

Osmani, Ahmad Reshad

Article — Manuscript Version (Preprint)

Conditional Cash Incentive and Use of Health Care Services: New Evidence from a Household Experiment

Journal of Family and Economic Issues

Suggested Citation: Osmani, Ahmad Reshad (2020) : Conditional Cash Incentive and Use of Health Care Services: New Evidence from a Household Experiment, Journal of Family and Economic Issues, ISSN 1573-3475, Springer, Berlin, Heidelberg, Iss. forthcoming, <https://doi.org/10.1007/s10834-020-09740-6>

This Version is available at:

<https://hdl.handle.net/10419/226352>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Conditional Cash Incentive and Use of Health Care Services: New Evidence from a Household Experiment

Abstract

Studies have recently provided insights into the effects of incentive modalities in the health care sector. However, there is insufficient evidence on the underlying causes of the 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 on 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 the supply-side conditional incentive on the utilization for health care services. After two years, the conditional incentive increased the use of pre-targeted maternal and children health care services among the households at lower levels and at contracted-out health facilities. Additionally, the incentive scheme is associated with sizable efficiency gains at the facility level. These gains are realized at the expense of deterring service users' satisfaction with physicians' communication qualities. This study establishes that margins of improvement do exist in the supply-side performance conditioning on an organizational structure and the service contractual arrangements of health facilities. This work provides a framework for the plausible implementation of incentive policies in the health care sector.

JEL codes: D1, C93, I12, J41

Keywords: Conditional cash incentive, field experiment, target-income hypothesis, noncompliance, instrumental variables

Introduction

Applications of financial incentives in the health systems of developing countries have been increasing in prominence among the policy instruments intended to scale up the demand and supply of health care services (Ellis and McGuire 1993). Although increases in the use and provision of 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, the impacts of an incentive package on the delivery of services could differ by payment mechanism, type of health care facility, and involvement of multiple stakeholders in the health system. These differences can have important public policy implications for the impactful provision of incentives in resource-constrained health systems. This study evaluates the causal effects of supply-side monetary incentives on the use of health care services, by exploiting a large-scale, carefully designed 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's and children's health outcomes across different levels of health facilities (HFs). Although different dimensions of HF may influence the outcome of an incentive policy, the main focus of this study was on the size and service composition that have played a substantial role in the empirical and theoretical literature (Baker et al. 2016). After relaxing the compliance assumption in the experimental design, the empirical model estimated the casual effects of supply-side conditional incentives on the use of health care services. In addition to evaluating the incentive effects across various levels of the health system, the study sample was bifurcated into users affiliated with the contracted-in (government-regulated)

and contracted-out (nongovernment-regulated) HFs. The goal was to investigate whether these contracting arrangements resulted in differences in medical staff behavior after receiving conditional monetary incentives.

The second part of this study investigates the experimental variation in the effects of conditional incentives on beneficiaries' satisfaction with the quality indicators of the provision of health care services; specifically, both physicians' behavioral and HF-related quality indicators were evaluated. The last part of this analysis explores the effects of the incentive design on the efficiency improvement of HFs. To perform this analysis, a nonparametric estimation method was used to elicit the technical efficiency of facilities pre- and post-realization of the monetary incentives.

The field experiment of this work was conducted with a sample of 11 provinces (6,484 households, 143 different levels of HFs, and 286 villages) in two time periods (baseline and endline surveys) with financial and technical support from the World Bank (WB) and the Ministry of Public Health of Afghanistan. The experiments and randomizations are conducted at the HF, village, and household levels. Initially, all 374 HFs in nine provinces¹ of the country were stratified by the facility type. The facilities were matched in pairs based on their outputs and staffing, and subsequently, eight pairs were randomly selected for each province. The treatment and control facilities were randomly chosen within each pair. Next, the total number of villages were identified using the catchment area of each facility². For each facility, two villages (treatment and control) were selected. Finally, all the households in the selected villages were enlisted, and the required number of households were randomly identified. The incentive payments are reimbursed quarterly

¹ The incentive scheme was implemented in 11 provinces. However, two provinces were dropped because of insecurity.

² Catchment area of each health facility indicated the number of people served by that health facility.

to the health care workers at the facility level and conditional on performance. The administered monetary incentive if received was between 11%–28% of a worker's base monthly salary.

The first set of estimates showed that conditional incentives for the health workers (physicians, nurses, midwives, and administrative workers of the HFs) had small but statistically significant effects on the use of health care services by the beneficiaries of treatment at lower-level HFs. On average, at a subcenter (SC) facility, the incentive pay was associated with a 2.2 and 1.6 percentage point increase in the use of women's pre- and postnatal health services, respectively. By the same token, the incentive pay increased the likelihood of skilled births and institutional deliveries at SC facilities. Notwithstanding, the magnitude of these estimates proportionally decreased as the size of HFs increased, to the extent that the estimated effects of an incentive had a reversed sign at the district hospitals (DHs). The treatment effects on children's health outcomes superseded the pattern of maternal effects across various levels of the HFs, such that at the SC and basic health center (BHC) levels, the probability of visiting HFs to seek care and obtaining the required types of childhood vaccinations increased considerably. These point estimates declined at upper-level health centers. Taken together, the findings revealed that the use of both women's and children's services widely differed when the effects of an incentive were evaluated at each HF.

The second set of results presented how incentivizing health care providers changed the quality indices from beneficiaries' perspective. Specifically, the effects of a conditional incentive on patient centeredness and facility accessibility were documented as the key quality signals of the health services provision. The beneficiaries assigned to the treatment facilities reported mostly lower satisfaction in the patient-related quality indices than the beneficiaries at the endline survey did. The normalized scores of physician's respectfulness with patients were 2.4 percentage points

higher at the incentivized facilities' SC, and for the baseline survey, these scores were not statistically different between the comparison and treatment facilities. By contrast, the empirical models could barely capture or imprecisely estimated the effects of an incentive on HF related indices such as cleanliness and availability of medicine at treatment facilities. The final set of findings documented a small but significant effect of supply-side incentives on the efficiency gradients of treatment HFs. Using the piecewise linearity concept of data envelopment analysis (DEA), a 6%–7% increase in the input-oriented efficiency scores of treatment facilities was detected compared with these facilities at the endline survey.

The association between monetary incentive and the service provider's performance can be explained by standard behavioral economics theory within the framework of the target-income and work-leisure trade-off hypotheses (McGuire and Pauly 1991; Thornton and Eakin 1997). Under either framework, the increase in the number of services seems to be driven by the relative strength of income or the price effect.

The contribution of this paper to the literature is threefold. First, unlike the literature, this work estimates the effects of supply-side incentives at various levels of HFs. This extension is crucial for understanding the feasibility of incentive schemes in developing country contexts where provision of services is partitioned into hierarchical levels. In addition, this paper investigates the effectiveness of incentive pay schemes with regard to the health care services' outsourcing arrangements. Inadequate organizational and institutional capacities in low-income settings have resulted in a focus on the role of contracting-out mechanisms. The estimates of this work provide new evidence on the contribution of outsourcing mechanisms in the feasibility and success of incentive-based mechanisms. Second, a simple theoretical model based on the target-income hypothesis within the principal–agent problem setting is crafted to explicitly determine the

underlying mechanism of supply-side financial incentives in the health care system. Last and most importantly, this work contributes to the growing body of research on empirical modeling of experimental datasets. Because this study evaluates the effects of incentives across different levels of HFs, there are at least two sources of selection bias. The first source is related to a noncompliance problem; the second source might be related to the population heterogeneity in the catchment area of each HF. Although the experimental design of this study partially addresses the selection bias, the executed empirical model eliminates the selection bias associated with the noncompliance problem. Consequently, the bias-corrected effects of the conditional incentive on the use of health services are obtained.

The remainder of this paper is organized as follows; the next two sections discuss the country context and related literature on the effects of incentive payments in health care sector, followed by an introduction to the theoretical construct and research design of the study; next, the empirical results are presented; and finally, the last section discusses the results and concludes by assessing the findings and policy implications.

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 significantly improving 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 87%, and it was less than 10% in 2001 (DHS 2017). IN addition to relative macroeconomic stability, these gains are the consequence of the smooth implementation of the

basic package of health services (BPHS) and the essential package of hospital services (EPHS) as the country's foundation for its health system. BPHS and EPHS, with nearly 2,300 HFs across 34 provinces of Afghanistan, are the initial points of contact for receiving preventive and curative health care services (NHS 2016). These HFs are partitioned into provincial and DHs, comprehensive health centers (CHCs), BHCs, SCs, and mobile health teams to ensure a hierarchical and well-defined referral system. Each of the aforementioned HFs has a defined population catchment area serving 2,500–150,000 individuals. To increase the organizational and institutional capacities in the public sector and the existence of macro-level funding agencies (e.g., the WB, United States Agency for International Development, and European Commission), the Afghan government has implemented the provision of health care services through nongovernmental organizations (NGOs), based on a contractual arrangement, namely, the contracted-out approach. Approximately 77% of the population receive health care services through the latter one, and less than 25% of individuals are served by the government-regulated mechanism.

Although initial efforts toward system building and improving health outcomes have been implemented, the country several challenges concerning health care coverage, quality, and financing. These obstacles could negatively affect every citizen but mostly women and children because they are the disadvantaged groups in Afghanistan. Evidently, utilization of maternal and child health care services has lagged. Characterized by Afghanistan's health and demographic survey DHS (2017) estimates, 59% and 51% of pregnant women have received one skilled prenatal care and have delivered in presence of a skilled birth attendant, respectively. Similarly, use of modern contraceptives and postnatal care have been stagnant since 2010. Additionally, the government contribution to the total health expenditure (THE) is 5%, and Afghan's out-of-pocket

health expenses is approximately three-quarters of THE. The insufficient quality of service provisions 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 regional countries (NHA 2014).

Since early 2010, increasing political instability and security concerns have consistently threatened the efficient delivery of health care services throughout the country. In addition, the reduction in foreign aid for the health system and the economic downturn in the country represent other obstacles to the provision of responsive health services and their sustainability (NHS 2016). These unfavorable circumstances have started a laudable public policy debate to maintain current gains and promote health service delivery without quality compromise. In response, the government of Afghanistan with financial and technical support from the WB has piloted a supply-side conditional incentive scheme to further improve accessible, equitable, and quality health services, with a focus on women's and children's health outcomes. More specifically, incentive pay that interfaces health care utilization outcomes with the supply-side (health care providers) incentives was initiated in 2010. Primarily, this paper documented the evidence on the effects of the current performance-based incentive initiative on the use of health care services by women and children. Additionally, the impact of an incentive scheme on patient satisfaction based on physician communication qualities was evaluated. The rich experimental design of this work attempted to obtain the bias-corrected estimated effects of the incentive program.

Pay-for-Performance and Related Literature

Pay for performance (P4P) has been defined as the 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 the utilization of services and overall system performance (Van de Poel et al. 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 et al. 2015; Handa et al. 2016; Powell-Jackson et al. 2015; Renmans et al. 2016; Trani et al. 2017). P4P has been associated with the demand- and supply-sides. On the demand-side, a conditional cash transfer has been a well-established mechanism to incentivize users of services (e.g., households, women, and children) when utilizing the available services as necessary (De Janvry and Sadoulet 2006; Gertler 2004). On the supply side, the performance-based incentive to health care providers (e.g., medical doctors, nurses, midwives, and other health personals) has been widely implemented (de Hennin and Rozema 2011; Ireland et al. 2011). This paper is related to a stream of literature on the supply-side of P4P initiatives.

Essentially, the production of health care services has been determined by the 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 et al. 2015). Although structural inputs are sufficiently provided, in most low-income settings, the quality of the processes has been unsatisfactory because of the poor performance of health care providers. Monetary and nonmonetary rewards aiming to cause behavioral changes among the health care providers might improve their performance (Eichler 2006). Performance-based incentives target providers at the individual or facility level. The former has been implemented predominately in developed economies (Vujicic 2009). Facility-based incentives, which improve efficient resource allocation especially when service production entails a strong

interdependence among group members (Gaynor et al. 2004), have been a dominant strategy in low-income countries. Unlike other empirical health economics studies, there is a large amount of literature related to P4P that has used an experimental research framework because of 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 of 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 and no statistically significant effect on the number of prenatal visits and child immunization services. One of the potential limitations of their study was the simultaneous effects of decentralization policies during the study periods. Relying on a quasi-experimental design, De Walque et al. (2015) showed that a P4P scheme in Rwanda increased HIV testing among married and discordant couples by 10.2 and 14.17 pp, respectively. Using two waves of demographic and health surveys from the same setting, Sherry et al. (2017b) analyzed the effects of a PBF mechanism on health care utilizations, outcomes, and the program's unintended consequences. Although they provided insights into the positive effects of the program on the utilization of services, the impact of the scheme on health outcomes was indistinguishable from zero.

Observing no statistically significant effects of the incentive program on prenatal and infant vaccination rates, Van de Poel et al. (2016) researched the impact of PBF in Cambodia. They found that P4P schemes improved both the service utilization and quality indicators in the long term, especially when heavy reliance was placed on the user's fee. Their findings were limited because of the selection bias due to there being an insufficient experimental research design. Additionally,

the parallel implementation of other reforms in the health care system of Cambodia could be a driver of the results of its PBF program.

To assess the impact of incentives on the clinical quality of health care services in the Philippines, Peabody et al. (2011) investigated the effects of P4P on the individual and system levels by 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, a physician's bonus, unlike a system-level bonus, had no statistically significant effect on the utilization of services. Improving the extrinsic but deterring the intrinsic motivation with 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.

This paper deviates from the literature in at least three ways. First, this study evaluates the causal effects of a supply-side conditional incentive at different levels of care was a policy relevant and less studied topic. Second, in contrast to the literature, the empirical modeling of this paper relaxes the perfect compliance assumption in the experimental setting. Third, in addition to estimating the treatment effects on the efficiency of the contracted-in and contracted-out HFs, this study provides causal estimates of an incentive on the use of health care services among the recipients under both types of contractual arrangement.

Response to Monetary Incentive: Theory

A health care provider's behavioral response to incentive pay could be rationalized within the target-income and work-leisure trade-off hypotheses (McGuire and Pauly 1991; Thornton and Eakin 1997). An earlier theoretical framework suggested that providers customize the composition

and volume of the offered services to reach to a targeted level of income, and the latter proposed that a physician as a utility-maximizing agent trades the volume of services for leisure time. In either framework, the increased supply of service was associated with the relative strength of income or the price effect. To cement this knowledge, a simple theoretical model akin to Garen (1994) was crafted to explicitly represent the health care worker's response to the incentive payments. Assuming that an implementer (e.g., NGOs or government) and provider at an HF were the principal and agent, respectively, the agent determined the level of effort necessary to produce a Q quantity of health care services to the service beneficiaries. Quantity Q was a vector comprising several types of services $Q = (q_1, q_2, q_3, \dots, q_n)$. Increasing the quantity of health care services Q by the agent was the potential benefit to the principal and was revealed in the forms of increased health care services $Y = (y_1, y_2, y_3, \dots, y_n)$ used by beneficiaries (e.g., women, children). Implicit and explicit incurred costs to the agent to provide Y services were shown by $C(Y)$, which is a twice differentiable and positive function. Provision of Y services depended upon the HF's fixed resources $\theta(R)$, the health worker's effort $\pi(E)$, and other HF or beneficiary's unobservable factors:

$$Y = \pi(E) + \theta(R) + \zeta. \quad (1)$$

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

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

however, the principal could incentivize the health worker to maximize her level of effort $P \geq C(Y)$, conditional on achieving a predetermined target S_t . Then, the payment structure \mathbb{R} for the agent in the presence of an incentive pay scheme was

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

Substituting Eq. (3) into Eq. (1) yielded the incentivized provision of services, and it was written as

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

The risk-neutral agent solved the following profit maximization problem while considering an incentive package:

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

Further, Eq. (5) could be rewritten using the cumulative density function of ε (F):

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

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

Experimental Design

A conditional monetary incentive was executed in the health system of Afghanistan with specific objectives³. The pay scheme was administered and implemented nationally⁴ at contracted-in and contracted-out HFs. The payments associated with the incentive were reimbursed quarterly to the health care workers at each facility level, conditional upon their performance and increase in the quantity of pre-targeted health care services. Initially, HFs reported their monthly provided services to their provincial managing office. These reports were then subject to quarterly verification process by three different methods. The following were then conducted: random interviews of individuals who were reported as beneficiaries (users of services) of the HF, random visits to the HF by a group of independent monitors, and finally matching the facility monthly reports with the registries of the health information system.

NGOs had the 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 the geographical location of the facility. Once verified by the Ministry of Public Health (MoPH), the incentive payments were transferred to the managing offices of contracted-in and contracted HFs along their fixed budget. Distribution of the incentive pay at each HF solely depended on the health care providers in that facility. Providers' salary scale, level of an individual's effort, and/or an equal share of payment to every member of the facility

³ (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.

⁴ In 11 out of 34 provinces in the country.

are some of the reported incentive distributional mechanisms. Nonetheless, the bonus if received was between 11% and 28% of the worker's base monthly salary. To evaluate the impact of this incentive scheme, large-scale field experiments have been performed with financial support from the WB and technical support from the MoPH.

The experimental design of this study was based on randomization at the HF, village, and household levels. In the first stage, all 374 HFs in nine provinces of the country were stratified by the facility type (DH, CHC, BHC, and SC), and the proportion of each facility type was calculated in a particular province⁵. For each province, the HFs were matched in pairs based on similarities (average number of outpatient visits in one month and number of staffs in the facility), and then eight pairs were randomly selected. Treatment and comparison facilities were chosen within each pair—one facility is a comparison, and the other facility is the treatment to ensure that the comparison facility within one pair is very similar to the treatment. In total, 140 HFs in nine provinces, which is approximately 16 facilities for each province, were randomized. In the second stage, the catchment area⁶ of each HF was identified to properly determine the total number of villages. Thereafter, two villages⁷ (comparison and treatment village) were randomly selected for each HF for all 140 HFs. In total, 280 villages were randomly selected. In the third stage, all the households in the randomized villages (280 villages) for each province were listed. The required⁸ number of households was randomly selected from those comparison and treatment villages (Table 1). In summary, the sample size for this study comprised 140 HFs, 280 villages, and 6,848

⁵ The incentive scheme was implemented in 11 provinces. However, two provinces were dropped because of insecurity.

⁶ Catchment area of each health facility indicated the number of people served by that health facility.

⁷ Villages with less than 50 households had to be linked with the nearest village in that facility's catchment area and treated as one unit of randomization.

⁸ Security, logistical, and statistical facts are considered in choosing the final sample.

households.

Econometric Model

Considering the focus of policy context in the health care sector, the main outcome of interest was the utilization indicators for different types of services. The empirical model of this paper exploited the random variation in the use of these services, induced by performance-based pay, to estimate the effects of a supply-side conditional incentive. A natural place to initiate the empirical analysis of experimental data was by comparing the outcomes in the premises of treatment status. Assuming perfect compliance in the experiment setting, the random assignment of observations allowed us to obtain the average causal effect of treatment on the treated (ATET) as shown:

$$E[y_i | d_i = 1] - E[y_i | d_i = 0] = E[y_{1i} - y_{0i} | d_i = 1] = E[y_{1i} - y_{0i}] \quad (7)$$

where y_i is the outcome variable for observation i , and d_i is a binary variable that shows whether observation is assigned to treatment or not. However, the aforementioned formulation is unable to recover the ATET when the compliance assumption is violated (Balke and 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, by contrast, might transform the comparisons by 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, the first step is 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 as

$$y_i = \alpha + \phi d_i + \varphi_i. \quad (8)$$

The treatment effects ϕ cannot be consistently estimated when there is a correlation between the idiosyncratic error φ_i and the treatment status d_i . However, the random assignment of intended treatment status Z_i could provide a solution to disentangle the bias-corrected causal estimates of the incentive. Realizing 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

$$\{E[y_i | Z_i = 1] - E[y_i | Z_i = 0]\} / \{E[d_i | Z_i = 1] - E[d_i | Z_i = 0]\} = \phi. \quad (9)$$

The treatment effects obtained from Eq. (9) with the constant-treatment effects assumption are sufficient, but in the real world, this practice is unrealistic. For instance, the 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), and others may be little affected or not at all. Additionally, there might be heterogeneity in the treatment effects among beneficiaries. To overcome these hurdles, J. D. Angrist et al. (1996) proposed 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 LATE, we let the potential treatment assignments be d_{0i} and d_{1i} for observation i , and Z_i takes a binary value of 0 or 1. Then, the observed treated status could be

written as

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

The aforementioned 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 et al. 2005). Exclusion restriction implies that the instrument does not have a direct effect on the potential outcome of interest. Second, the incentive offer might affect the use of health care 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 manner (Hoderlein and Mammen 2007). To formalize the idea, the following models are empirically estimated:

$$d_i = \beta_0 + \beta_1 Z_i + \beta_2 X_{it} + \vartheta_i \quad (11)$$

$$Y_i = \gamma_0 + \gamma_1 \hat{d}_i + \gamma_2 X_i + v_i. \quad (12)$$

Equations (11) and (12) show 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 health care services, patients' satisfaction) could be obtained even if noncompliance existed (Heckman 1995; Moffitt 1996). In the empirical results 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.

Empirical Results

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 might be correlated with the health care services outcomes (Burde and Linden 2013).

Table 1 Balance of and differences in the treatment and comparison groups at the baseline and endline surveys

	Baseline survey			Endline survey		
	Comparison	Treatment	Difference	Comparison	Treatment	Difference
	(1)	(2)	(3)	(4)	(5)	(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) declines.

Table 1 provides a measurement of differences between the composition of the comparison and treatment groups at the baseline and endline surveys. In both surveys, there was a strong balance between the number of HFs, villages, households, women, and infants. Generally, all the differences were small and consistent over two study periods; therefore, the randomization appears to be successful. Table 2 reported the result of the equality test for the observable characteristics of the women included in the study sample. In the baseline survey, none of the differences between the comparison and treatment groups were statistically distinguishable from zero, except for family size. For the endline survey, however, all the differences in the socioeconomic characteristics of respondents were statistically not significant. Apparently, the lack of differential attrition in the control and treatment groups might suggest that the estimates associated with the incentive were not biased (J. Angrist 1995). However, noncompliance of respondents in the endline survey raised an important concern when estimating the effects of a treatment program.

Table 2 Observable characteristics of women and infants in the treatment and control groups before and after the 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 (0.23)	31.50	31.76	0.259 (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 (Mean)	9.62	10.16	0.542* (0.299)	8.14	8.31	0.174 (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 comprises 8,174 women of child-bearing age. Significance levels are indicated as ***1%, **5%, and *10%.

Incentive and maternal health outcomes

Table 3 shows the causal effects of the incentive on the use of health care services by women across HFs. The odd and even columns present the estimated coefficients of the ordinary least square⁹ (OLS) of the IV models, respectively. The main outcomes of interest were prenatal and postnatal care, skilled birth and birth at an HF, and contraceptive use. Column (2) corresponds to the IV estimated effects of the incentive at an SC facility. Except for contraceptive use, all other

⁹ OLS estimates were provided for comparison purpose.

outcomes revealed a statistically and economically significant and positive estimate associated with the incentive. An increase of 2.2 and 1.6 was implied in the number of pre-and postnatal care visits, respectively, after providing incentives to HFs. In addition, skilled births and births at the HF were induced by 9.1 and 3.2 pp, respectively. Again, the incentive effects in column (4) remained statistically significant at the BHC facilities, but the magnitude was smaller than the SC estimates. Column (6) indicates that the positive effects of the incentive on the skilled birth and births at a HF faded out at CHC level, although these effects declined by more than half for pre-and postnatal care outcomes. Notably, the estimated effects of the incentive reversed signs at a DH. The column (8) estimates indicated that the number of pre- and postnatal care use decreased by 2.7 and 4.6, respectively. Likewise, skilled birth and birth in an HF decreased by 7.4 and 17.9 pp, respectively. In contrast to other HFs, the effect of incentive on the use of contraceptives was precisely estimated at the DH level, and it lowered the use of contraceptives by 19.7 pp. Taken together, these results suggest that the effect of the supply-side incentive was positive and larger at smaller HFs and negative and relatively smaller at large HFs. Finally, the IV estimator was likely to provide a consistent estimate because the unobserved utilization determinants were likely not uniformly distributed across women living in the catchment areas of each HF.

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

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

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

Incentive and children health indicators

We have established that providing the supply-side incentive significantly affected women's health outcomes across HFs; next, the effects of the incentive scheme on child health indicators are investigated. Table 4 presents the results of OLS and IV regressions on pre-targeted child health outcomes (number of visits at an HF, Bacille Calmette-Guérin [BCG], oral polio vaccine (OPV), Penta and measles vaccines, and use of vitamin A). Starting with SCs, incentivizing health care providers significantly increased the use of HFs and receiving certain types of vaccines by children. OPV and Penta vaccination rates were improved by 26.2 and 17.9 pp, respectively. The likelihood of visiting an SC for receiving care increased by 14.7% among children who lived in the catchment area of a treated facility. These positive effects remained statistically significant at BHCs but with a relatively smaller magnitude than at SCs. Although the incentive scheme induced the rates for the measles vaccine by 5.5 pp at the comprehensive HFs, it negatively affected both the likelihood of visiting health center and rate of Penta vaccine. Similar to the effects of the incentive on maternal health outcomes, children's health indicators were negatively associated with the incentive scheme at DHs, emphasizing that the incentive decreased the use of health care services as the size of HFs increase.

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

	Subcenter (SC)		Basic Health Center (BHC)		Comprehensive Health Center (CHC)		District Hospitals (DHs)	
Dependent Variables	OLS (1)	IV (2)	OLS (3)	IV (4)	OLS (5)	IV (6)	OLS (7)	IV (8)
Visit health facility	0.111*** (0.046)	0.147** (0.052)	0.116** (0.041)	0.127*** (0.054)	-0.152*** (0.032)	-0.146*** (0.035)	-0.1633 [†] (0.088)	-0.155 [†] (0.091)

BCG vaccine	0.039** (0.017)	0.025 (0.019)	0.021* (0.01)	0.027 ** (0.013)	0.045** (0.019)	0.043 (0.121)	0.065 (0.049)	-0.091** (0.046)
OPV vaccine	0.119 (0.107)	0.262** (0.112)	-0.331** (0.155)	0.186*** (0.079)	-0.238 (0.213)	-0.262 (0.236)	-0.423 (0.553)	-0.283 (0.583)
Penta vaccine	0.810** (0.388)	0.179*** (0.049)	0.126** (0.054)	0.119** (0.059)	-0.288** (0.129)	-0.296** (0.145)	-0.051 (0.141)	-0.005 (0.149)
Measles vaccine	-0.027 (0.024)	-0.039 (0.027)	0.019 (0.016)	0.029 [†] (0.018)	0.045* (0.025)	0.055** (0.028)	0.055 (0.069)	-0.111 [†] (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, educational background, household size, and wealth quantile, and infant's gender. The dependent variables are responses to endline survey questions elicited after the incentive scheme. Standard errors are reported in parentheses clustered at the village level. OLS and IV denote ordinary least square and instrumental variables models, respectively. The sample comprises 7,806 infants under five years old.

[†] p < .1, *p < .05, **p < .01, and ***p < .001 show the significance levels.

Incentive and service users' satisfaction

Next, this section presents the results related to the effects of the supply-side incentive on the satisfaction of service users across each HF. Table 5 presents the estimates corresponding to the patient and HF categorical outcomes that encompassed both a physician's behavioral and HF-related indicators. Respectfulness, patient privacy, time spent with patient, and explanation of diseases to the patient were a physician's quality indicators. The HF-related quality outcomes were the availability of medicine and the cleanliness of the facility. For each outcome across the HFs, the estimated coefficients of the ordered probit and the instrumental variables-ordered probit were reported. The results in column (2) suggest that offering incentives to service providers at an SC facility negatively affected physician's respectfulness toward service users. The incentive increased service users' experience with time spent with patient outcome by 4.2 pp. However, the causal effects of the incentive on the remaining outcomes were indistinguishable from zero at an SC facility. Except for the negative effects on the beneficiary's satisfaction with the patient privacy outcome, the provider's incentive appeared to have no meaningful impact on users' satisfaction outcomes at a basic HF. Strikingly, the incentive negatively affected many of the user's satisfaction outcomes at comprehensive HFs. Compared with the control group, on

average, doctors spent 11% less time with the service users at treatment HFs. Additionally, patients were less satisfied with a physician's respectfulness at the time of service delivery. Furthermore, both cleanliness and availability of medicine declined by approximately 3.5 and 4 pp, respectively. Finally, similar to comprehensive HFs, the estimates in column (8) illustrate that the conditional incentive lowered the user's satisfaction with the physician's respectfulness and explanation outcomes at DHs. Likewise, facility-related outcomes were worsening among treated facilities.

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

	Subcenter (SC)		Basic Health Center (BHC)		Comprehensive Health Center (CHC)		District Hospitals (DHs)	
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 (0.012)	0.024 [†] (0.011)	0.003 (0.005)	0.0001 (0.007)	-0.027 [†] (0.014)	0.017*** (0.001)	-0.014 (0.031)	-0.009 (0.036)
Dissatisfied	0.008 (0.015)	0.030 [†] (0.018)	0.009 (0.017)	0.003 (0.021)	-0.027 [†] (0.014)	0.080*** (0.025)	-0.033 (0.071)	-0.021 (0.086)
Satisfied	-0.002 (0.003)	-0.005 (0.005)	-0.004 (0.007)	-0.001 (0.009)	0.016 [†] (0.009)	0.019 (0.018)	0.008 (0.019)	-0.115*** (0.020)
Very satisfied	-0.012 (0.024)	-0.049* (0.021)	-0.008 (0.015)	-0.003 (0.019)	0.037 [†] (0.011)	0.029 (0.118)	0.039 (0.084)	0.025 (0.102)
Explanation								
Very dissatisfied	-0.007 (0.013)	0.0003 (0.016)	0.007 (0.007)	0.0003 (0.009)	-0.013 (0.013)	-0.085** (0.016)	-0.028 (0.028)	-0.078 (0.056)
Dissatisfied	-0.001 (0.017)	0.0003 (0.021)	0.013 (0.015)	0.0003 (0.0004)	-0.017 (0.016)	-0.078** (0.020)	-0.107 (0.081)	0.157** (0.073)
Satisfied	0.006 (0.001)	-0.0002 (0.012)	-0.007 (0.008)	-0.017 (0.014)	0.009 (0.008)	0.009 (0.010)	0.063 (0.052)	-0.192*** (0.063)
Very satisfied	0.012 (0.011)	-0.0004 (0.025)	-0.011 (0.014)	-0.009 (0.014)	0.022 (0.021)	0.024 (0.026)	0.071 (0.058)	0.112 (0.145)
Availability of medicine								
Very dissatisfied	-0.001 (0.025)	0.019 (0.030)	0.034** (0.016)	0.022 (0.020)	0.031** (0.020)	0.048 [†] (0.025)	-0.121 (0.122)	-0.224 (0.182)
Dissatisfied	-0.004 (0.001)	0.008 (0.012)	0.019** (0.009)	0.012 (0.011)	0.026** (0.013)	0.031** (0.016)	-0.099 (0.098)	0.126 (0.094)
Satisfied	0.006 (0.016)	-0.012 (0.019)	-0.026** (0.012)	-0.016 (0.015)	-0.021** (0.014)	-0.035** (0.018)	0.061 (0.068)	-0.085 [†] (0.046)
Very satisfied	0.008 (0.019)	-0.015 (0.023)	-0.027** (0.013)	-0.020 (0.016)	-0.036** (0.018)	-0.044 [†] (0.023)	0.159 (0.159)	0.184 (0.101)
Patient privacy								
Very dissatisfied	-0.007 (0.007)	0.002 (0.008)	-0.005* (0.003)	-0.005** (0.002)	0.005 (0.007)	0.007 (0.008)	0.044 (0.041)	0.012 (0.048)
Dissatisfied	-0.012 (0.011)	0.003 (0.014)	-0.0025 [†] (0.014)	-0.027** (0.011)	0.010 (0.015)	0.017 (0.019)	0.041 (0.036)	0.010 (0.043)
Satisfied	-0.015 (0.014)	0.003 (0.017)	-0.009 [†] (0.005)	-0.001 (0.007)	0.005 (0.006)	0.006 (0.007)	0.038 (0.036)	0.008 (0.032)
Very satisfied	0.035 (0.032)	-0.007 (0.038)	0.038* (0.021)	0.042 (0.027)	-0.011 (0.027)	-0.030 (0.034)	-0.123 (0.102)	-0.030 (0.122)

Time spent with Patient								
Very dissatisfied	-0.029** (0.012)	-0.023 (0.015)	-0.002 (0.006)	-0.005 (0.008)	-0.0167 (0.013)	-0.017 (-0.017)	-0.025 (0.039)	-0.088 (0.062)
Dissatisfied	-0.031*** (0.017)	-0.035 (0.020)	-0.006 (0.016)	-0.011 (0.019)	-0.0192 (0.015)	-0.011 (0.019)	-0.037 (0.055)	0.031 (0.020)
Satisfied	0.016** (0.008)	0.012 (0.009)	0.003 (0.007)	0.006 (0.009)	0.017 (0.013)	-0.117*** (0.013)	0.009 (0.018)	-0.098** (0.048)
Very satisfied	0.053*** (0.023)	0.042*** (0.017)	0.006 (0.015)	0.011 (0.018)	0.019 (0.015)	-0.113*** (0.011)	0.054 (0.080)	0.154 (0.087)
Cleanliness of HF								
Very dissatisfied	-0.011 (0.008)	-0.003 (0.009)	0.007 [†] (0.004)	0.004 (0.005)	0.0160** (0.008)	0.015 (0.010)	-0.080 (0.051)	-0.071 (0.052)
Dissatisfied	-0.028 (0.019)	-0.008 (0.023)	0.027* (0.016)	0.017 (0.020)	0.038** (0.019)	0.035 (0.024)	-0.155*** (0.056)	.246** (0.092)
Satisfied	0.777 (0.069)	-0.00009 (0.0007)	-0.005 (0.003)	-0.003 (0.004)	-0.018 [†] (0.009)	-0.036** (0.001)	-0.010 (0.059)	-0.021 (0.059)
Very satisfied	0.031** (0.011)	0.010 (0.033)	-0.028 [†] (0.017)	-0.018 (0.021)	-0.037** (0.018)	-0.033** (0.013)	0.246*** (0.075)	-0.145** (0.059)

All regressions include the following observables: women's age, educational background, household size, and wealth quantile. The dependent variables are responses to endline survey questions elicited after the incentive scheme. Standard errors are reported in parentheses clustered at the village level. OP and IV-OP denote 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 comprises 8,174 women.

[†] p < .1, *p < .05, **p < .01, and ***p < .001 show the significance levels.

Heterogenous effects by the type of financing mechanism

In addition to the effects of the incentive scheme across HFs, this study hypothesized whether the incentive effects varied based on the types of financing modalities. Tables 6 and 7 report the estimated effects of the conditional incentive on women's and children's health outcomes, respectively, across contracted-out (NGO-regulated) and contracted-in (government-regulated) HFs.

Table 6 Effects of the 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 (1.145)	1.798 (1.552)	0.884* (0.519)	0.517*** (0.026)
Postnatal care	-5.799 [†] (3.282)	-1.964 (4.403)	2.91*** (0.994)	2.763** (1.182)

Skilled birth attendance	-0.161 (0.153)	0.067 (0.118)	0.517 (0.466)	1.806** (0.473)
Birth in health facility	0.120 (0.13)	0.139 (0.134)	0.129 (0.453)	0.760* (0.456)
Contraceptive use	-0.125** (0.055)	-0.083 (0.078)	-0.060** (0.021)	-0.059† (0.035)

All regressions include the following observables: women's age and educational background, household size, and wealth quantile. The dependent variables are responses to endline survey questions elicited after the incentive scheme. Standard errors are reported in parentheses clustered at the village level. OLS and IV denote ordinary least square and instrumental variables models, respectively. The sample comprises 8,174 women of child-bearing age.

† p < .1, *p < .05, **p < .01, and ***p < .001 show the significance levels.

Based on the estimated coefficient of the IV models in column (2) of Table 6, the effects of the incentive on maternal health outcomes at the government-regulated HF was indistinguishable from zero at all levels of significance, although the estimates of the OLS models in the column (1) show the negative effects of the incentive on the use of postnatal services and contraceptives. Except for contraceptive use, the incentive program was positively associated with an increase in the use of maternal health outcomes at NGO-regulated HF.

Table 7 Effects of the incentive on children's 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† (0.052)	0.102 (0.080)	-0.051 (0.056)	0.047† (0.028)
BCG vaccine	-0.014 (0.017)	-0.032* (0.011)	0.046** (0.001)	0.051** (0.010)
OPV vaccine	-0.118 (0.109)	-0.131 (0.121)	-0.084 (0.103)	-0.036 (0.112)
Penta vaccine	0.296 (0.383)	0.509 (0.449)	0.027 (0.108)	0.043 (0.119)
Measles vaccine	-0.025 (0.024)	-0.031 (0.029)	0.026† (0.013)	0.034* (0.015)
Use of Vitamin A	0.001 (0.023)	0.017 (0.027)	0.011 (0.011)	0.025* (0.012)

All regressions include the following observables: mother's age, educational background, household size, and wealth quantile and infant's gender. The dependent variables are responses to endline survey questions elicited after the incentive scheme. Standard errors are reported in parentheses clustered at the village level. OLS and IV denote ordinary least square and instrumental variables models, respectively. The sample comprises 7, 806 infants under five years old.

† p < .1, *p < .05, **p < .01, and ***p < .001 show the significance levels.

Coinciding with the incentive effects on maternal health outcomes, Table 7 shows that the incentive program had no impactful results on children's health outcomes across government-

regulated HFs. Even the incentive program lowered the probability of receiving BCG vaccines at these facilities. By contrast, the utilization of children's health care services was significantly improved among service users at NGO-affiliated HFs. Administrative complexities, a higher marginal cost of production, and weak institutional and organizational capacities at contracted-in HFs might have eliminated any positive, meaningful impact of the incentive programs (Milliman and Prince 1992; Sherry et al. 2017a).

Incentive and Health Facility Efficiency Analysis

Table 8 provides the estimated input-oriented technical efficiency scores of the comparison and treatment HFs at the baseline and endline surveys. At the baseline survey, there was no statistically significant difference among the efficiency scores of comparison and treatment SCs, BHCs, and DHs. However, on average, the treatment CHCs were 4.9% less technically efficient before the implementation of the incentive. By contrast, the technical efficiency scores of treatment HFs improved at the endline survey. Human resource constraints in running efficient HFs has been a debated topic of interest among researchers to understand rational combinations of medical resources (Banerjee et al. 2004; Rosenzweig and Wolpin 1986).

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

Type of health facility	Baseline Survey		Difference	Endline Survey		Difference
	Comparison	Treatment		Comparison	Treatment	
Subcenter (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 (DHs)	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 are clustered at the village level. The distribution of test statistics is bootstrapped using the wild-cluster bootstrap technique (Cameron et al. 2008). Input-oriented technical efficiency scores are estimated using the data envelopment approach (Tone 2001).

[†] $p < .1$, * $p < .05$, ** $p < .01$, and *** $p < .001$ show the significance levels.

Discussions and Conclusions

To align the socially desirable goals of health service delivery and the behaviors of users and providers, developing countries have implemented both demand- and supply-side performance-based financing initiatives. Application of these initiatives have provoked debates 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 burgeoning body of empirical literature, the identification strategy of the this work is based upon the following: Empirical analysis of field experiments might provide biased results if the compliance assumption of the experiment setting is violated (J. D. Angrist and Krueger 2001; J. D. Angrist and Pischke 2008). The bias-corrected estimates of this study show that the supply-side conditional incentive is associated with an increase in the utilization of health care services among women and children, particularly at the smaller HFs. Investigating facility-level effects is essential to having a full understanding of the impact of any incentive policy that aims to improve the production of health care services (Miller and Babiarz 2013). HFs as the building blocks of health systems play a significant role in improving the quantity and quality of services (Mwabu 2007).

The effects of the incentive scheme on the use of maternal and child health outcomes, concordant with much of the literature, are modestly positive among users at smaller facilities but strikingly negative at larger HFs. 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 et al. 2011). The indirect psychological effect is associated with intrinsic and extrinsic motivations. The incentive for health care providers is an extrinsic motivator. Psychologically, intrinsic motivation always pushes an individual's behavior in the same direction as extrinsic motivation (Deci et al.

1999). A key explanation for the contrasting effects of the incentive at small and large facilities is the divergent directions of these two effects. At smaller health centers, the number of providers is limited, and the predetermined delivered services are comparable to those of larger facilities. The effects of income associated with the incentive are sufficiently sizable to generate a convergent price and psychological effects. By contrast, at the larger HFs, although rewards could produce a trivial income, their psychological effect results in a minimum increase in the provider's motivation. The theoretical basis for predicting differences between small and large facilities in the level and manner in which they incentivize workers is weak. Turning first to the total amount of incentive pay, arguments can be gathered to determine that small facility workers will earn as much as, less than, or more than their large facility counterparts. Parity in incomes for workers of equal productivity would result if job markets were competitive and HF workers were indifferent to working at small or large HF. Lower earnings might result if some service providers were willing to supply their labor to small facilities at a lower price, perhaps because they prefer to associate with or influence the behavior of a small facility that was engaged in particular set of service provisions. The latter idea could be explained under the framework of the "labor donation" theory of wages (Handy and Katz 1998).

Relatedly, the differential effects of the incentive across the HFs can also be attributed to the patients' health status and the facility-related binding constraints. In general, patients receiving care at smaller facilities are probably 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 an outpatient service. Therefore, incentivized schemes generate effective results at 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 on the outcomes might be limited.

The findings of this study have important public policy implications for both the design and underlying economic theory of incentive schemes. First, the supply-side pay-for-performance mechanism improves the use of health care services by families in the lower levels of the 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. Last, the smaller HFs surrounded by a relatively smaller number of households are more efficient in terms of their resource allocations than the larger facilities, and this could determine an important policy direction in the redistribution of health care resources in resource-constrained contexts.

ACKNOWLEDGMENTS

An earlier version of this manuscript was competitively selected for presentation at the 2020 American Economic Association (AEA) annual meetings in San Diego, CA. Also, a preliminary draft of this work was first presented at the 2018 Missouri Valley Economic Association (MVEA) annual meetings in Memphis, TN. I further thank the session attendees for their constructive comments. I would like to express my special thanks of gratitude to my advisor Dr. Albert Okunade, who provided me the excellent guidance throughout this research project. I thank the anonymous reviewers whose comments helped improve and clarify this manuscript. Finally, I thank the Health Economics and Health Care Financing Directorate (HEFD) of Afghanistan's Ministry of Public Health for providing the datasets for this study. The standard disclaimer applies.

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, doi:10.3368/jhr.XL.4.791.
- Angrist, J. (1995). Conditioning on the probability of selection to control selection bias. 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, doi:10.1257/jep.15.4.69.
- 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, doi:10.1257/app.20130126.
- 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, doi:10.1016/j.jhealeco.2016.08.006.
- Balke, A., & Pearl, J. (1997). Bounds on treatment effects from studies with imperfect compliance. *Journal of the American Statistical Association*, 92(439), 1171-1176, doi:10.1080/01621459.1997.10474074.
- Banerjee, A., Deaton, A., & Duflo, E. (2004). HEALTH, HEALTH CARE, AND ECONOMIC DEVELOPMENT: Wealth, Health, and Health Services in Rural Rajasthan. *American Economic Review*, 94(2), 326-330, doi:10.1257/0002828041301902.
- 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. *Lancet*, 377(9775), 1421-1428, doi:10.1016/S0140-6736(11)60177-3.
- 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, doi:10.1257/app.5.3.27.
- 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.
- 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, doi:10.1093/wber/lhj002.
- De Walque, D., Gertler, P. J., Bautista-Arredondo, S., Kwan, A., Vermeersch, C., de Dieu Bizimana, J., et al. (2015). Using provider performance incentives to increase HIV testing and counseling services in Rwanda. *Journal of Health Economics*, 40, 1-9.

- Deci, E. L., Koestner, R., & Ryan, R. M. (1999). A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological bulletin*, 125(6), 627.
- 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, doi:10.1257/jep.7.4.135.
- Garen, J. E. (1994). Executive compensation and principal-agent theory. *Journal of Political Economy*, 102(6), 1175-1199, doi:10.1086/261967.
- Gaynor, M., Rebitzer, J. B., & Taylor, L. J. (2004). Physician incentives in health maintenance organizations. *Journal of Political Economy*, 112(4), 915-931, doi:10.1086/421172.
- Gertler, P. (2004). Do conditional cash transfers improve child health? Evidence from PROGRESA's control randomized experiment. *American Economic Review*, 94(2), 336-341, doi:10.1257/0002828041302109.
- 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, doi:10.1257/jep.25.4.191.
- 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, doi:10.1002/hec.3136.
- Handy, F., & Katz, E. (1998). The wage differential between nonprofit institutions and corporations: Getting more by paying less? *Journal of comparative economics*, 26(2), 246-261, doi:10.1006/jcec.1998.1520.
- Heckman, J. J. (1995). Randomization as an instrumental variable. 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, doi:10.1111/j.1468-0262.2007.00801.x.
- 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, doi:10.1093/qje/qjz010.
- McGuire, T. G., & Pauly, M. V. (1991). Physician response to fee changes with multiple payers. *Journal of Health Economics*, 10(4), 385-410, doi:10.1016/0167-6296(91)90022-F.
- Miller, G., & Babiarz, K. S. (2013). Pay-for-performance incentives in low-and middle-income country health programs. National Bureau of Economic Research.
- 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, doi:10.1016/0095-0696(92)90035-U.

- Moffitt, R. A. (1996). Identification of causal effects using instrumental variables: Comment. *Journal of the American Statistical Association*, 91(434), 462-465, doi:10.2307/2291630.
- Mwabu, G. (2007). Health Economics for Low-income Countries. *Handbook of Development Economics*, 4, 3305-3374, doi:10.1016/S1573-4471(07)04053-3.
- NHA (2014). Afghanistan National Health Accounts with Disease Accounts. Kabul, Afghanistan.
- NHS (2016). National Health Strategy 2016–2020. In MoPH (Ed.). Kabul, Afghanistan.
- Peabody, J., Shimkhada, R., Quimbo, S., Florentino, J., Bacate, M., McCulloch, C. E., et al. (2011). Financial incentives and measurement improved physicians' quality of care in the Philippines. *Health Affairs*, 30(4), 773-781, doi:10.1377/hlthaff.2009.0782.
- 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, doi:10.1086/683989.
- 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, doi:10.1016/j.jhealeco.2015.07.001.
- 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, doi:10.1093/heapol/czw045.
- Roodman, D. (2011). Fitting fully observed recursive mixed-process models with cmp. *The Stata Journal*, 11(2), 159-206, doi:10.1177/1536867X1101100202.
- 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, doi:10.1162/rest.1997.79.2.348
- 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*, 3(2), 192-226.
- 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*, doi:10.1162/AJHE_a_00072.
- Thornton, J., & Eakin, B. K. (1997). The utility-maximizing self-employed physician. *Journal of Human Resources*, 98-128, doi:10.2307/146242.
- 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. *Lancet Glob Health*, 5(8), e828-e837, doi:10.1016/S2214-109X(17)30251-6.
- 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, doi:10.1002/hec.3219.
- Vujicic, M. (2009). How you pay health workers matters: A primer on health worker remuneration methods. *World Bank, Washington DC*.

