	-.525	0.014**	.1990995	1.511479
Infrastructure index^^ (Principal Component Analysis)	Health Facility	110	.184	0.372	-.1715342	1.425423
Equipment index^^^ (Principal Component Analysis)	Health Facility	116	-.639	0.026**	.052816	2.226755
Number of types of vaccine currently available (between 0 and 5)	Health Facility	118	-.744	0.034**	4.16129	1.738603
Number of types of vaccine that have been unavailable at some point in the last 12 months (between 0 and 5)	Health Facility	118	.036	0.929	1.52381	1.740014
Number of types of drug currently available (between 0 and 9)	Health Facility	117	.236	0.646	6.7	3.185241
Number of types of drug that have been missing once in the last 12 months (between 0 and 9)	Health Facility	111	-.276	0.589	5.333333	3.445148

Data Source: Endline survey. ***, **, * indicate significance at 1, 5, and 10%. Error terms are clustered at the health area level. We control for the urban/rural location of the unit of observation.

^The quality index includes observation on building quality, waiting room, consultation room, lavabo, soap, clean towels, bathrooms, sterilization material, permanent display of user fees and drugs' costs, use of an examination table and ordinogram.

^^The infrastructure index includes six items: phone ownership, motorized transportation mean ownership, access to clean water, toilet and electricity, and hard roof.

^^^The equipment index includes the quantity of fifteen types of medical equipment owned by the health facility: generator, sterilizer, tensiometer, stethoscope, baby-scales, weighing scale, height gauge, microscope, gynecological examination table, fridge, delivery boxes, fuel, kerosene, bed and solar panel.

Table 7: Effects on Staff Well-Being and Intrinsic Motivation

	Unit of Observation	Number of Observations	Average Treatment Effect (ATE)	p-value (ATE=0)	Mean of Dep. Var. (Control)	St.dev. of Dep. Var. (Control)
Job Satisfaction						
Level of satisfaction of the facility staff for his job (from 0 to 10)	Facility Staff	455	-.681	0.038**	5.705394	2.783944
Conflicts, Stress and Anxiety						
Facility staff founds his workload heavy in the last 7 days	Facility Staff	454	-.062	0.186	.5291666	.5001917
Facility staff reports too much work in the last 7 days	Facility Staff	444	-.128	0.014**	.6092437	.4889482
Facility staff felt tired due to the job in the last 7 days	Facility Staff	445	-.088	0.096*	.5606695	.4973471
Subjective Workload Index	Facility Staff	446	-.193	0.019**	-.0001799	.7782683
Facility staff worries about insecure / volatile remuneration	Facility Staff	451	.029	0.64	.3430962	.4757396
Facility staff worries about low remuneration	Facility Staff	451	-.013	0.815	.4142259	.4936216
Facility staff reports that the facility is in competition with other facilities	Facility Staff	454	-.06	0.292	.3583333	.4805129
Facility staff reports that payment allocation is a source of conflict in the facility	Facility Staff	453	.044	0.19	.0829876	.2764375
Level of conflicts among workers as perceived by facility staff (from 0 to 10)	Facility Staff	453	-.284	0.26	1.717842	2.203041
Free-riding						
% agents who did not do any outreach activity for targeted services in the last 12 months	Health Worker	326	-.004	0.947	0.482954	0.501135
Among agents who did some outreach activities for targeted services, 25th percentile	Health Worker	172	3	0.006***	2	na
Among agents who did some outreach activities for targeted services, 50th percentile	Health Worker	172	6	0.092*	6	na
Among agents who did some outreach activities for targeted services, 75th percentile	Health Worker	172	9	0.077*	24	na
Facility level stand. dev. of the number of outreach activities for targeted services per agent (if more than 1 agent)	Health Facility	87	9.083	0.136	14.74143	21.81978
Importance Attached to Job Remuneration						
Facility staff elicits financial benefits as the main advantage or disadvantage of his position	Facility Staff	454	.117	0.025**	.3833333	.4872145
Facility staff elicits financial benefits as the main advantage of his job	Facility Staff	452	.065	0.075*	.1087866	.3120247
Facility staff elicits financial benefits as the main disadvantage of his job	Facility Staff	454	.063	0.155	.2916667	.4554796
Staff Effort (Attendance) after PBF was withdrawn						
Number of workers in the facility on unnanounced visit 4 (enline survey)	Health Facility	123	-1.354	0.032**	3.84127	3.418198
% workers present in the facility on unannounced visit 4 (endline survey)	Health Facility	123	-.121	0.099*	.5741979	.3109018
Av. attendance rate of workers in the facility in the last 7 days (as reported by the facility head)	Health Facility	123	-.09	0.155	.7752835	.1929815
Attendance rate in the facility in the last 7 days (as reported by the Health Worker)	Health Worker	331	-.067	0.042**	.7799358	.1429585

Data Source: Endline survey. ***, **, * indicate significance at 1, 5, and 10%. Error terms are clustered at the health area level. We control for the urban/rural location of the health facility.
Unit of Observation: Facility Staff = health workers + facility head
Subjective Workload Index is the equally weighted average of z-scores of its components. The z-scores are calculated by subtracting the control group mean and dividing by the control group standard deviation.
The components of the index are dummies for facility staff founds his workload heavy, facility staff reports too much work in the last 7 days, and facility staff felt tired in the last 7 days.

Appendix Table 1: Relative Prices of Targeted Health Services

Service	Indicator	Relative Price (USD)
Services targeted at health centers and referral health centers		
Curative care	Per new curative consultation	$0.6
Institutional delivery	Per delivery at the health center	$5
Obstetric referral	Per pregnant woman referred to the referral center/hospital	$5
Full childhood immunization	Per fully immunized child	$3.5
Prenatal care	Per prenatal care consultation	$1.2
Tetanus toxoid vaccination	Per 5th dose of tetanus toxoid vaccination	$2
Family planning	Per woman that uses a modern method of family planning	$4.5
Additional services targeted only at referral health centers:		
Caesarean section	Per caesarean section delivery (and decision-tree has been followed)	$30
Blood transfusion, when appropriate	Per transfusion episode	$5
Obstetric referral	Per delivery referred to the referral center/ hospital"	$5

Appendix Table 2: Endline Sample

Endline Sample, by Payment Status

	PBF Group	Comparison Group	Total
Health areas	44	43	87
Health Facilities	60	63	123
Facility Staff	154	178	332
Patients	470	544	1,014
Households	859	849	1,708
Household members	4,578	4,656	9,234
Women 15-49	939	957	1,896
Pregnant Women*	571	560	1,131
Mothers**	479	489	968
Children 0-5	1,228	1,285	2,513

*Pregnant Women = Women who have been pregnant in the last 12 months
**Mothers = Women who gave birth in the last 12 months

Appendix Table 3: Descriptive Statistics at Endline

	Mean	Standard Deviation	Nb. of Observations
A. HEALTH FACILITY			
The facility is a "Centre de Santé de Référence"	0.11	0.31	123
The facility is a "Centre de Santé"	0.69	0.46	123
The facility is a "Poste de Santé"	0.20	0.40	123
The facility is public	0.66	0.48	123
The facility is religious	0.15	0.36	123
The facility is private	0.19	0.39	123
The facility is urban or semi-urban	0.17	0.38	123
The facility is rural	0.83	0.38	123
Served population size	12872.76	11570.57	123
B. HEALTH WORKERS			
The health worker is a female	0.57	0.50	332
Age of the health worker (years)	42.14	11.20	332
The health worker is a doctor	0.06	0.23	332
The health worker is a nurse	0.57	0.50	332
Number of years of experience	12.56	10.13	331
C. PATIENT			
The patient is a female	0.67	0.47	1006
Age of the patient (years)	18.61	17.39	1002
D. HOUSEHOLDS			
The household member is a female	0.50	0.50	9225
Age of the household member (years)	17.17	16.13	9135
The household member is literate (if aged 15 or more)	0.57	0.49	4166

Appendix Table 4: Balance Checks

	Unit of Observation	Number of Observations	Mean in the control group	Standard Deviation in the control group	Difference in mean in the treatment group	p-value (difference=0)
Health Facilities						
Health Center (versus Health Post)	Health Facility	129	.78125	.4166667	-.038	0.623
Public	Health Facility	129	.59375	.4950148	-.029	0.734
Private	Health Facility	129	.28125	.4531635	-.086	0.276
Religious	Health Facility	129	.125	.3333333	.115	0.119
Number of years of activity	Health Facility	122	20.18333	22.42539	.266	0.948
Catchement population	Health Facility	122	11129.3	15802.48	1255.156	0.669
Catchement area (km2)	Health Facility	109	368.963	826.58	-19.957	0.892
Number of beds	Health Facility	129	8.953125	13.43229	1.23	0.536
Number of workers	Health Facility	129	6.359375	5.524454	-.162	0.866
Infrastructure Index^	Health Facility	129	1.00e-08	.5614322	.057	0.525
Equipment Index^	Health Facility	128	-.0031828	.524324	.116	0.352
Medical Material Index^	Health Facility	129	-5.01e-09	.4828278	-.164	0.147
Stock of Vaccines Index^	Health Facility	125	-7.71e-10	.8745002	-.17	0.144
Health Workers						
Female	Health Worker	457	.4810127	.5006968	-.101	0.042**
Age (years)	Health Worker	456	40.31224	10.94959	-.101	0.932
Doctor	Health Worker	457	.0421941	.2014572	.016	0.29
Qualified Nurse	Health Worker	457	.2362869	.4256995	-.013	0.708
Non-qualified Nurse	Health Worker	457	.3122363	.4643864	-.011	0.801
Midwife	Health Worker	457	.1561181	.3637355	.006	0.849
Adjunct	Health Worker	457	.1687764	.375347	.003	0.952
No education	Health Worker	457	.0759494	.2654777	-.014	0.747
Primary Education	Health Worker	457	.0801688	.272129	-.016	0.612
Secondary Education	Health Worker	457	.3122363	.4643864	.061	0.262
Higher Education	Health Worker	457	.2278481	.4203318	.012	0.822
Experience (years)	Health Worker	455	10.64255	10.16824	-1.489	0.137
Income (Francs Congolais)	Health Worker	304	69508.57	69909.81	2084.831	0.837
Satisfied in the current position	Health Worker	457	.4767933	.5005182	.057	0.262

Appendix Table 4 (continued): Balance Checks

	Unit of Observation	Number of Observations	Mean in the control group	Standard Deviation in the control group	Difference in mean in the treatment group	p-value (difference=0)
Households						
Muslim	Household	1059	.0288462	.1675352	.001	0.931
Christian	Household	1059	.9076923	.2897385	-.039	0.065*
Animist	Household	1059	.0192308	.1374674	.005	0.668
Housing Index^	Household	1059	-.0003307	.6473751	-.076	0.349
Female	Household Member	6816	.4944524	.5000404	-.005	0.593
Maried	Adult 15+	3845	.5153217	.4998929	.032	0.154
Single person	Adult 15+	3845	.4147089	.4927976	-.017	0.445
Attended some school	Adult 15+	5431	.7348323	.4415013	-.033	0.266
Completed primary school	Adult 15+	2091	.4817245	.4999002	-.031	0.349
If was ever pregnant, visited a health facility during the last pregnancy	Pregnant Women	1017	.7447217	.4364366	-.015	0.691
If ever had a child, the last delivery was attended	Mother	862	.7112676	.4537061	-.046	0.379
If ever had a child, visited a health facility after the last delivery	Mother	846	.2805755	.4498201	.011	0.785
If ever had a child, the last child is still alive	Mother	859	.9270588	.2603464	-.005	0.777
Age in months	Children 0-10	2654	57.73658	36.12342	-.756	0.537
Immunization card seen by the interviewer	Children 0-5	1351	.122093	.3276314	.02	0.409
If immunization card seen, had BCG	Children 0-5	177	.797619	.4041878	-.034	0.653
If immunization card seen, had Polio	Children 0-5	172	.5432099	.501233	.017	0.859
If immunization card seen, had DTC	Children 0-5	177	.7261904	.4485906	.08	0.348
If immunization card seen, had Measles	Children 0-5	175	.4938272	.503077	-.015	0.878
If immunization card seen, had Vitamins	Children 0-5	178	.2738095	.4485906	-.05	0.489

Data Source: Baseline survey. ***, **, * indicate significance at 1, 5, and 10%. Error terms are clustered at the health area level. We control for the urban/rural location of the health facility.
^Each Summary Index is the equally weighted average of z-scores of its components. The z-scores are calculated by subtracting the control group mean and dividing by the control group standard deviation. The components of the Infrastructure Index are dummies of whether the health facility has a water tap, electricity, a waste disposal, a sewage disposal, and a pharmacy. The components of the Equipment Index are dummies for whether the health facility has a phone, a radio, the number of electricity generators, the number of examination tables, the number of refrigerators, the number of fuel liters, and the number of kerosene liters. The components for the Medical Material Index are dummies for whether the health facility has autoclaves, tensiometers, sthetoscopes, scales, gauges, microscope, and a delivery kit. The Stock of Vaccines Index components are the number of BCG vaccines, DTC vaccines, Polio vaccines, measles vaccines and anti-amariale vaccines in the facility. The Housing Index components are dummies for whether the household housing has a water tap, sanitation, garbage collection, and some energy (fuel or electricity).