Do Supply-Side Incentives Improve the Use of Healthcare Services? New Evidence from a Field Experiment

Abstract

Studies have recently kindled light on the effects of incentive modalities in the healthcare sector. However, there is insufficient evidence on underlying causes of partial effectiveness of these strategies in the health systems of developing countries. This study presents results from a large-scale randomized experiment across 6,848 households in Afghanistan that evaluates the impact of a conditional incentive pay scheme to the health facilities. Supported by the target income hypothesis framework and relaxing the compliance assumption in the empirical modeling, the estimated coefficients yield causal effects of supply-side conditional incentive on the demand for healthcare services. After two years, conditional incentive led to induce the demand for the pre-targeted maternal and children healthcare services among the households at lower levels and contracted-out health facilities. Further, the incentive scheme is associated with sizable efficiency gains at facility level. These gains are realized at the expense of deterring service users' satisfaction with the physician's communication qualities. This study establishes that margins of improvement do exist on the supply-side performance conditioning on organizational structure and the service contractual arrangements of health facilities. The current work provides a framework for plausible implementation of incentive policies in the healthcare sector.

JEL codes: C93, I12, J41, M52

Keywords: Financial incentive; field experiment; target income hypothesis; noncompliance; conditional mixed process; instrumental variables

1. Introduction

Applications of financial incentives in the health systems of developing countries have risen in prominence among the policy instruments intended to scale up both the demand and supply of healthcare services (Ellis & McGuire, 1993). While increase in the use and provision of the health care services are of first-order importance in implementing these incentivized schemes, there are explicit policy and practice gaps in understanding the organizational and institutional structure of incentive recipients. For example, impacts of an incentive package on delivery of services could be different given the payment mechanisms, type of healthcare facilities and involvement of multiple stakeholders in the health system. These differences can have important public policy implications for the impactful provision of incentives in the resource constrained health systems. In this paper, I study the causal effects of supply-side monetary incentive on the use of healthcare services by exploiting a carefully designed and large-scale field experiment in Afghanistan.

The first part of this paper investigates the overall effects of a supply-side conditional incentive on the demand for the pre-targeted women and children health outcomes across different levels of health facilities. Investigating facility level effects is essential to a full understanding of the impact of any incentive policy that aims to improve production of healthcare services (G. Miller & Babiarz, 2013). Health facilities as the building blocks of health systems play a significant role in improving both quantity and quality of services (Mwabu, 2007). While different dimensions of a health facility may influence the outcome of an incentive policy, the main focus of this study is on the size and service composition that have played a substantial role in the empirical and theoretical literature (Baker, Bundorf, & Kessler, 2016). After relaxing the compliance assumption in the experimental design, this empirical study estimates the casual effects of supply-side conditional incentive on the use of healthcare utilization measures. In addition to evaluating the incentive effects across various levels of the health system, the study sample is bifurcated into the users affiliated with the contracted-in (government regulated) and contracted-out (non-governmental organizations regulated) health facilities. The goal is to investigate whether these contracting arrangements translate into differences in the medical staff behaviors after receiving conditional incentives.

The second part of this study evaluates the experimental variation in the effects of conditional incentives on beneficiaries' satisfaction with the quality indicators of healthcare services provision. Specifically, both physician and health facility related quality outcomes are evaluated. Satisfaction from provider's communication qualities is evaluated at different levels of care. Provider's payment generosity could influence service users' satisfaction regarding the accessibility and quality of services (Brunt & Jensen, 2014).

The last part of this analysis explores the effects of the incentive design on the efficiency improvement of health facilities. For that, a non-parametric estimation method is used to elicit the technical efficiency of facilities pre- and post-realization of the performance incentive. Human resource constraints in running efficient health facilities has been a heated topic of interest among researchers to understand rational combination of medical resources (A. Banerjee, Deaton, & Duflo, 2004; Rosenzweig & Wolpin, 1986).

The field experiment of this work is conducted with a sample of eleven provinces (6,484 households, 143 different levels of health facilities and 286 villages) in two different time periods (baseline and endline surveys) with the financial and technical support of the World Bank (WB) and the Ministry of Public Health of Afghanistan. The experiments and randomizations were performed at village, health facility and household levels. The rich design of the current experiment allows me to scrutinize the causal effects of supply-side incentives after correcting noncompliance problem.

The first set of estimates shows that conditional incentives to the health workers (i.e., physicians, nurses, midwives and administrative workers of the health facilities) have small but statistically significant effects on the use of healthcare services by the beneficiaries of treatment health facilities at lower level health facilities. On average, at a sub center (SC) facility, the incentive pay is associated with a 2.2 and 1.6 percentage points (pp) increase in the use of women's pre- and post-natal health services, respectively. By the same token, the incentive pay increases the likelihood of skilled births and institutional deliveries at SC facilities. Notwithstanding, the magnitude of these estimates proportionally decreases as the size of health facilities increases. To the extent that the estimated effects of incentive reverses sign at the district hospitals (DH). The treatment effects on the children health outcomes supersede the pattern of maternal effects across various levels of the health facilities. So that, at SC and basic health center (BHC) levels, the probability of visiting health facilities to seek care, obtaining required types of childhood vaccinations rises considerably. These point estimates decline at upper level health centers. Taken together, the findings reveal that the use of both women and children services widely differ when the effect of incentives is evaluated at each health facility.

The second set of results presents how incentivizing healthcare providers changes quality indices from beneficiaries' perspective. Specifically, the effects of conditional incentive on the patient centeredness and facility accessibility is documented as the key quality signals of the health services provision. Beneficiaries assigned with the treatment facilities report mostly lower satisfaction in the patient-related quality indices when compared with the comparison arm at the endline survey. The normalized scores of physician's *respectfulness* with patients are 2.4 percentage points higher at the incentivized facilities SC while at the baseline survey these scores were not statistically different between comparison and treatment facilities. In contrast, the empirical models could barely capture or imprecisely estimate the effects of incentive on health-facility related indices such as *cleanliness* and *availability of medicine* at treatment facilities.

The final set of findings documents a small yet significant effect of supply-side incentives on the efficiency gradients of treatment health facilities. Using the piecewise linearity concept of data envelopment analysis (DEA), 6-7% increase in the input-oriented efficiency scores of treatment facilities is detected when compared with the comparison facilities at the endline survey.

This analysis on the effects of supply-side conditional incentive offers several contributions to a growing body of literature. First, while the existing literature evaluates the overall effects of incentive schemes on the provider's performance, this work helps to extend incorporating the supply-side incentive effects on use of healthcare services at various levels of care. This extension is crucial for understanding the feasibility of incentive pay in the most developing country contexts where provision of services is partitioned into different levels. Second, this paper investigates the effectiveness of incentive pay schemes with regards to the healthcare services outsourcing arrangements. Inadequate organizational and institutional capacities in low-income settings have brought the role of contracting-out mechanisms to the center of attention. The estimates of current work provide new evidence on the contribution of outsourcing mechanisms in the feasibility and success of incentive schemes when designing and implementing. Third, a simple theoretical model based upon target income hypothesis within the principal-agent problem setting is crafted to explicitly determine the underlying mechanism of supply-side financial incentives in the healthcare system. Lastly but most importantly, this work contributes to the growing body of research on empirical modeling of experimental datasets. In contrast to other related works, the underlying assumption of compliance in the experiment design is relaxed in this exercise. Using instrumental variables (IV) modeling strategy, the bias-corrected effects of conditional incentive on the use of health services are obtained.

The remainder of this paper is organized as it follows. The next two sections discuss country context and the related literature on the effects of incentive payments in healthcare sector. Followed by an introduction to the theoretical construct and research design of the study. Then, the empirical results are presented. Last two sections discuss the results and conclude the paper by assessing the findings and policy implications.

2. Institutional setting and country context

With the highest fertility rate and the lowest maternal and child health indicators in Asia, Afghanistan is globally recognized for achieving significant improvements in the certain health deliverables since 2001. Maternal and infant mortalities have declined from 1,600 to 396 per 100,000 and from 66 to 45 per 1,000 live births, respectively. Access to health care services within two hours of walking distance has reached to 87% while it was less than 10% in 2001 (DHS, 2017). Besides a relative macroeconomic stability, these gains are the consequence of smooth implementation of the basic package of health services (BPHS) and the essential package of hospital services (EPHS) as the country's foundation for the health system. BPHS and EPHS with nearly 2,300 health facilities (HFs) across 34 provinces of Afghanistan are the initial point of contact for receiving preventive and curative health care services (NHS, 2016). These HFs are partitioned into provincial and district hospitals (DH), comprehensive health centers (CHC), basic health centers (BHC), sub centers and mobile health teams (SC) to ensure a hierarchical and well-defined referral system. Each of the above HF has a defined population catchment area serving 2,500-150,000 individuals. In an effort to build up the organizational and institutional capacities in the public sector and the existence of macro level funding agencies (e.g., the World Bank, United States Agency for International Development and European Commission), the Afghan government has implemented the provision of health care services through non-governmental organizations (NGOs) based on a contractual arrangement; namely, contracted-out approach. Approximately, 77% of the population deliver healthcare services through the latter one while less than 25% of people are served by the government regulated mechanism.

Although initial efforts toward system building and improving health outcomes have been taken, the country is still faced with several challenges concerning healthcare coverage, quality and financing. These obstacles could negatively affect every citizen but mostly women and children as the desperate groups of population in the country. Evidently, utilization of maternal and child healthcare services has lagged behind. Characterized by Afghan's health and demographic survey DHS (2017) estimates, only 59% and 51% of pregnant women have received one skilled pre-natal care and have had delivery in presence of a skilled birth attendant, respectively. Similarly, use of modern contraceptive and post-natal care are stagnant since 2010. Not to mention, with only 5% government contribution to the total health expenditure (THE), Afghan's out-of-pocket health expenses substitute approximately three-quarters of THE. Lack of quality in the service provision is another equally important challenge that contributes to the lower use of services and significant escape of monetary resources as a result of medical tourism to the regional countries (NHA, 2014).

Since early 2010, rising political instability and security concerns have constantly threatened efficient delivery of healthcare services throughout the country. In addition, reduction in the foreign aid to the health system, and economic downturn in the country represent other obstacles to provision of responsive health services and their sustainability (NHS, 2016). These unfavorable circumstances have raised a laudable public policy debate to maintain current gains and to promote health service delivery without quality compromise. In response, the government of Afghanistan with the financial and technical supports of the World Bank (WB) has piloted supply-side conditional incentive scheme aiming to further improve accessible, equitable and quality health services with a special attention to women and children health outcomes. More Specifically, incentive pay which interfaces healthcare utilization outcomes with the supply-side (health care providers) incentives was initiated in 2010. This paper is intended to document the evidence on the effects of current performance-based incentive initiative both on the use of women and children health indicators and patient satisfaction from physician communication qualities. The rich design of the field experiment of the study allow me to obtain the bias-corrected estimated effects of incentive program. The empirical findings of this analysis have certain public policy implications that could enhance the technical and allocative efficiency of resource allocation in donor-dependent and resource- limited settings of developing countries.

3. Related literature

Pay for performance (P4P) is defined as conditional transfer of monetary or material incentives when a pre-target performance measure is achieved (Eichler, 2006). In the health sector, P4P has gained significant popularity in improving utilization of services and overall system performance (Van de Poel, Flores, Ir, & O'donnell, 2016). In the low-income countries with weak institutional and organizational capacities and heavy dependence on foreign developmental aid, relying on the effectiveness of available resources to improve allocative and technical efficiencies of health systems has been a key public policy agenda (Attanasio, Oppedisano, & Vera-Hernández, 2015; Handa, Peterman, Seidenfeld, & Tembo, 2016; Powell-Jackson, Mazumdar, & Mills, 2015; Renmans, Holvoet, Orach, & Criel, 2016; Trani, Kumar, Ballard, & Chandola, 2017). P4P is associated with both the demand- and supply-sides. On the demand-side, conditional cash transfer is a well-established mechanism to incentivize users of services (e.g., households, women and children) upon utilizing available services as necessary (De Janvry & Sadoulet, 2006; Gertler, 2004). The performance-based incentive to the health care providers (e.g., medical doctors, nurses, midwives and other health personals) is recognized on the supply side of the market (de Hennin & Rozema, 2011; Ireland, Paul, & Dujardin, 2011). This paper is related to a stream of literature focusing on the supply-side of P4P initiatives.

Essentially, the production of healthcare services is determined by availability of the structural inputs (health care providers, medical devices, supplies and drugs) and the effective processes that change the inputs into outputs initially and into outcomes ultimately (Polachek, Das, & Thamma-Apiroam, 2015). Even though structural inputs are sufficiently provided, in most low-income settings, the quality of the processes is unsatisfactory as a result of poor performance of healthcare providers. Monetary and non-monetary rewards aiming to bring behavioral changes among the health care providers could potentially improve the performance (Eichler, 2006). Performance-based incentives target providers at individual or facility levels. The former has been implemented predominately in developed economies (Vujicic, 2009). facility-based incentive which improves efficient resource allocation especially when service production entails a strong interdependence among group members (Gaynor, Rebitzer, & Taylor, 2004) is a dominant strategy in the low-income countries. Unlike other empirical health economics studies, the literature related to P4P is affluent with the experimental research framework given the complexity of the schemes.

Exploiting a well-designed controlled randomized experiment and the difference-in-differences empirical model, Basinga et al. (2011) studied the effect of P4P on the use maternal and child health indicators in Rwanda. They found a positive effect of the program on the number of institutional deliveries and children's preventive care but no statistically significant effect on the number of prenatal visits and child's immunization services. One of the potential limitations of their study could be the simultaneous effect of decentralization policies during the study periods. Relying on a quasi-experimental design, De Walque et al. (2015) showed that P4P scheme in Rwanda has increased HIV testing among married and discordant couples by 10.2 and 14.17 pp, respectively. Using two waves of demographic and health surveys for the same setting, Sherry, Bauhoff, and Mohanan (2017a) analyzed the effects of PBF mechanism on the healthcare utilizations, outcomes and the program's unintended consequences. Although they shed light on the positive effects of the program on utilization of services, the impact of the scheme on health outcomes is indistinguishable from zero.

Observing no statistically significant effect of the incentive program on pre-natal and infant's vaccination rates, Van de Poel et al. (2016) researched the impact of PBF in Cambodia. They cautiously concluded about the possible long-term effects of PBF programs in improving the quantity and quality of healthcare services in the low-income settings especially when heavily relied on the users' fee for receiving healthcare services. Their findings may suffer from the selection bias due to lack of an experimental research design. Also, the parallel implementation of other reforms in the healthcare system of Cambodia might have driven the results of PBF program.

Assessing the impact of incentives on the clinical quality of healthcare services in the Philippines, Peabody et al. (2011) investigated the effects of P4P on individual and system levels using a randomized experiment. Bonus payments to the physicians and increased compensation to the hospitals improved the clinical performance by 9.8 and 9.1 pp, respectively. However, physician's bonus, unlike the system-level bonus, had no statistically significant effect on the utilization of services. Boosting the extrinsic but deterring the intrinsic motivation with the minimum or no coverage improvement after implementing a PBF scheme in the Democratic Republic of Congo, Huillery and Seban (2013) suggested a permanent supply-side policy shift combined with other innovative demand-side schemes after analyzing experimental evidence. The current paper deviates from the past literature in at least three different ways. First, evaluating the casual effects of supply-side conditional incentive at different levels of care is a policy relevant and unique initiative. Second, in contrast to the past work, the empirical modeling of this paper relaxes the perfect compliance assumption in the experimental setting. Estimated coefficients of the treatment effects for both the continuous and categorical outcomes are bias-corrected. Third, besides estimating the treatment effects on efficiency of the contracted-in and contracted-out health facilities, I provide the causal effects of incentive on the use of healthcare services among the recipients of both types of contractual arrangement.

4. Response to monetary incentive: Theory

Healthcare provider's behavioral response to incentive pays could be rationalized within the target income and work-leisure trade-off hypotheses (McGuire & Pauly, 1991; Thornton & Eakin, 1997). The earlier suggests that providers customize both the composition and volume of the offered services to reach to a targeted level of income while the later proposes that physician as the utility maximizing agent trade-offs the volume of services for the leisure time. In either framework, the increased supply of service is associated with the relative strength of income or the price effect. To cement this knowledge, I modify and craft a simple theoretical model akin to Garen (1994) of health worker's response to the incentive payments. Assuming that implementer (e.g., NGOs or government) and provider at a health facility are the principal and agent, respectively; agent determines the level of effort needed to produce Q quantity of health care services to the service beneficiaries. Quantity Q is a vector comprised of several types of services $Q=(q_1,q_2,q_3,...,q_n)$. Increasing the quantity of healthcare services Q by the agent is the potential benefit to the principal and is revealed in the forms of increased health care services $Y=(y_1,y_2,y_3,...,y_n)$ utilized by beneficiaries (e.g., women, children). Implicit and explicit incurred cost to the agent to provide Y services is shown by $C(Y)$ which is a twice differentiable and positive function. Provision of Y services depend upon the health facility fixed resources $\theta(R)$, the health worker' effort $\pi(E)$ and other health facility or beneficiary's unobservable factors ζ as written below:

$$Y=\pi(\mathrm{E})+\theta(R)+\zeta. \tag{1}$$

It is assumed that the agent's actual level of effort is unobservable to the principal. In the absence of any incentive, the agent's effort is compensated by his monthly income P and the marginal increase in the expected level of services resulting from an increase in the worker's effort is:

$$\frac{\partial Y}{\partial P}=\mathrm{P}<\mathrm{C(Y)}, \tag{2}$$

however, the principal could incentivize the health worker to maximize his level of effort $P \geq C(Y)$ conditional upon achieving pre-determined targets S_t. Then, the payment structure $\mathbb{R}$ for the agent in presence of an incentive pay scheme is:

$$\mathbb{R} = P_0 + \sum_{t=1}^{T} P_1 Z(Y \geq S_t), \tag{3}$$

substituting Eq. (3) into Eq. (1) yields the incentivized provision of services and it is written as:

$$Y^* = \mathbb{R} + \theta(R) + e. \tag{4}$$

The risk neutral agent solves the following profit maximization problem while taking into consideration of an incentive package:

$$E(R) - C(Y) = P_0 + \sum_{t=1}^{T} P_1 \Pr(Y \geq S_t) - C(Y) \tag{5}$$

further, Eq. (5) can be rewritten using cumulative density function of ε (F) as:

$$E(R) - C(Y) = P_0 + \sum_{t=1}^{T} P_1 \left\{ F_t(y^*) - S_t \right\} - C(Y). \tag{6}$$

The first-order condition $\partial C / \partial Y_i \ (i..., I)$ affirms that the agent chooses the level of services (Y) by equalizing the marginal cost of service provision to expected level of profit received. In the empirical section, Eq. (4) is estimated to obtain bias-corrected coefficients for the treatment effect.

5. Research design

5.1 Experimental setup

Incentive payment linked to metric-driven outcome improvement is initiated in the health system of Afghanistan with the precise objectives[1]. The pay scheme is administered and implemented nationally[2] both at contracted-in and contracted-out health facilities. The payments associated with the incentive are reimbursed four times annually to the health care workers at each facility level conditional upon their performance and increase in the quantity of pre-targeted healthcare services. Initially, each health facility reports the monthly supplied services to its provincial managing office. Then, these reports are subject to quarterly verification process by three different methods. Random interviews of individuals who are reported as beneficiaries (users of services) of the health facility, random visits to the health facility by a group of independent monitors and finally matching the facility monthly reports with the registries of health information system (HIS). NGOs have negotiation power to discuss and adjust the incentive payments based on improvements in the service provision at each facility given the availability of exogenous constraints such as insecurity or geographical location of the facility. Once verified by the ministry of public health (MoPH), the incentive payments are transferred to the managing offices of contracted-in and contracted health facilities along their fixed budget. Distribution of the incentive pays at each health facility solely depends upon the health care providers in that facility. Providers salary scale, level of individual's effort and/or equal share of payment to every member of the facility are some of the reported incentive distributional mechanisms. Nonetheless, the bonus if received is between 11-28% of the worker's base monthly salary. To evaluate the impact of this incentive scheme, a large-scale field experiments have been performed by the financial support of

[1] (1) To increase key maternal and child health indicators (2) to improve the quality of services (3) to maximize patients and communities' involvement and satisfaction

[2] In 11 out of 34 provinces of the country

the World Bank (WB) and the technical support of MoPH. The experimental design of this study is based on randomization at health facility, village and household levels. In the first stage, all 374 HFs in nine provinces of the country are stratified by the facility type (DH, CHC, BHC and SC) and the proportion of each facility type is calculated in a particular province[3]. For each province, the HFs are matched in pairs based on similarities (average number of outpatient visits in a month and number of staffs in the facility) and then eight pairs are randomly selected. Treatment and comparison facilities are chosen within each pair – one facility is comparison and the other is treatment to ensure that comparison facility within a pair is very similar to treatment. In total, 140 HFs in nine provinces which is approximately 16 facilities for each province is randomized. In the Second stage, the catchment area[4] of each HF is identified to properly determine the total number of villages. Thereafter, two villages[5] (comparison and treatment village) are randomly selected for each health facility for all of 140 HFs. In total, 280 villages are randomly selected. In the third stage, all the households in the randomized villages (280 villages) for each province is listed. The required[6] number of households is randomly selected from those comparison and treatment villages (see Table 1). In summary, the sample size for this study consists of 140 HFs, 280 villages and 6,848 households.

5.2 Empirical model

Considering the focus of policy context in the healthcare sector, the main outcome of interest is the utilization indicators for different types of services. The empirical model of this paper exploits the random variation in the use of these services induced by a performance-based

[3] The incentive scheme is implemented in 11 provinces. However, two provinces were dropped due to insecurity.
[4] Catchment area of each health facility indicates the number of people being served by that health facility.
[5] Villages with less than 50 households have to be linked with the nearest village in that facility's catchment area and treated as one unit of randomization.
[6] Security, logistical and statistical facts are considered in choosing the final sample.

pay to estimate the effects of supply-side conditional incentive. A natural place to initiate the empirical analysis of an experimental data is by comparing the outcomes in the premises of treatment status. Assuming perfect compliance in the experiment setting, the random assignment of observations allows us to obtain the average casual effect of treatment on the treated (ATET) as shown:

$$E\left[y_i \mid d_i = 1\right] - E\left[y_i \mid d_i = 0\right] = E\left[y_{1i} - y_{0i} \mid d_i = 1\right] = E\left[y_{1i} - y_{0i}\right] \tag{7}$$

where y_i is the outcome variable for observation i, d_i is a binary variable shows whether observation is assigned to treatment or not. howbeit, the above formulation is unable to recover the ATET when the compliance assumption is violated (Balke & Pearl, 1997). The selection bias as a consequence of the generated relationship between the treatment status and outcome of interest confounds the estimated effects of treatment. The instrumental variables (IV) framework, on the other hand, has the potential to transform the comparisons using the intended random assignment into the bias-corrected causal estimates of the treatment (Heckman, 1995; Moffitt, 1996). To understand the intuition behind the IV model application in solving the compliance problem, I start with constructing a model with the constant treatment effects $\phi = [y_{1i} - y_{0i}]$, where $y_{0i} = \alpha + \varphi_i$, and $\alpha = E(y_{0i})$, akin to J. D. Angrist and Pischke (2008). Then, the outcome model is written:

$$y_i = \alpha + \phi d_i + \varphi_i. \tag{8}$$

The treatment effects ϕ could not be consistently estimated if there is correlation between the idiosyncratic error φ_i and the treatment status d_i.Yet, the random assignment of intended treatment status Z_i could provide us with a solution to disentangle the bias-corrected causal estimates of the incentive. Realizing the facts that assigned treatments have no direct effect on the outcome of interest other than through obtaining treatments, the instrumental variable Z_i and the outcome of interest y_i are independent and therefore the zero-expectation assumption $E[\varphi_i | Z_i] = 0$ does hold. Using the instrument Z_i, we can write the conditional expectation of y_i as

$$\left\{E[y_i | Z_i = 1] - E[y_i | Z_i = 0\right\} \big/ \left\{E[d_i | Z_i = 1] - E[d_i | Z_i = 0\right\} = \phi. \tag{9}$$

The treatment effects obtained from the Eq. (9) is sufficient with the constant-treatment effects assumption though in the real-world practice it is unrealistic. For instance, health seeking behavior of some women and children could be improved as a direct result of the supply-side incentive (e.g., supply-induced demand) while others may get affected little or not at all. Also, there is possible heterogeneity in the treatment effects among beneficiaries. To overcome these hurdles, J. D. Angrist, Imbens, and Rubin (1996) propose that in presence of heterogeneity in the treatment effects, the IV estimates capture the average causal effects for a subset of treated observations and this effect is called local average treatment effects (LATE). To better understand the intuition behind the LATE, let the potential treatment assignments be d_{0i} and d_{1i} for observation i while Z_i takes a binary value of 0 or 1. Then, the observed treated status could be written as

$$d_i = d_{0i} + Z_i(d_{1i} - d_{0i}). \tag{10}$$

The above setup necessitates a number of key underlying assumptions to provide casual inference. First, the random assignment of observations into comparison and treatment groups satisfies the *exclusion restriction* assumption of the instrument (Altonji, Elder, & Taber, 2005). Exclusion restriction implies that the instrument does not have a direct effect on the potential outcome of interest. Second, the incentive offer could potentially affect the use of healthcare services by the treated recipients and this confirms the *relevance assumption* (Shea, 1997). Third, the instrumental variable is randomly assigned to the units of randomization and it explicitly satisfies the *exogeneity condition* of IV in the model. Finally, all the observations affected by the instrument, must be affected in the same way (Hoderlein & Mammen, 2007). To formalize the idea, the following models are empirically estimated:

$$d_i = \beta_0 + \beta_1 Z_i + \beta_2 X_{it} + \vartheta_i \tag{11}$$

$$Y_i = \gamma_0 + \gamma_1 \hat{d}_i + \gamma_2 X_i + \upsilon_i \ . \tag{12}$$

Equations (11) and (12) shows the first and second stages of the 2SLS model, respectively. Within this empirical strategy, the bias-corrected causal effects of incentive payments on the outcome variables (e.g., utilization of healthcare services, patients' satisfaction) could be obtained even if noncompliance existed (Heckman, 1995; Moffitt, 1996). In the empirical result section, the results of equations (12) and (7) are provided side by side to compare the estimated coefficients of the models. For the ordered outcome variables, the conditional mixed mixed-process (CMP) modeling framework (Roodman, 2011) is executed. The CMP allows the appropriate estimation of two or more equations with any possible linkage among their error processes and their discrete outcome variables.

6. Results

6.1 Validity of randomization

The main objective of a randomization process is to ensure that the assignment of observations into treatment is orthogonal to other observable characteristics of the study sample that could possibly be correlated with the health services outcomes (Burde & Linden, 2013).

Table 1 Balance and difference of treatment and comparison groups at baseline and endline surveys

	Baseline survey			Endline survey		
	Comparison (1)	Treatment (2)	Difference (3)	Comparison (4)	Treatment (5)	Difference (6)
Health Facilities						
Identified	102	113	-11	102	113	-11
Randomized	72	72	0	71	72	-1
Villages (Clusters)						
Identified	204	226	-22	204	226	-22
Randomized	144	144	0	142	144	-2
Households						
Identified	4,080	4,520	-440	4,080	4,520	-440
Surveyed	3,443	3,341	102	3,427	3,421	6
Women aged 15-49 years†						
Identified	3,866	3,931	-65	4,079	4,166	-87
Surveyed	3,865	3,929	-64	4,042	4,132	-90
Infants under 5 years						
Identified	4,397	4,590	-193	3,895	3,898	-3
Surveyed (Mothers responded)	4,355	4,525	-170	3,911	3,895	16
Infant's gender						
Male	2,126	2,229	103	2,001	1,910	-81
Female	2,229	2,296	70	1,966	1,929	-37
HF affiliation						
Contracting-out (NGOs)	2,958	2,956	-2	3,233	3,089	-144
Strengthening Mechanism (Government)	608	975	367	899	953	54

† Although the number of treatment women in the endline survey remains the same, the number of compliers (Who

actually treated) have been declined.

Table 1 provides a measurement of differences between composition of comparison and treatment groups at baseline and endline surveys. In both surveys, there is a strong balance between the number of health facilities, villages, households, women and infants. Generally, all the differences are small and consistent over two study periods and therefore the randomization appears to be successful. Table 2 reports the result of equality test for the observable characteristics of the women included in the study sample. In the baseline survey, none of the differences between comparison and treatment groups are statistically distinguishable from zero with an exception of family size. At the endline survey, however, all the differences in the socio-economic characteristics of respondents are statistically not significant. Apparently, lack of differential attrition in the control and treatment groups could suggest that the estimates associated with the incentive do not suffers from biasness (J. Angrist, 1995). However, noncompliance of respondents in the endline survey raise an important concern when estimating the effects of a treatment program.

Table 2 Observable characteristics of Women and Infants at treatment and control groups before and after incentive program

	Before incentive			After incentive		
	Control	Treatment	Difference	Control	Treatment	Difference
	(1)	(2)	(3)	(4)	(5)	(6)
average age	32.50	32.51	0.013	31.50	31.76	0.259
			(0.23)			(0.254)
Education						
No school	3,694	3,710	16	3,781	3,717	-64
Primary	110	136	26	273	238	-35
Higher	62	85	23	78	86	-8

Pregnancy						
Yes	494	573	79	624	646	22
No	3,315	3,285	-30	3,451	3,342	-109
Not sure	41	47	6	50	44	-6
Household size	9.62	10.16	0.542*	8.14	8.31	0.174
(Mean)			(0.299)			(0.195)
Household wealth						
Poorest	648	700	52	665	637	-28
Second	788	671	-117	1,019	798	-221
Third	865	744	-121	1,025	908	-117
Fourth	785	789	4	799	826	27
Richest	779	1027	248	624	873	249

Standard errors are reported in parentheses. The sample is consisted of 8, 174 women at child-bearing age.
Significance levels are indicated as ***1%, **5%, *10%.

6.2 Incentive and maternal health outcomes

Table 3 shows the casual effects of incentive on the use of healthcare services by women across different health facilities. Odd and even columns present the estimated coefficients of OLS[7] the IV models, respectively, outlined in section 5.2. The main outcomes of interest are prenatal and postnatal care, skilled birth and birth at a health facility, and contraceptive use. Column (2) corresponds to the IV estimated effects of incentive at SC facility. Except contraceptive use, all other outcomes reveal a statistically and economically significant and positive estimate associated with the incentive. Implying an increase of 2.2 and 1.6 in the number of pre-and post-natal care, respectively, after providing incentives to health facilities. In addition, skilled births and births at the health facility induced by 9.1 and 3.2 percentage points (pp), respectively. Again, the incentive

[7] OLS estimates are provided for comparison purpose.

effects shown in column (4) remain statistically significant at the BHC facilities, but the magnitude is smaller than SC estimates. Column (6) indicates that the positive effects of incentive on the skilled birth and births at a health facility fade out at CHC level though these effects decline by more than a half for pre- and post-natal care outcomes. Notably, the estimated effects of incentive reverse signs at a DH. Column (8) estimates indicate that the number of pre- and post-natal care use shrinks by 2.7 and 4.6, respectively. Likewise, skilled birth and birth in a health facility decrease by 7.4 and 17.9 pp, respectively. In contrast to other health facilities, the effect of incentive on the use of contraceptive is precisely estimated at DH level and it lowers the use of contraceptive by 19.7 pp. Taken together, my results suggest that the effect of supply-side incentive is positive and larger at smaller health facilities while negative and relatively smaller at large health facilities. Finally, the IV estimator is likely to provide a consistent estimate as the unobserved utilization determinants are likely not be uniformly distributed across women living in the catchment areas of each health facility.

Table 3 Effects of incentive scheme on the use of maternal health indicators across different health facilities

	Sub Center (SC)		Basic Health Center (BHC)		Comprehensive Health Center (CHC)		District Hospitals (DH)	
Dependent Variables	OLS (1)	IV (2)	OLS (3)	IV (4)	OLS (5)	IV (6)	OLS (7)	IV (8)
Prenatal care	1.192	2.161**	0.861	1.094***	0.914***	0.707**	-2.406	-2.702*
	(1.011)	(0.857)	(0.722)	(0.165)	(0.269)	(0.337)	(2.931)	(1.485)
Postnatal care	-6.007**	1.587**	-2.922**	1.352***	-2.024	0.709*	-4.874	-4.635**
	(2.761)	(0.636)	(1.425)	(0.125)	(2.008)	(0.346)	(4.398)	(2.069)
Skilled birth attendance	0.059*	0.091*	0.045	0.042*	-0.029	-0.061	0.085	-0.074*
	(0.032)	(0.058)	(0.031)	(0.023)	(0.048)	(0.045)	(0.081)	0.043
Birth in health facility	0.053	0.032**	0.129	0.024***	-0.052	-0.087	-0.118	-0.179**
	(0.038)	(0.016)	(0.196)	(0.036)	(0.082)	(0.230)	(0.192)	(0.082)
Contraceptive use	-0.147***	-0.107	-0.026	0.712	-0.034	-0.009	-0.198*	-0.197*
	(0.054)	(0.071)	(0.037)	(0.629)	(0.058)	(0.072)	(0.117)	(0.102)

All regressions include the following observables: Women's age an educational background, household size, wealth quantile. The dependent variables are responses to endline survey questions elicited post incentive scheme. Standard errors are reported in parentheses clustered at village level. OLS and IV stand for ordinary least square and instrumental variables models, respectively. The sample is consisted of 8, 174 women at child-bearing age.
Significance levels are indicated as ***1%, **5%, *10%.

6.3 Incentive and children health indicators

Having established that providing supply-side incentive significantly affects women health outcomes across different health facilities, I next turn my attention to the effects of incentive scheme on child health indicators. Table 4 presents the results of OLS and IV regressions on pre-targeted child health outcomes (visit health facility, BCG, OPV, Penta and measles vaccines, and use of vitamin A). Starting with subcenters, incentivizing healthcare providers significantly increases the use of health facility and receiving certain types of vaccines by children. OPV and Penta vaccination rates are improved by 26.2 and 17.9 pp, respectively. The likelihood of visiting a subcenter for receiving care is increased by 14.7% among children who live in the catchment area of a treated facility. These positive effects remain statistically significant at basic health centers but with a relatively smaller magnitude than subcenters. Although the incentive scheme induces the rates for measles vaccine by 5.5 pp at the comprehensive health facility, it negatively affects both the likelihood of visiting health center and rate of Penta vaccine. Akin to the effects of incentive on maternal health outcomes, children health indicators occur to be negatively associated with the incentive scheme at district hospitals, emphasizing that the incentive does appear to hurt the use of healthcare services as the size of health facility expands.

Table 4 Effects of incentive scheme on the use of children health indicators across different health facilities

	Sub Center (SC)		Basic Health Center (BHC)		Comprehensive Health Center (CHC)		District Hospitals (DH)	
Dependent Variables	OLS (1)	IV (2)	OLS (3)	IV (4)	OLS (5)	IV (6)	OLS (7)	IV (8)
Visit health facility	0.111***	0.147**	0.116**	0.127***	-0.152***	-0.146***	-0.1633*	-0.155*
	(0.046)	(0.052)	(0.041)	(0.054)	(0.032)	(0.035)	(0.088)	(0.091)
BCG vaccine	0.039**	0.025	0.021*	0.027 **	0.045**	0.043	0.065	-0.091**
	(0.017)	(0.019)	(0.01)	(0.013)	(0.019)	(0.121)	(0.049)	(0.046)
OPV vaccine	0.119	0.262**	-0.331**	0.186***	-0.238	-0.262	-0.423	-0.283
	(0.107)	(0.112)	(0.155)	(0.079)	(0.213)	(0.236)	(0.553)	(0.583)
Penta vaccine	0.810**	0.179***	0.126**	0.119**	-0.288**	-0.296**	-0.051	-0.005
	(0.388)	(0.049)	(0.054)	(0.059)	(0.129)	(0.145)	(0.141)	(0.149)
Measles vaccine	-0.027	-0.039	0.019	0.029*	0.045*	0.055**	0.055	-0.111*
	(0.024)	(0.027)	(0.016)	(0.018)	(0.025)	(0.028)	(0.069)	(0.071)

Use of Vitamin A	0.047** (0.019)	0.066*** 0.022	-0.011 (0.014)	-0.002 0.016	-0.005 (0.011)	0.022 0.022	0.148** (0.061)	0.154** 0.073

All regressions include the following observables: Mother's age an educational background, household size, wealth quantile and infant's gender. The dependent variables are responses to endline survey questions elicited post incentive scheme. Standard errors are reported in parentheses clustered at village level. OLS and IV stand for ordinary least square and instrumental variables models, respectively. The sample is consisted of 7, 806 infants who are under five years old.
Significance levels are indicated as ***1%, **5%, *10%.

6.4 Incentive and service users' satisfaction

Let me now turn to the results related to the effects of supply-side incentive on satisfaction of service users across each health facility. Table 5 elaborates on the estimates corresponding to the patient and health facility categorical outcomes that encompass both physician's behavioral and health facility related indicators. For each outcome across different health facilities, the estimated coefficients of ordered probit and the instrumental variables ordered probit are reported. The results in column (2) suggest that offering incentives to service suppliers at a subcenter facility negatively affects *physician's respectfulness* toward service users. Incentive increases service users experience with *time spent with patient* outcome by 4.2 pp. However, the casual effects of incentive on the remaining outcomes are indistinguishable from zero at a subcenter level. Except the negative effects on the beneficiary's satisfaction with *patient privacy* outcome, provider's incentive appears to have no meaningful impact on users' satisfaction outcomes at basic health facility. Strikingly enough, incentive has negatively affected most of the user's satisfaction outcomes at the comprehensive health facility. Compared with the control group, on average, doctors spend 11% less time with the service users at treatment health facility. Also, patients are less satisfied with a physician's respectfulness at the time of service delivery. Furthermore, both *cleanliness* and *availability of medicine* are declined by approximately 3.5 and 4 pp, respectively. Finally, akin to comprehensive health facilities, estimates in column (8) illustrate that conditional incentive has lowered user's satisfaction with physician's *respectfulness* and *explanation* outcomes at DH. Likewise, facility related outcomes are worsening among treated facilities.

Table 5 Marginal effects of incentive scheme on satisfaction of service users across different health facilities

	Sub Center (SC)		Basic Health Center (BHC)		Comprehensive Health Center (CHC)		District Hospitals (DH)	
Dependent Variables	OP (1)	IV-OP (2)	OP (3)	IV-OP (4)	OP (5)	IV-OP (6)	OP (7)	IV-OP (8)
Respectfulness								
Very dissatisfied	0.006	0.024*	0.003	0.0001	-0.027*	0.017***	-0.014	-0.009
	(0.012)	(0.011)	(0.005)	(0.007)	(0.014)	(0.001)	(0.031)	(0.036)
Dissatisfied	0.008	0.030*	0.009	0.003	-0.027*	0.080***	-0.033	-0.021
	(0.015)	(0.018)	(0.017)	(0.021)	(0.014)	(0.025)	90.071)	(0.086)
Satisfied	-0.002	-0.005	-0.004	-0.001	0.016*	0.019	0.008	-0.115***
	(0.003)	(0.005)	(0.007)	(0.009)	(0.009)	(0.018)	(0.019)	(0.020)
Very satisfied	-0.012	-0.049*	-0.008	-0.003	0.037*	0.029	0.039	0.025
	(0.024)	(0.021)	(0.015)	(0.019)	(0.011)	(0.118)	(0.084)	(0.102)
Explanation								
Very dissatisfied	-0.007	0.0003	0.007	0.0003	-0.013	-0.085***	-0.028	-0.078
	(0.013)	(0.016)	(0.007)	(0.009)	(0.013)	(0.016)	(0.028)	(0.056)
Dissatisfied	-0.001	0.0003	0.013	0.0003	-0.017	-0.078***	-0.107	0.157**
	(0.017)	(0.021)	(0.015)	(0.0004)	(0.016)	(0.020)	(0.081)	(0.073)
Satisfied	0.006	-0.0002	-0.007	-0.017	0.009	0.009	0.063	-0.192***
	(0.001)	(0.012)	(0.008)	(0.014)	(0.008)	(0.010)	(0.052)	(0.063)
Very satisfied	0.012	-0.0004	-0.011	-0.009	0.022	0.024	0.071	0.112
	(0.011)	(0.025)	(0.014)	(0.014)	(0.021)	(0.026)	(0.058)	(0.145)
Availability of Medicine								
Very dissatisfied	-0.001	0.019	0.034**	0.022	0.031**	0.048*	-0.121	-0.224
	(0.025)	(0.030)	(0.016)	(0.020)	(0.020)	(0.025)	(0.122)	(0.182)
Dissatisfied	-0.004	0.008	0.019**	0.012	0.026**	0.031**	-0.099	0.126
	(0.001)	(0.012)	(0.009)	(0.011)	(0.013)	(0.016)	(0.098)	(0.094)
Satisfied	0.006	-0.012	-0.026**	-0.016	-0.021**	-0.035**	0.061	-0.085*
	(0.016)	(0.019)	(0.012)	(0.015)	(0.014)	(0.018)	(0.068)	(0.046)
Very satisfied	0.008	-0.015	-0.027**	-0.020	-0.036**	-0.044*	0.159	0.184
	(0.019)	(0.023)	(0.013)	(0.016)	(0.018)	(0.023)	(0.159)	(0.101)
Patient Privacy								
Very dissatisfied	-0.007	0.002	-0.005*	-0.005**	0.005	0.007	0.044	0.012
	(0.007)	(0.008)	(0.003)	(0.002)	(0.007)	(0.008)	(0.041)	(0.048)
Dissatisfied	-0.012	0.003	-.0025*	-0.027**	0.010	0.017	0.041	0.010
	(0.011)	(0.014)	(0.014)	(0.011)	(0.015)	(0.019)	(0.036)	(0.043)
Satisfied	-0.015	0.003	-0.009*	-0.001	0.005	0.006	0.038	0.008
	(0.014)	(0.017)	(0.005)	(0.007)	(0.006)	(0.007)	(0.036)	(0.032)
Very satisfied	0.035	-0.007	0.038*	0.042	-0.011	-0.030	-0.123	-0.030
	(0.032)	(0.038)	(0.021)	(0.027)	(0.027)	(0.034)	(0.102)	(0.122)
Time spent with Patient								
Very dissatisfied	-0.029**	-0.023	-0.002	-0.005	-0.0167	-0.017	-0.025	-0.088
	(0.012)	(0.015)	(0.006)	(0.008)	(0.013)	(-0.017)	(0.039)	(0.062)
Dissatisfied	-0.031***	-0.035	-0.006	-0.011	-0.0192	-0.011	-0.037	0.031
	(0.017)	(0.020)	(0.016)	(0.019)	(0.015)	(0.019)	(0.055)	(0.020)
Satisfied	0.016**	0.012	0.003	0.006	0.017	-0.117***	0.009	-0.098**
	(0.008)	(0.009)	(0.007)	(0.009)	(0.013)	(0.013)	(0.018)	(0.048)
Very satisfied	0.053***	0.042***	0.006	0.011	0.019	-0.113***	0.054	0.154
	(0.023)	(0.017)	(0.015)	(0.018)	(0.015)	(0.011)	(0.080)	(0.087)

Cleanliness of HF								
Very dissatisfied	-0.011	-0.003	0.007*	0.004	0.0160**	0.015	-0.080	-0.071
	(0.008)	(0.009)	(0.004)	(0.005)	(0.008)	(0.010)	(0.051)	(0.052)
Dissatisfied	-0.028	-0.008	0.027*	0.017	0.038**	0.035	-0.155***	.246***
	(0.019)	(0.023)	(0.016)	(0.020)	(0.019)	(0.024)	(0.056)	(0.092)
Satisfied	0.777	-0.00009	-0.005	-0.003	-0.018*	-0.036***	-0.010	-0.021
	(0.069)	(0.0007)	(0.003)	(0.004)	(0.009)	(0.001)	(0.059)	(0.059)
Very satisfied	0.031**	0.010	-0.028*	-0.018	-0.037**	-0.033**	0.246***	-0.145**
	(0.011)	(0.033)	(0.017)	(0.021)	(0.018)	(0.013)	(0.075)	(0.059)

All regressions include the following observables: Women's age an educational background, household size, wealth quantile. The dependent variables are responses to endline survey questions elicited post incentive scheme. Standard errors are reported in parentheses clustered at village level. OP and IV-OP stand for ordered probit and instrumental variables-ordered probit models, respectively. The conditional Mixed Process (CMP) framework is used to estimate the IV-OP models. The sample is consisted of 8,174 women.
Significance levels are indicated as ***1%, **5%, *10%.

6.5 Heterogenous effects by the type of financing mechanism

In addition to the effects of incentive scheme across different health facilities, I ask whether the incentive effects vary based on the types of financing modalities. Tables 6 and 7 report the estimated effects of conditional incentive on women and children health outcomes, respectively, across contracted-out (NGO-regulated) and contracted-in (government-regulated) health facilities.

Table 6 Effects of conditional incentive on maternal health outcomes by the type of financing mechanisms

	Contracted-in facility beneficiaries		Contracted-out facility beneficiaries	
	OLS (1)	IV (2)	OLS (3)	IV (4)
Prenatal care	0.636	1.798	0.884*	0.517***
	(1.145)	(1.552)	(0.519)	(0.026)
Postnatal care	-5.799*	-1.964	2.91***	2.763**
	(3.282)	(4.403)	(0.994)	(1.182)
Skilled birth attendance	-0.161	0.067	0.517	1.806***
	(0.153)	(0.118)	(0.466)	(0.473)
Birth in health facility	0.120	0.139	0.129	0.760*
	(0.13)	(0.134)	(0.453)	(0.456)
Contraceptive use	-0.125**	-0.083	-0.060**	-0.059*
	(0.055)	(0.078)	(0.021)	(0.035)

All regressions include the following observables: Women's age and educational background, household size, wealth quantile. The dependent variables are responses to endline survey questions elicited post incentive scheme. Standard errors are reported in parentheses clustered at village level. OLS and IV stand for ordinary least square and instrumental variables models, respectively. The sample is consisted of 8, 174 women at child-bearing age. Significance levels are indicated as ***1%, **5%, *10%.

Based on the estimated coefficient of IV models in the column (2) of Table 6, the effects of conditional incentive on maternal health outcomes at the health facilities associated with the contracted-in mechanism is indistinguishable from zero at all levels of significance though the estimates of OLS models in the column (1) show the negative effects of incentive on the use of postnatal services and contraceptives. On the other hand, the incentive program is associated with an increase in the use of maternal health outcomes but contraceptives among the users of services at all the health facilities affiliated with contracted-out financing mechanism.

Table 7 Effects of incentive on children health outcomes by the type of financing mechanisms

	Contracted-in facility beneficiaries		Contracted-out facility beneficiaries	
Dependent Variable	OLS (1)	IV (2)	OLS (3)	IV (4)
Visit health facility	0.089*	0.102	-0.051	0.047*
	(0.052)	(0.080)	(0.056)	(0.028)
BCG vaccine	-0.014	-0.032*	0.046***	0.051***
	(0.017)	(0.011)	(0.001)	(0.010)
OPV vaccine	-0.118	-0.131	-0.084	-0.036
	(0.109)	(0.121)	(0.103)	(0.112)
Penta vaccine	0.296	0.509	0.027	0.043
	(0.383)	(0.449)	(0.108)	(0.119)
Measles vaccine	-0.025	-0.031	0.026*	0.034**
	(0.024)	(0.029)	(0.013)	(0.015)
Use of Vitamin A	0.001	0.017	0.011	0.025**
	(0.023)	(0.027)	(0.011)	(0.012)

All regressions include the following observables: Mother's age an educational background, household size, wealth quantile and infant's gender. The dependent variables are responses to endline survey questions elicited post incentive scheme. Standard errors are reported in parentheses clustered at village level. OLS and IV stand for ordinary least square and instrumental variables models, respectively. The sample is consisted of 7, 806 infants who are under five years old.
Significance levels are indicated as ***1%, **5%, *10%.

Coinciding with the incentive effects on maternal health outcomes, Table 7 provides that the incentive program has no impactful results on the children health outcomes across government-regulated health facilities. Even the incentive program lowers the probability of receiving BCG vaccines at these facilities. In contrast, the utilization of children healthcare services is significantly improved among service users at NGO affiliated health facilities. Administrative complexities, higher marginal cost of production and weak institutional and organizational capacities at contracted-in health facilities could eliminate any positive and meaningful impact of the incentive programs (Milliman & Prince, 1992; Sherry, Bauhoff, & Mohanan, 2017b).

6.6 Incentive and health facility efficiency analysis

Table 8 provides the estimated input-oriented technical efficiency scores of both comparison and treatment health facilities at baseline and endline surveys. At baseline survey, there is no statistically significant difference among the efficiency scores of comparison and treatment SCs, BHCs and DHs. However, on average, the treatment CHCs are 4.9% less technically efficient prior to incentive implementation. On the other hand, the technical efficiency scores of treatment health facilities experience an improvement at the endline survey.

Table 8 Effects of incentive on the technical efficiency of health facilities

Type of health facility	Baseline Survey		Difference	Endline Survey		Difference
	Comparison	Treatment		Comparison	Treatment	
Sub center (SC)	0.637	0.646	0.0091	0.653	0.709	0.056*** (0.061)
Basic Health Center (BHC)	0.814	0.824	0.011 (0.001)	0.809	0.869	0.060*** (0.143)
Comprehensive Health Center (CHC)	0.859	0.810	-0.049* (0.0017)	0.846	0.913	0.067** (0.140)
District Hospitals (DH)	0.882	0.868	0.014 (0.012)	0.991	0.991	0.008 (0.0008)

Independent group *t-test* is performed to test the mean difference of technical efficiency between treatment and comparison health facilities at baseline and endline surveys. Standard errors clustered at the village level. The distribution of test statistics is bootstrapped using wild-cluster bootstrap technique (Cameron, Gelbach, & Miller, 2008). Input-oriented technical efficiency scores are estimated using data envelopment approach (Tone, 2001).

7. Discussions and Conclusions

In an effort to align the socially desirable goals of health services delivery and behaviors of users and providers, developing countries have implemented both demand- and supply-sides performance-based financing initiatives. Application of these initiatives have provoked heated discussions among economists and public policy decision makers. This paper examines the effects of a supply-side incentive scheme on the demand for the pre-targeted maternal and children health services by exploiting data from a field experiment conducted in Afghanistan, a developing country. In contrast to the past burgeoning body of empirical literature, the identification strategy of the current work is based upon the fact that empirical analysis of field experiments could provide biased results if the compliance assumption of the experiment setting is violated (J. D. Angrist & Krueger, 2001; J. D. Angrist & Pischke, 2008). The bias-corrected estimates of this study show that supply-side conditional incentive is associated with an increase in the utilization of health care services among women and children, particularly at the smaller health facilities.

The effects of incentive scheme on the use of maternal and child health outcomes, concordant to much of the past literature, are modestly positive among users at smaller but strikingly negative at larger health facilities. What explains the differential effects of the incentive program at each level of care? Standard behavioral economics theory has recognized two main effects of a monetary incentive; the direct price effect and an indirect psychological effect (Gneezy, Meier, & Rey-Biel, 2011). A key explanation on the contrasting effects of incentive at the small and large facilities is the divergent directions of these two effects. At smaller health centers, number of providers is limited while the predetermined delivered services are comparable to those of larger facilities. The effects of income associated with the incentive is sizable to generate a convergent price and psychological effects. On the other hand, at the larger health facilities, although rewards could produce a trivial income, its psychological effect is minimum to galvanize the level of efforts. Relatedly, the differential effects of incentive across different health facilities could also be attributed to the patients' health status and the facility-related binding constraints. In general, patients receiving care at the smaller facilities are likely to be healthier than those admitted to the larger facilities. Moreover, unlike the upper level facilities, the required clinical procedures at lower level facilities are mostly performed by a primary care provider as outpatient. Therefore, incentivized schemes generate effective results at the lower levels of care (Gaynor et al., 2004). In the same line of reasoning, the limited resources at upper level facilities could be the source for some binding constraints (Mbiti et al., 2019). Conditional Incentive schemes could be partially effective in alleviating some of these constraints and consequently the impact might be limited on the outcomes.

Findings of this study have important public policy implications for both the design and underlying economic theory of incentive schemes. First, supply-side pay for performance mechanism improves the use of health care service by households at the lower levels of health system. Second, this study can potentially inform the decisions on the feasibility and execution of P4P modalities in the resource constrained contexts where organizational and institutional capacities are under development. Lastly, the smaller health facilities surrounded by a relatively smaller number of households are more efficient in terms of their resource allocations than the larger facilities. This could determine an important policy direction in redistribution of health care resources in the resource constrained contexts.

References

Altonji, J. G., Elder, T. E., & Taber, C. R. (2005). An evaluation of instrumental variable strategies for estimating the effects of catholic schooling. *Journal of Human Resources, 40*(4), 791-821.

Angrist, J. (1995). Conditioning on the probability of selection to control selection bias. In: National Bureau of Economic Research Cambridge, Mass., USA.

Angrist, J. D., Imbens, G. W., & Rubin, D. B. (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association, 91*(434), 444-455.

Angrist, J. D., & Krueger, A. B. (2001). Instrumental variables and the search for identification: From supply and demand to natural experiments. *Journal of Economic Perspectives, 15*(4), 69-85.

Angrist, J. D., & Pischke, J.-S. (2008). *Mostly Harmless Econometrics: An Empiricist's Companion*: Princeton university press.

Attanasio, O. P., Oppedisano, V., & Vera-Hernández, M. (2015). Should cash transfers be conditional? Conditionality, preventive care, and health outcomes. *American Economic Journal: Applied Economics, 7*(2), 35-52.

Baker, L. C., Bundorf, M. K., & Kessler, D. P. (2016). The effect of hospital/physician integration on hospital choice. *Journal of Health Economics, 50*, 1-8.

Balke, A., & Pearl, J. (1997). Bounds on treatment effects from studies with imperfect compliance. *Journal of the American Statistical Association, 92*(439), 1171-1176.

Banerjee, A., Deaton, A., & Duflo, E. (2004). Wealth, health, and health services in rural Rajasthan. *American Economic Review, 94*(2), 326-330.

Basinga, P., Gertler, P. J., Binagwaho, A., Soucat, A. L., Sturdy, J., & Vermeersch, C. M. (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.

Brunt, C. S., & Jensen, G. A. (2014). Pricing distortions in medicare's physician fee schedule and patient satisfaction with care quality and access. *Health Economics, 23*(7), 761-775.

Burde, D., & Linden, L. L. (2013). Bringing education to Afghan girls: A randomized controlled trial of village-based schools. *American Economic Journal: Applied Economics, 5*(3), 27-40.

Cameron, A. C., Gelbach, J. B., & Miller, D. L. (2008). Bootstrap-based improvements for inference with clustered errors. *The Review of Economics and Statistics, 90*(3), 414-427.

Cashin, C., Fleisher, L., & Hashemi, T. (2015). Verification of performance in results-based financing (RBF): The case of Afghanistan. *Washington, DC: World Bank.*

de Hennin, C., & Rozema, H. (2011). Study on Results-Based Programming and Financing, in Support of Shaping the Multi-Annual Financial Framework after 2013. Final Report.

De Janvry, A., & Sadoulet, E. (2006). Making conditional cash transfer programs more efficient: designing for maximum effect of the conditionality. *The World Bank Economic Review, 20*(1), 1-29.

De Walque, D., Gertler, P. J., Bautista-Arredondo, S., Kwan, A., Vermeersch, C., de Dieu Bizimana, J., . . . Condo, J. (2015). Using provider performance incentives to increase HIV testing and counseling services in Rwanda. *Journal of Health Economics, 40*, 1-9.

DHS. (2017). *Afghanistan Demographic and Health Survey*. Kabul, Afghanistan: Central Statistics Organization.

Eichler, R. (2006). Can "pay for performance" increase utilization by the poor and improve the quality of health services. *Background papers for the Working Group on Performance Based Incentives*.

Ellis, R. P., & McGuire, T. G. (1993). Supply-side and demand-side cost sharing in health care. *Journal of Economic Perspectives, 7*(4), 135-151.

Garen, J. E. (1994). Executive compensation and principal-agent theory. *Journal of Political Economy, 102*(6), 1175-1199.

Gaynor, M., Rebitzer, J. B., & Taylor, L. J. (2004). Physician incentives in health maintenance organizations. *Journal of Political Economy, 112*(4), 915-931.

Gertler, P. (2004). Do conditional cash transfers improve child health? Evidence from PROGRESA's control randomized experiment. *American Economic Review, 94*(2), 336-341.

Gneezy, U., Meier, S., & Rey-Biel, P. (2011). When and why incentives (don't) work to modify behavior. *Journal of Economic Perspectives, 25*(4), 191-210.

Handa, S., Peterman, A., Seidenfeld, D., & Tembo, G. (2016). Income transfers and maternal health: evidence from a national randomized social cash transfer program in Zambia. *Health Economics, 25*(2), 225-236.

Heckman, J. J. (1995). Randomization as an instrumental variable. In: National Bureau of Economic Research Cambridge, Mass., USA.

Hoderlein, S., & Mammen, E. (2007). Identification of marginal effects in nonseparable models without monotonicity. *Econometrica, 75*(5), 1513-1518.

Huillery, E., & Seban, J. (2013). Performance-based financing for health: experimental evidence from the Democratic Republic of Congo. *ENS-France Working paper*.

Ireland, M., Paul, E., & Dujardin, B. (2011). Can performance-based financing be used to reform health systems in developing countries? *Bulletin of the World Health Organization, 89*(9), 695-698.

Mbiti, I., Muralidharan, K., Romero, M., Schipper, Y., Manda, C., & Rajani, R. (2019). Inputs, incentives, and complementarities in education: Experimental evidence from tanzania. *The Quarterly Journal of Economics, 134*(3), 1627-1673.

McGuire, T. G., & Pauly, M. V. (1991). Physician response to fee changes with multiple payers. *Journal of Health Economics, 10*(4), 385-410.

Miller, G., & Babiarz, K. S. (2013). *Pay-for-performance incentives in low-and middle-income country health programs*. Retrieved from

Milliman, S. R., & Prince, R. (1992). Firm incentives to promote technological change in pollution control: Reply. *Journal of Environmental Economics and Management, 22*(3), 292-296.

Moffitt, R. A. (1996). Identification of causal effects using instrumental variables: Comment. *Journal of the American Statistical Association, 91*(434), 462-465.

Mwabu, G. (2007). Health economics for low-income countries. *Handbook of Development Economics, 4*, 3305-3374.

NHA. (2014). *Afghanistan National Health Accounts with Disease Accounts*. Kabul, Afghanistan.

NHS. (2016). *National Health Strategy 2016–2020*. Kabul, Afghanistan Retrieved from http://moph.gov.af/en/Page/579.

Peabody, J., Shimkhada, R., Quimbo, S., Florentino, J., Bacate, M., McCulloch, C. E., & Solon, O. (2011). Financial incentives and measurement improved physicians' quality of care in the Philippines. *Health Affairs, 30*(4), 773-781.

Polachek, S. W., Das, T., & Thamma-Apiroam, R. (2015). Micro-and macroeconomic implications of heterogeneity in the production of human capital. *Journal of Political Economy, 123*(6), 1410-1455.

Powell-Jackson, T., Mazumdar, S., & Mills, A. (2015). Financial incentives in health: New evidence from India's Janani Suraksha Yojana. *Journal of Health Economics, 43*, 154-169.

Renmans, D., Holvoet, N., Orach, C. G., & Criel, B. (2016). Opening the 'black box'of performance-based financing in low-and lower middle-income countries: a review of the literature. *Health Policy and Planning, 31*(9), 1297-1309.

Roodman, D. (2011). Fitting fully observed recursive mixed-process models with cmp. *The Stata Journal, 11*(2), 159-206.

Rosenzweig, M. R., & Wolpin, K. I. (1986). Evaluating the effects of optimally distributed public programs: Child health and family planning interventions. *The American Economic Review, 76*(3), 470-482.

Shea, J. (1997). Instrument relevance in multivariate linear models: A simple measure. *Review of Economics and Statistics, 79*(2), 348-352.

Sherry, T. B., Bauhoff, S., & Mohanan, M. (2017a). Multitasking and Heterogeneous Treatment Effects in Pay-for-Performance in Health Care: Evidence from Rwanda. *American Journal of Health Economics*.

Sherry, T. B., Bauhoff, S., & Mohanan, M. (2017b). Multitasking and heterogeneous treatment effects in pay-for-performance in health care: evidence from Rwanda. *American Journal of Health Economics, 3*(2), 192-226.

Thornton, J., & Eakin, B. K. (1997). The utility-maximizing self-employed physician. *Journal of Human Resources*, 98-128.

Tone, K. (2001). A slacks-based measure of efficiency in data envelopment analysis. *European journal of operational research, 130*(3), 498-509.

Trani, J.-F., Kumar, P., Ballard, E., & Chandola, T. (2017). Assessment of progress towards universal health coverage for people with disabilities in Afghanistan: a multilevel analysis of repeated cross-sectional surveys. *The Lancet Global Health, 5*(8), e828-e837.

Van de Poel, E., Flores, G., Ir, P., & O'donnell, O. (2016). Impact of performance‐based financing in a low‐resource setting: a decade of experience in Cambodia. *Health Economics, 25*(6), 688-705.

Vujicic, M. (2009). How you pay health workers matters: A primer on health worker remuneration methods. *World Bank, Washington DC.*