

IMPACT OF PERFORMANCE-BASED FINANCING IN A LOW-RESOURCE SETTING: A DECADE OF EXPERIENCE IN CAMBODIA

ELLEN VAN DE POEL^{a,*}, GABRIELA FLORES^b, POR IR^{c,d} and OWEN O'DONNELL^{e,f,g}

^a*Institute of Health Policy and Management, Erasmus University Rotterdam, Rotterdam, The Netherlands*

^b*Institute of Health Economics and Management, University of Lausanne, Lausanne, Switzerland*

^c*National Institute of Public Health, Phnom Penh, Cambodia*

^d*Institute of Tropical Medicine, Antwerp, Belgium*

^e*Erasmus School of Economics, Erasmus University Rotterdam, Rotterdam, The Netherlands*

^f*Tinbergen Institute, Amsterdam, The Netherlands*

^g*University of Macedonia, Thessaloniki, Greece*

ABSTRACT

This paper exploits the geographic expansion of performance-based financing (PBF) in Cambodia over a decade to estimate its effect on the utilization of maternal and child health services. PBF is estimated to raise the proportion of births occurring in incentivized public health facilities by 7.5 percentage points (25%). A substantial part of this effect arises from switching the location of institutional births from private to public facilities; there is no significant impact on deliveries supervised by a skilled birth attendant, nor is there any significant effect on neonatal mortality, antenatal care and vaccination rates. The impact on births in public facilities is much greater if PBF is accompanied by maternity vouchers that cover user fees, but there is no significant effect among the poorest women. Heterogeneous effects across schemes differing in design suggest that maintaining management authority within a health district while giving explicit service targets to facilities is more effective in raising utilization than contracting management to a non-governmental organization while denying it full autonomy and leaving financial penalties vague. Copyright © 2015 John Wiley & Sons, Ltd.

Received 1 October 2014; Revised 24 June 2015; Accepted 29 June 2015

KEY WORDS: health financing; performance-based financing; maternity care; vaccination; Cambodia

1. INTRODUCTION

Performance-based financing (PBF) is increasingly deployed in low-income and middle-income countries with the aim of raising the quantity and quality of health care delivered. As the name suggests, PBF involves paying providers, at least partially, on the basis of what they deliver, rather than allocating budgets in relation to historic costs or inputs. It offers incentives intended to redress the underperformance, notably high worker absenteeism, frequently observed in poorly funded public health systems that lack accountability (Meeseen *et al.*, 2011; Gertler and Vermeersch, 2012). It is spreading rapidly, particularly in sub-Saharan Africa, where more than 30 countries have linked or are linking the payment of healthcare providers to their performance, mostly in relation to delivery of maternal and child health care (Fritsche *et al.*, 2014).

While there is considerable enthusiasm for the policy, particularly among donor agencies, there is relatively little evidence on its effectiveness in low-resource settings (Soeters, 2011; Witter *et al.*, 2012). Evidence is particularly lacking on the effects of alternative designs of PBF contracting arrangements and

*Correspondence to: Institute of Health Policy and Management, Erasmus University Rotterdam, PO Box 1738, 3000 DR Rotterdam, The Netherlands. E-mail: vandepoel@bmg.eur.nl

incentive structures, and on its distributional effects. Given the complexity and scale of PBF interventions, this evidence is difficult to obtain through experimental studies. The largest and most carefully designed randomized experiment of PBF in a low-resource setting, which was carried out in Rwanda, indicated that PBF increased the number of institutional deliveries by 24%, but had no effect on antenatal care and immunization rates (Basinga *et al.*, 2011). Evaluation of a randomized pilot in Cambodia also found significant and large effects on some, but not all, incentivized outcomes (Bhushan *et al.*, 2007). However, the internal validity of both of these experiments was somewhat compromised by the initial randomized allocation of districts to treatment and control groups not being fully maintained. External validity can be threatened by the evaluation adding an additional layer of performance monitoring that may influence providers' behaviour. Experiments cannot provide evidence of the effect of PBF when it is extended beyond selected sites without adherence to study protocols. This paper adds to the evidence base on the impact of PBF in low-resource settings (Witter *et al.*, 2012) by evaluating the impact of operational, as opposed to experimental, programmes implemented over a decade across Cambodia.

Cambodia was the first low-income country to experiment with PBF of public health care. Since 1999, a variety of programmes have contracted the management of district health authorities to non-governmental organizations (NGOs), made the funding of districts and facilities contingent on performance targets or directly linked revenues to services delivered. These arrangements are intended to increase healthcare provision, particularly of maternal and child health services, either by incentivizing health authority managers to exert effort in ensuring that care is delivered, for example by monitoring that facilities are open, staffed and stocked, or by giving financial incentives directly to providers to deliver care. There is variation in the strength of the incentives offered. Some districts are given a performance target with little credible threat of punishment for failure to meet it. The impact of such a scheme is expected to be weaker than another that pays providers for each service delivered. We exploit variation in the timing of the introduction across districts of PBF schemes that differ in the nature of the contracting, the placement of management authority and the explicitness with which finance is tied to performance to identify and compare their effects on the delivery of maternal and child health care.

There have been substantial improvements in indicators of population health and health service utilization in Cambodia since 2000 (National Institute of Statistics, Directorate General for Health, ICF Macro, 2011). PBF may have contributed to these gains, but so too could a multitude of economic, social and health system developments. We identify the effect of PBF by comparing changes in rates of utilization of maternal and child health care in districts in which the financing instrument is introduced with changes occurring in districts that remain without it. We control for the introduction of demand-side schemes and test robustness both to restricting the control districts to those most similar to the treated with respect to pre-treatment trends in the outcomes and to weighting the controls according to their degree of similarity. For the initial phase of PBF implementation, we are able to conduct the difference-in-differences analysis on treatment and control districts that were randomly assigned in a pilot experiment.

By increasing the reliance of facilities on revenue derived from user fees, it is feared that PBF may weaken the incentives for poorer patients to seek care and exacerbate inequity (Bonfrer, Soeters *et al.*, 2014). We examine the legitimacy of this concern not only by allowing for heterogeneous effects on utilization by poverty but also by testing whether the effect is greater when PBF operates alongside vouchers that eliminate user fees for the incentivized maternal health services.

We estimate that any type of PBF introduced between 1999 and 2010 raised the proportion of births occurring in a public facility by 7.5 percentage points (pp) (25%). For the other two incentivized outcomes examined – antenatal care and infant vaccinations – the point estimates of the effects are positive but not significant. Consistent with the effectiveness of PBF being constrained by the capacity of patients to pay for incentivized services, the effect on public facility births is smaller for the poorest women and is greatly reduced when PBF is not accompanied by vouchers that cover user fees.

The impact on the rate of institutional deliveries is considerably smaller than that on births in public clinics, indicating substantial switching from private to the incentivized public facilities with no significant impact on

deliveries supervised by a skilled birth attendant. Possibly because of this, but perhaps also because of a lack of equipment and personnel that can save life in the case of a complicated birth, we find no significant effect on neonatal mortality. This is consistent with evidence that PBF in Rwanda raised institutional deliveries, but this did not help reduce neonatal mortality (Chari and Okeke, 2014).

Comparison of the effects of different PBF schemes on the rate of public facility births suggests that contracting the management of a health district to an NGO works better when the latter is given more autonomy to manage. Maintaining management within the public sector while specifying service targets for facilities is found to be more effective than contracting management to an NGO, but giving it limited managerial autonomy and only vague financial incentives.

2. PERFORMANCE-BASED FINANCING IN CAMBODIA

The trauma and devastation wrought by the Khmer Rouge and years of war left Cambodia with untold need for medical care and no healthcare system to speak of (Fujita *et al.*, 2013). The scale of this imbalance created the impetus for contracting the provision of district health services to NGOs. Different forms of contracting at the Operational District (OD) level have been adopted over the years (Table I), but they share a common characteristic of making funds received contingent, with varying degrees of clarity, on performance in the delivery of defined services. Over time, reliance on NGOs to manage the delivery of health care in Cambodia has been reduced, as the capacity of the health system has strengthened.

Table I. Description of PBF schemes in the 1999–2010 period

Scheme	Acronym	Period	# ODs to 2010	OD management responsibility	OD incentives	Facility and staff incentives
PBF pilot	Pilot-OUT Pilot-IN	1999–2003	2	NGO	Fixed price contract with progress payments contingent on performance against service targets	Level and nature of incentive payments at discretion of NGO contractor
			3	NGO within MoH employment and procurement rules		
PBF extension	IN	2004–2008	11	As for Pilot-IN	As for Pilot	NGO contractor obliged to implement performance management but given discretion in level and nature of incentive payments
	Internal	2005–2010	8	MoH with NGO advisors	Performance contract with MoH specifying targets but w/o incentive payments	Bonuses for reaching targets and abstention from illegitimate practices, plus fee-per-case
	GAVI-HSS	2007–present	10	MoH (PBF implemented directly at facility not through OD)	None	Payment per unit of service delivered
PBF consolidation	SOA	2009–present	22	SOA - semi-autonomous within MoH	Performance contract with MoH with incentive payments	Facility enters performance contract with SOA with incentive payments

OD, Operational District; NGO, non-governmental organization; MoH, Ministry of Health; OUT, contracting out; IN, contracting in; SOA, Special Operating Agencies; GAVI-HSS, Global Alliance for Vaccination and Immunization Health Systems Strengthening; PBF, performance-based financing.

2.1. Performance-based financing pilot, 1999–2003

Between 1999 and 2003, there was a pilot experiment of contracting the management of health districts to NGOs.¹ Eight out of 12 potentially eligible ODs in three provinces were randomly selected for implementation of contracting, and the remaining four were used as controls. Since the tenders received in three of the treatment ODs were considered technically incomplete or too expensive, contracting was introduced in only five ODs (Bhushan *et al.*, 2007). In each OD, an international NGO was contracted to manage the delivery of basic preventive care through health centres and of basic curative care at the referral hospital. Two versions of contracting that differed in the degree of management responsibility acceded to NGOs were tried. One model, which we will label *Pilot-OUT*, implemented in two ODs, gave the NGO full autonomy to hire and fire staff on conditions of employment defined by the contractor and to procure medicines and supplies. Most staff were Ministry of Health employees who could be assigned to another OD if their conduct was considered unsatisfactory. The less radical scheme, which we will refer to as *Pilot-IN*, did not grant the NGO the right to hire and fire, and required that it operate within Ministry rules on terms of employment and use standard procurement channels.

In both varieties of contracting, the contract between the Ministry of Health and the NGO stated a fixed price, with progress payments being contingent on performance towards the realization of targets specified for delivery of mainly maternal and child services. In the *Pilot-OUT* model, the contractor was free to set incentive payments for facilities and staff as it liked. Under *Pilot-IN*, employees were supposed to be paid a standard Ministry salary. Because of staff resistance, this proved impossible to implement. All three *Pilot-IN* ODs used supplementary funding to make additional payments to staff.

An unpublished evaluation of the pilot finds a positive point estimate of the effect of any form of contracting (*IN* or *OUT*) for all but one of the targeted outcomes and a significant (5%) average effect size (Bhushan *et al.*, 2007). The effect size is slightly larger for the *Pilot-OUT* model, but the null of equal effects is not rejected. The improvements in service delivery cannot be attributed solely to response to PBF incentives because treatment and control ODs differed not only in the structure of financing but also in the levels. The *Pilot-OUT* ODs received 142–178% more funding per head than the control ODs (Loevinsohn and Harding, 2005; Bhushan *et al.*, 2007). The *Pilot-IN* ODs received 52–90% more.

2.2. Performance-based financing extension, 2004–2008

Despite initial claims that the *Pilot-OUT* model was more effective (Keller and Schwartz, 2001; Soeters and Griffiths, 2003; Loevinsohn and Harding, 2005), the greater burden it imposed on the public purse, plus an unwillingness on the part of the Ministry of Health to devolve service delivery entirely to international NGOs, resulted in the adoption of a version closer to the *Pilot-IN* model in four of the five treatment ODs included in the pilot, plus seven others. In this extended programme, which ran from 2004 to 2008 and we will refer to as the *IN* model, OD staff were seconded to the NGO contractor but remained civil servants. They could not be fired, but if performing poorly could be referred to the provincial health department, although this agency had no financial incentive to penalize recalcitrants (Ministry of Health Cambodia, 2006; Braye, 2007). Although subject to constraints of working within the Ministry of Health administration and procurement rules, the NGO contractor was obliged to implement performance management and could make incentive payments to facilities and/or staff at its own discretion.² It was intended that the share of staff income derived from user fees would increase over

¹Contracting was initiated by the Swiss Red Cross at a single provincial hospital in 1997, leading to a scheme referred to as the New Deal in two rural ODs. Because this scheme did not link payments to specific services and only a few observations in our data were exposed to it, we do not include it in the definition of PBF.

²There was a great deal of heterogeneity across ODs in the incentive structures implemented. Incentive payments were made at three levels within each OD – the OD administration office, the referral hospital and health centres. Each contract specified the maximum incentive payment at the respective level, the scoring method to ascertain attainment of targets and the proportion of user fee revenue to be paid to staff (Braye, 2007).

time in order that incentive payments could be gradually withdrawn. Given this plan, there was a policy to encourage the location of funded user fee waiver schemes – so-called Health Equity Funds (Flores *et al.*, 2013) – in the contracting ODs in order to protect the poor from the burden of user fees and to provide a source of revenue. As with the pilot, the contracting ODs were more generously financed.³

From 2005 up to 2009, bilateral aid from the Belgian Technical Cooperation supported another form of PBF in eight ODs. Unlike the *IN* model, this scheme did not devolve management of the OD to an NGO. The province health department, OD administration office, referral hospitals and health centres were all managed through performance contracts within the Ministry of Health. These contracts specified targets at all levels, but offered incentive payments only at the facility level. We refer to this as *internal contracting*.

Besides the location of management responsibility, an important difference between the two PBF models in this period is that *internal contracting* did not promise a capitation payment contingent on performance targets being met. Rather, facilities were offered bonuses for reaching targets, were paid a fee per case for certain procedures and were paid a further bonus if illegitimate practices, such as staff taking informal payments or not being present at facilities, were avoided (Annear *et al.*, 2008). The incentives were thus structured in the form of additional payments that could be earned, as opposed to the *IN* model's threat not to pay a promised amount.

Since 2007, ten ODs have received funding from the Global Alliance for Vaccines and Immunization Health System Strengthening (*GAVI-HSS*) programme. This includes incentive payments for the delivery of five maternal and child preventive care services, including vaccinations and antenatal care (Pearson and Pun, 2009). Unlike the contracting approaches described earlier, this scheme offers fee-for-service payments to facilities, rather than finance linked in some way to performance targets. In terms of per capita funding, it is much less substantial than the other schemes.

At the end of 2007, the government introduced the nationwide Midwife Incentive Scheme that pays midwives \$10 (15) for each live birth they attended at a referral hospital (health centre), on top of any fee charged to the patient (Liljestrand and Sambath, 2012).

2.3. Performance-based financing consolidation, since 2009

The *internal contracting* model has proven to be the forerunner of the entirely publicly managed system that increasingly operates through the use of performance-based contracts to pass down responsibilities and incentives for service delivery to lower levels of the health system. The 22 ODs that have become Special Operating Agencies since 2009 have semi-autonomous status within the public sector. They are given discretion to manage resources as considered best towards the realization of performance targets agreed with a higher level within the public administration. We do not evaluate the effects of this specific form of internal contracting separately because of the very limited exposure to it within the estimation period (up to 2010).

The main characteristics of the various schemes are summarized in Table I. The spread of performance-based financing in Cambodia across ODs and over time is summarized in Table AI in Supporting Information. Of the 77 ODs in the country, 45 had no experience of PBF by 2010. There is clear path dependence among the 32 ODs exposed to PBF; once some form of PBF is introduced, it is generally followed by another.

3. DATA

We use data from the Cambodian Demographic and Health Survey (CDHS) fielded in 2000, 2005 and 2010. Each of the three cross-section surveys interviews a nationally representative sample of women of reproductive

³Per capita, public health expenditure in the 11 contracting ODs averaged \$7.30 compared with a national average of \$4.11 (Braye, 2007). Administration costs were high. On average, they were estimated to be one-third of total spending in the contracting districts (Ministry of Health Cambodia, 2006).

age. Women are asked about their pregnancies and birth-related healthcare use in the 5 years preceding each survey. Combining these data gives us information on maternal and child healthcare use for 80 004 births in the 1995–2010 period.

The main performance-based financing programmes implemented in Cambodia since 1999 have consistently specified performance targets for indicators of the following: (i) child vaccination; (ii) antenatal care (at least two visits); (iii) delivery in a public facility; and (iv) birth-spacing use.⁴ While all four indicators are consistently measured in all three of the cross-section surveys, utilization of birth-spacing methods can only be identified at the time of interview, and not with respect to the starting date of any PBF scheme operating in the OD in which the women are located. This indicator is therefore not included in the analyses. The *GAVI-HSS* scheme makes payments for vaccinations and antenatal care, but not for deliveries in facilities.

Full vaccination is defined by an indicator that equals one if a child, aged between 12 and 24 months at the time of the survey, is fully vaccinated (BCG, OPV3, DPV3 and measles). The indicator for antenatal care is equal to one if the mother had at least two antenatal visits during the most recent pregnancy. The third outcome indicates whether each child the woman has given birth to in the previous 5 years was born in a public health facility. To check whether there is any crowd-out of births in private facilities, we also examine the impact on births in any facility.

We identify the effects by examining service utilization in relation to whether there was a PBF scheme in operation in the OD in which the mother or child was located at the time of pregnancy or birth, depending on the specific outcome. The OD in which each mother/child is located is not coded in the data. We identify it from the Global Positioning System codes that are given for the sample clusters. These codes are purposefully slightly offset to protect confidentiality. This makes it impossible to match 231 communities mainly in the 2005 and 2010 surveys, which represent 14.1% of the total number of communities surveyed over the 3 years. In 2005 and 2010, these clusters are mainly located in the region of the capital city, Phnom Penh, which causes the sample of unmatched births (13% of the total⁵) to be somewhat better-off and more urban than the estimation sample. No other important differences in covariates exist between the matched and unmatched sample.

Figure 1 shows steep upward trends in the rates of utilization of all three incentivized services both in the ODs in which some form of PBF was introduced by 2010 and in those that remained without any PBF. The overall increase in the use of maternal and child health care reflects the very low baseline and is likely the result of strong economic growth, strengthening of the public health system and interventions aimed at removing user fees for the poorest. Prior to the introduction of the first PBF schemes in 1999, utilization rates are approximately flat for all three indicators, and there is no obvious difference between the two groups of ODs. After 2000, rates begin to increase rapidly. For the vaccination rate and antenatal care, there are no obvious differential trends between the ODs that obtain PBF sometime in the period and those that do not. From 2004, when the largest expansion of PBF schemes took place, the rate of delivery in public facilities appears to rise more rapidly in the ODs that experience PBF. Separate figures for the PBF pilot and extension phases suggest that especially the PBF pilot schemes may have been effective in increasing the rate of institutional deliveries (see Supporting Information).

4. EMPIRICAL STRATEGY

4.1. Effect of any form of performance-based financing

We identify the effect of PBF by comparing changes in utilization rates when PBF is implemented in a district with changes occurring in the same time period in districts that remain without PBF. Identification relies on the common trends assumption that if PBF had not been introduced in what will be

⁴Note that most schemes had a much longer list of incentivized outcomes. These are the ones common to all schemes.

⁵The unmatched births represent 1% of the CDHS 2000, 25% of CDHS 2005 and 13% of CDHS 2010 sample.

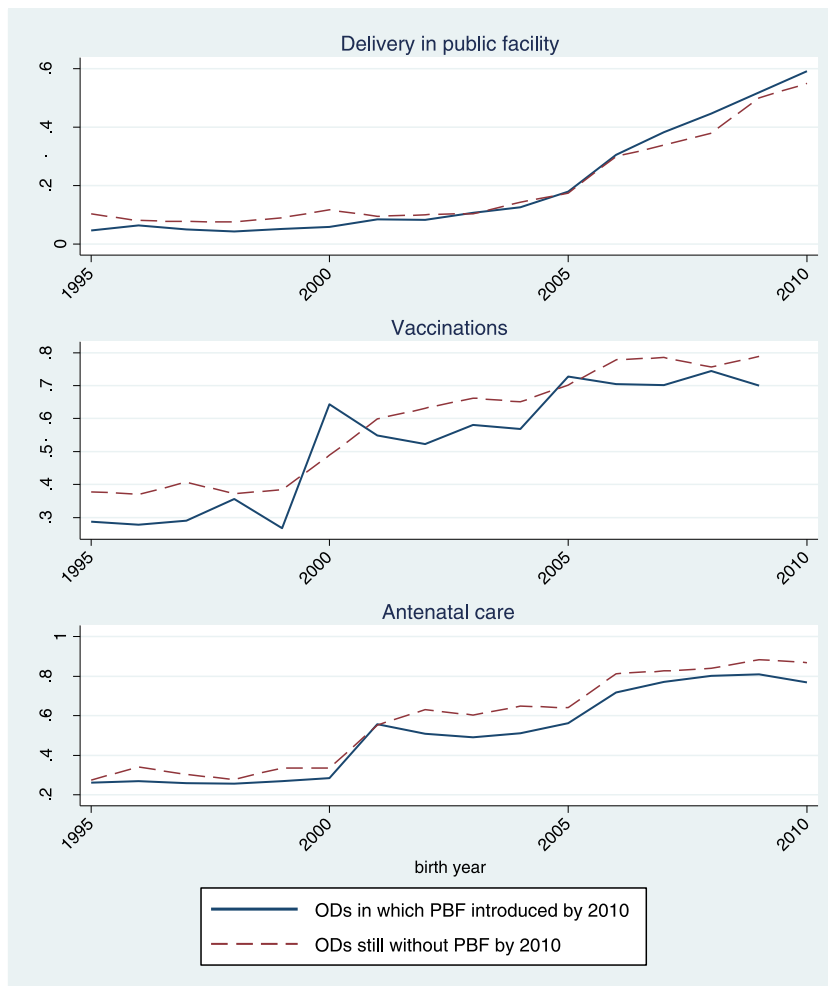


Figure 1. Trends in proportions utilizing incentivized services, 1996–2010

referred to as treated ODs, then the outcome would have changed in those ODs in the same way as it is observed to have changed in the ODs that remain without PBF. Given the non-random allocation of PBF schemes across ODs, it is important to weaken this assumption by controlling for time-varying observable factors that may have generated differential trends across the two groups of ODs. In particular, the policy of favouring the location of fee waiver schemes in ODs with contracting could result in the PBF effect being confounded by that of the fee waiver if there was no control for the latter.⁶ Maternity care vouchers that can be used to obtain antenatal care and to deliver in a public facility without incurring user fees have been shown to impact positively on utilization (Van de Poel *et al.*, 2014).

⁶Donor sponsored Health Equity Funds are disproportionately concentrated in ODs with PBF. But the government-funded fee waiver scheme (SUBO) is almost entirely located in ODs without PBF. As a result, ODs with and without PBF differ little in their exposure to some form of fee waiver scheme. Of the 32 ODs in which PBF is introduced, 25 have a Health Equity Fund at some time and 1 has the government scheme. Of the 45 ODs without PBF, 17 have an equity fund and 10 have the government scheme.

Table II. Means of covariates at baseline and test of equality of change in mean across treated and control ODs

	Baseline: 1998			Change:
	ODs in which PBF introduced by 2010 (Treated)	ODs remaining without PBF (Control)	Test: treated = control <i>p</i> -value	1998–2010 Test: Δ treated = Δ control <i>p</i> -value
birth interval < 24 months	0.194	0.189	0.796	0.294
First born	0.184	0.169	0.456	0.186
> = 4th born	0.333	0.345	0.604	0.067
Mother's age at birth < 20 years	0.115	0.083	0.033	0.285
20 ≤ mother's age at birth < 35	0.677	0.712	0.146	0.741
Mother's age at marriage	19.005	19.414	0.032	0.535
Age of the household head	37.903	38.044	0.813	0.090
Household head is male	0.877	0.840	0.035	0.556
Mother has primary education	0.470	0.539	0.006	0.471
Mother has secondary education	0.097	0.124	0.090	0.299
Wealth index				
< 20th quantile	0.336	0.279	0.021	0.175
20–39th quantile	0.225	0.279	0.021	0.026
40–59th quantile	0.189	0.168	0.300	0.001
60–79th quantile	0.148	0.158	0.595	0.314
urban	0.162	0.128	0.059	0.568

PBF, performance-based financing; OD, Operational Districts.

All covariates are dummy variables apart from *mother's age at marriage* and *age of head of household*. First two columns give the means for the respective groups in 1998 before the introduction of any PBF. Third column gives the *p*-value from a *t*-test that the mean of each covariate is equal across the Treatment and Control ODs in 1998. Fourth column gives the *p*-value from a *t*-test of the null that the change in the mean between 1998 and 2010 is the same across treated and control ODs. This involved restricting the data to births in 1998 and 2010 and regressing each covariate on a treatment indicator, a year indicator and an interaction between the two, and testing whether the coefficient on the latter is zero.

Potentially, they relax the demand-side constraint on the response to the provider incentives offered by PBF.

We estimate the following difference-in-differences linear probability model for each outcome:

$$y_{idt} = \beta PBF_{dt} + \mathbf{X}_{idt}\boldsymbol{\Omega} + \lambda_1 Voucher_{dt} + \lambda_2 HEF_{dt} + \lambda_3 SUBO_{dt} + \mathbf{OD}_d + \tau_t + \varepsilon_{idt} \quad (1)$$

where y_{idt} is an indicator of whether a service (in turn, vaccination, antenatal care or delivery in a public facility) is utilized for the child/birth/pregnancy i in OD d and month t .⁷ PBF_{dt} is 1 if a PBF scheme that ties financial incentives to provision of the particular service is in operation in OD d at time t . For antenatal care, this indicator is *turned on* if PBF was in operation 9 months before the birth. For delivery in a public facility and vaccination, it is 1 if PBF is operating at the time of birth.⁸ The vector \mathbf{X}_{idt} contains child, mother and household characteristics (Table II).⁹ $Voucher_{dt}$, HEF_{dt} and $SUBO_{dt}$ indicate operation of a maternity voucher scheme, a

⁷In addition to the standard arguments commonly made for using least squares with binary outcomes (Angrist and Pischke, 2009), we have two reasons to prefer ordinary least squares in this application. First, the assumption required to identify the effect with a nonlinear estimator differs from the common trends in outcomes assumption required with least squares (Puhani, 2012). We assess the plausibility of the latter in Section 4.3 and find no strong reason to doubt it. Related, the means of the binary outcomes are far from the upper bound of 100% participation. Second, least squares allows us to test for heterogeneous effects explicitly through interactions between treatment and covariates. With a nonlinear estimator, the effect would be heterogeneous by default.

⁸For this outcome, the PBF indicator is not switched on if the scheme operating is the GAVI-HSS programme because this did not pay fees for deliveries in facilities.

⁹While most household characteristics are measured at the time of the survey, we assume they reflect household living conditions at the time of birth/pregnancy. Given this assumption and to avoid clutter, we do not make explicit in equation 1 that some of the covariates are measured at a different time from the outcome and treatment.

Health Equity Fund and the government fee waiver scheme, respectively. OD_{dt} and τ_t represent fixed OD and time (month) effects, respectively.

Model (1) imposes homogeneity in the effect of PBF across covariates. There are reasons to doubt this. Because public health services are more heavily utilized by the poorest, the impact on this group may be anticipated to be greater. On the other hand, the PBF schemes encouraged facilities to generate user fee revenue to finance incentive payments to staff. In the absence of effective fee waivers, this may have raised the price of services and discouraged utilization, particularly by the poorest patients. Where there are fee waivers provided by vouchers for antenatal care and institutional delivery, the impact of PBF would be anticipated to be greater.¹⁰ Geographic variation in access to public health services may constrain the extent to which even incentivized providers can influence utilization rates. These hypotheses lead us to estimate the following extended model,

$$y_{idt} = \delta PBF_{dt} + \gamma_1 PBF_{dt} \times Voucher_{dt} + \gamma_2 PBF_{dt} \times Poor_{idt} + \gamma_3 PBF_{dt} \times Urban_{dt} + \theta_1 Voucher_{dt} + \theta_2 HEF_{dt} + \theta_3 SUBO_{dt} + \mathbf{X}_{idt} \Psi + OD_{dt} + \tau t + u_{idt} \quad (2)$$

where $Poor_{idt}$ indicates that the household has a wealth score that places it within the poorest 40 percent of the cross-section¹¹ and $Urban_{dt}$ is 1 if the household is resident in an urban location.

Under the common trends assumption, the least squares estimate of δ gives the average effect of PBF on non-poor, rural observations located in ODs in which there is no maternity voucher scheme operating alongside PBF. The average effect on the treated is given by averaging the partial effect of PBF across all observations located in ODs in which PBF operates. This involves combining the estimate of δ with estimates of the (weighted) coefficients on the interaction terms. The standard error is estimated by the delta method with adjustment made for clustering at the OD level. The average effect on a sub-group, for example the poor, is given by averaging the partial effect of PBF across all observations in that sub-group that are exposed to PBF. A t -test of the null of no interaction between the treatment and group indicator, for example $\gamma_1 = 0$, is given.

Note that if model (2) is correct, then the estimate of β obtained from model (1) is not consistent for the weighted average treatment effect across groups. For this reason, we focus on the average effect estimated from model (2), or a restricted version of this model that drops insignificant interactions.

4.2. Heterogeneous effects by type of performance-based financing

To estimate the effects of different PBF schemes, we break the analysis into sub-periods within which various programmes operated simultaneously. First, we limit the sample to births in the 1996–2003 period using data from the 2000 and 2005 surveys. We estimate model (1), but with two treatment indicators to identify OD-time periods in which the two pilot schemes (*Pilot-IN* and *Pilot-OUT*) operated. Because of the limited number of districts (five) that were exposed to PBF in this period, we do not allow for any heterogeneity in the effects by poverty status and urban/rural location.¹²

Second, we estimate the effect of the *IN* and of the *internal contracting* schemes that were introduced between 2004 and 2006 by restricting the sample to births in the 2000–2008 period using data from all three cross-sections. We drop observations in ODs that had been exposed to contracting in the pilot phase in order to identify the effect of the introduction of each PBF programme relative to a baseline in which financing was not tied to performance targets. This leaves us with data from seven ODs in which the *IN* model was introduced for the first time, and from another seven in which *internal contracting* was introduced. Model (1) is estimated, but with a treatment indicator for each of the two contracting schemes.¹³ In a model similar to

¹⁰The PBF effect is not expected to vary with the Health Equity Fund and SUBO fee waiver schemes as, during the estimation period, these mainly operated at hospitals, not health centres where most of the antenatal care and deliveries take place.

¹¹The wealth score is constructed on the pooled data from a principal component analysis of a set of household assets and dwelling characteristics (Filmer and Pritchett, 2001).

¹²There were no maternity voucher schemes in operation during this period, nor did the government fee waiver scheme exist.

¹³We control in this model for the operation of the GAVI-HSS scheme, but do not report the coefficient on the indicator because the observation period after the introduction this scheme is too short to reliably estimate its effect.

(2), an indicator of each scheme is interacted with the poverty and urban indicators, and the effect of *internal contracting* is allowed to differ depending on whether a voucher scheme is in operation. The *IN* programme never operated alongside vouchers.

4.3. Assessment and weakening of common trends assumption

The means of the covariates in 1998, before the introduction of PBF anywhere, are given in Table II both for ODs in which PBF was introduced by 2010, which are loosely referred to as treated, and for those remaining without PBF until 2010 (control). For the most part, the two groups are very similar with respect to these characteristics at baseline. Women in districts that eventually receive PBF initially had lower education and were more likely to become pregnant as teenagers. However, the change in the means of these characteristics between 1998 and 2010 does not differ between the treated and control ODs. At baseline, children born in the ODs that eventually obtain PBF are more likely to be among the very poorest. The change in the proportion in the second and third poorest groups does differ between the treatment and control groups. The only other covariate that displays differential trends across the two groups is a birth order above three and the age of the head of household. On the whole, with respect to observables, the groups do not differ markedly at baseline and they do not experience grossly disparate changes over time.

To further gauge the plausibility of the common trends identification assumption, we test whether trends in the outcomes in the period prior to PBF were similar across ODs that eventually obtain PBF and those that did not. In relation to estimation of the average effect of any PBF and of the pilot programmes, for which the starting year of treatment is 1999, we look at outcomes in the 1995–1998 period. With respect to estimation of the second phase *IN* and the *internal contracting* programmes, which start from 2004, we look at outcomes in the 2000–2003 period. From all tests conducted across three outcomes, three samples and two periods, the null of common trends is rejected at the 10% level in only two cases—antenatal care and vaccinations in the 1995–1998 period comparing the pilot treatment ODs with all others (see Table A2 in Supporting Information).¹⁴ Effects of the PBF pilot on these two outcomes should therefore be interpreted with some caution.

The similarity of trends in the period prior to that used for estimation builds confidence in the plausibility of the identifying assumption. Nevertheless, we check the robustness of the estimates to restricting the controls to observations located in ODs that are most similar to the treated ODs with respect to outcomes prior to the estimation period. To do this, we aggregate the data to the OD level and estimate propensity scores from a probit model of the treatment group indicator of whether or not a PBF scheme is implemented in the OD by the end of the estimation period on the OD average of all three outcomes (antenatal care, delivery in public facility and vaccinations) in each of the pre-estimation period years (1995–1998 for the pooled and pilot phase models and 2000–2003 for the extension phase model). We then use Nearest Neighbour matching without replacement to find a control OD for each of the 32 treated ODs that are used to estimate the effect of any PBF over the 1999–2010 period (Ho *et al.*, 2007), and similarly for the other two periods.¹⁵ We also check robustness to weighting the controls according to their degree of similarity to the treated using the estimated propensity scores.¹⁶

Finally, we check robustness of the estimated effects of the pilot schemes to restricting the control districts to the four that were used in the randomized experiment (Bhushan *et al.*, 2007).

¹⁴This test is implemented by regressing each outcome variable on year dummies, a treatment group dummy and interactions between these. The null hypothesis that the interactions are jointly insignificant is tested by an *F*-test. The two rejections of the null obtained do not arise if the controls are restricted to those used in the randomized pilot experiment.

¹⁵Even in the full sample, the overlap in the distribution of propensity scores across the treated and control ODs is reasonably good, and trimming the controls through matching makes it even better (see Figure A2 in Supporting Information).

¹⁶A weight equal to $\hat{p}(\mathbf{Z}_{dt}) / (1 - \hat{p}(\mathbf{Z}_{dt}))$ is given to all the control observations located in a particular OD, where $\hat{p}(\mathbf{Z}_{dt})$ is the estimated propensity score defined on the OD average of each of the three outcomes in each of the pre-estimation years (\mathbf{Z}_{dt}) (DiNardo *et al.*, 1996). For the extension phase (2004–2008), we observe covariates, in addition to outcomes, from the pre-estimation period and so can include these in the propensity score estimation. Treated observations are given a weight of 1.

Table III. Estimated effects of any form of PBF, 1999–2010

	Delivery in facility		Antenatal Care	Vaccination
	Public	Any		
Without interactions (model (1)), all control ODs	0.069* (0.040)	0.038 (0.041)	0.030 (0.032)	0.023 (0.024)
Sample size	21 325	21 247	15 549	19 519
With interactions (model (2))				
All control ODs	0.075** (0.035)	0.045* (0.036)		
Sample size	21 325	21 247		
Control ODs selected by matching	0.061* (0.036)	0.030 (0.036)	0.032 (0.029)	0.019 (0.025)
Sample size	18 941	18 875	13 788	17 303
Control ODs weighted by propensity scores	0.059 (0.037)	0.031 (0.036)	0.029 (0.028)	0.010 (0.026)
Sample size	21 325	21 247	15 549	19 519
Mean outcome for treated:				
Baseline (1998)	0.044	0.052	0.255	0.356
Counterfactual (1999–2010 if had been no PBF)	0.300	0.371	0.699	0.450

PBF, performance-based financing; OD, Operational Districts.

Least squares estimates from models including birth period and OD fixed effects and covariates. Standard errors in parentheses are adjusted for clustering at the OD level. Estimated on a sample of births from 1995 to 2010. Covariates are those given in Table II plus indicators of fee waiver schemes and (for delivery and antenatal care) maternity vouchers. For models with interactions (2), the treatment effects are calculated by averaging partial effects of the PBF indicator and standard errors computed by delta method. For antenatal care and vaccinations, all interactions with the treatment indicator were found to be insignificant and are omitted. Hence, estimates in the first and second panels are identical for these outcomes. Bottom row gives the mean prediction from the model (with interactions and all controls) for treated ODs in the period 1999–2010 with the treatment indicator turned off.

*, ** and *** indicate significance at the 10%, 5% and 1% level, respectively.

5. RESULTS

5.1. Effect of any form of performance-based financing

In Table III, we present estimates of the average treatment effect on the treated of any form of PBF implemented between 1999 and 2010 on the probability of utilizing each of the three incentivized services. Without allowing for heterogeneity (model (1)), the point estimate is positive for all three outcomes, but is only significant for delivery in a public facility (6.9 pp).¹⁷ The remaining panels give estimates from model (2) and differ in how control observations are selected and weighted. After controlling for covariates in a flexible manner that allows for heterogeneity, the estimated effect of PBF on the probability of a child being born in a public facility is a significant 7.5 pp increase. This is dramatic compared with the baseline mean of 4.4%, but the latter is not an appropriate standard against which to assess the magnitude of the effect because, as is clear in Figure 1, utilization rates had increased markedly before the largest expansion of PBF in 2004–2005. Using the model estimates, the counterfactual average probability of a child being born in a public facility between 1999 and 2010 in the treated ODs if there had been no PBF implemented is 30%. This suggests that contracting raised the rate of delivery in a public facility by about one quarter. Limiting the control observations to those in ODs that most resemble the treated ODs with respect to pre-estimation period outcomes (third panel) or weighting the controls (bottom panel) results in a reduction in the magnitude of the estimated effect to around 6 pp, but it remains highly significant.

¹⁷Notwithstanding the caveats regarding use of a nonlinear estimator stated in footnote 7, logit estimates also reveal a significant effect only on the probability of delivery in public facility. The estimated effect on this outcome is somewhat larger than the ordinary least squares estimate at 9.5 pp.

Table IV. Heterogeneous effects of any PBF on delivery in a public facility 1999–2010

	Wealth index		With maternity voucher scheme	
	Poor	Non-poor	Yes	No
	0.024 (0.050)	0.134*** (0.028)	0.260*** (0.070)	0.070** (0.036)
Homogeneity test	0.012		0.011	
<i>p</i> -value				
Sample size	11 421	9904	962	20 363

PBF, performance-based financing; OD, Operational Districts.

Derived from least squares estimates of model (2). Estimates obtained by averaging the partial effect of treatment (PBF) indicator over observations within the respective sub-group. Standard errors in parentheses are obtained by delta method with adjustment for clustering on the OD level. Homogeneity test is a *t*-test of the null that the coefficient on the interaction between the treatment and group indicators is zero. *Poor* are those in bottom two quintiles of household wealth score within cross-section.

*, ** and *** indicate significance at the 10, 5, and 1% level, respectively. Other details as in notes in Table III.

Irrespective of how control ODs are selected and weighted, there is still no significant effect on the rate of vaccinations or antenatal care.

Heterogeneous effects of any type of PBF on the probability of delivery in a public facility by poverty status and whether PBF operates alongside a maternity care voucher scheme are presented in Table IV. PBF is estimated to raise the probability of delivery in a facility by 13.4 pp among the non-poor, while there is no significant effect on the probability of the poorest women delivering in a public facility. Operation of a maternal voucher scheme alongside PBF is found to almost quadruple the effect of the latter to a 26 pp increase in the probability of giving birth in a public facility. This suggests a very strong synergy in the effects of supply-side and demand-side incentives.

Performance-based financing is estimated to increase the probability of birth in any facility (public or private) by 4.5 pp (Table III, panel 2, column 2), which is only 60% of the impact on the probability of birth in a public facility. Further, the estimated effect on any type of institutional birth is not robust to restricting or weighting the control ODs. These results suggest that a substantial part of the positive effect on the probability of delivering in a public facility arises from switching the location of institutional births from private to public providers. This is consistent with a significant effect being obtained only for women who are not among the poorest, who could better afford to give birth in a private clinic. There is also no significant impact on the probability of delivery to be attended by a skilled birth attendant (results not presented). This suggests that the women who are persuaded to deliver in a public health centre are not those who would otherwise have given birth at home under the supervision of, at most, a traditional birth attendant.

5.2. Heterogeneous effects by type of performance-based financing

The estimated effects of the pilot PBF schemes introduced in 1999 are presented in Table V. The estimates obtained using all control ODs weighted equally indicate that both schemes had a significant and substantial positive impact on the probability of delivery in a public facility (top panel). The effect of the *Pilot-OUT* arrangement was substantially and significantly (*p*-value = 0.077) greater than the *Pilot-IN* scheme. No matter what controls are used and how they are weighted, the *Pilot-OUT* scheme is estimated to have had a significant positive impact on the rate of delivery in public facilities. The magnitude of the estimated effect is substantially reduced when only the control ODs from the randomized experiment are used.¹⁸ But even then, the effect is large

¹⁸Comparison of the treated ODs with the controls from the randomized experiment does not necessarily give the estimate that is least likely to be subject to bias. In addition to the chance that differences exist across a relatively small number of randomly selected treatment and control ODs, differences could also arise from the fact that three of randomly selected treatment ODs were not awarded contracts for PBF. Comparison of the ODs that actually experienced PBF with those from all over the country and, in particular, those selected to resemble the treated ODs most with respect to pre-treatment outcomes could provide a less biased estimate. Note that none of the control ODs from the randomized experiment are selected by matching to be those most similar to the treated ODs.

Table V. Estimated effects of pilot PBF schemes 1999–2003

	Delivery in public facility		Antenatal care		Vaccination	
	<i>IN</i>	<i>OUT</i>	<i>IN</i>	<i>OUT</i>	<i>IN</i>	<i>OUT</i>
All controls	0.093*** (0.033)	0.164*** (0.023)	0.288*** (0.096)	0.135 (0.126)	0.054 (0.044)	−0.024 (0.038)
Sample size	10 634		7174		9527	
Control ODs from randomized experiment	0.051 (0.081)	0.114* (0.060)	0.189** (0.065)	−0.036 (0.131)	−0.087 (0.069)	−0.092 (0.074)
Sample size	820		569		738	
Control ODs selected by matching	0.066 (0.049)	0.161** (0.060)	0.128 (0.084)	−0.034 (0.126)	0.028 (0.076)	−0.041 (0.067)
Sample size	967		670		860	
Control ODs weighted by propensity scores	0.095*** (0.026)	0.179*** (0.028)	0.207** (0.093)	0.117 (0.143)	0.063 (0.045)	−0.047 (0.042)
Sample size	10634		7174		9527	
Mean outcome for treated:						
Baseline (1997–1998)	0.059	0.06	0.264	0.265	0.363	0.364
Counterfactual (1999–2003 if had been no PBF)	0.064	0.036	0.358	0.504	0.483	0.618

PBF, performance-based financing; OD, Operational Districts.

Least squares estimates of model (1) on a sample of births from 1995 to 2003. Covariates are those given in Table II. Bold indicates that effects of *Pilot-IN* and *Pilot-OUT* are significantly different at the 10% level. Other details as in Table III.

relative to the mean in the pre-treatment period and to the counterfactual. The estimated effect of the other pilot scheme, which did not allow the contracted NGO to hire and fire staff and received less generous per capita funding, falls in magnitude and is insignificant when the control ODs are limited either to those from the randomized experiment or to those matched on pre-treatment outcomes. The loss of significance is partly due to the very large reduction in sample size when these restrictions are made, although the point estimate is also reduced. The estimated effect is robust in magnitude and significance to using the full set of controls and weighting them by the propensity score.

The effects of both pilot schemes on the probability of an institutional delivery in any facility are very similar to those on delivery in a public facility (see Table A3 in Supporting Information). This indicates that in this early period, PBF encouraged a movement from home birth to delivery in a public facility with little or no crowd-out of the utilization of private clinics.

Comparison of the two pilot schemes with respect to the impacts on antenatal care differs markedly from that regarding the effects on institutional delivery. Only the *Pilot-IN* has a significant positive impact on the rate of utilization of antenatal care. Significance is maintained with all selection and weighting of the controls, except when they are chosen by matching. It is not obvious why only the less powered of the two schemes in terms of the ability of the contractor to act on incentives appears to have had an impact on the uptake of antenatal care. Bhushan *et al.* (2007) also found the *Pilot-IN* model to have greater impact on antenatal care. Neither scheme had an impact on the rate of vaccinations.

The estimated effects of the schemes introduced in the second phase of PBF between 2004 and 2006 are presented in Table VI. *Internal contracting* is estimated to have had a positive impact on the probability of delivery in a public facility that is almost 50% of the counterfactual prevalence, irrespective of the precise method of estimation. The impact on public facility deliveries of the *IN* scheme is estimated to be considerably and significantly smaller. According to the baseline estimate obtained from including covariates and using all control ODs (second panel), it is only 20% of the effect of *internal contracting*. This could well be because of the fact that *internal contracting* offered a large payment per delivery in a health centre, while the *IN* model only

Table VI. Estimated effects of extension phase PBF schemes 2004–2008

	Delivery in public facility		Antenatal care		Vaccination	
	<i>IN</i>	<i>Internal contracting</i>	<i>IN</i>	<i>Internal contracting</i>	<i>IN</i>	<i>Internal contracting</i>
Without interactions (model (1)), all control ODs	0.009 (0.019)	0.132** (0.031)	0.098*** (0.036)	-0.024 (0.043)	0.038* (0.028)	-0.028 (0.093)
Sample size		10 742		7722		9926
With significant interactions (model (2))						
All control ODs	0.028** (0.017)	0.140*** (0.033)				
Sample size		10 742				
Control ODs selected by matching	0.018 (0.018)	0.126*** (0.033)	0.105*** (0.038)	-0.022 (0.044)	0.025 (0.033)	-0.041 (0.096)
Sample size		8373		5952		7721
Control ODs weighted by propensity scores	0.026** (0.017)	0.137*** (0.033)	0.089*** (0.038)	-0.027 (0.049)	0.028 (0.037)	-0.035 (0.098)
Sample size		10742		7722		9926
Mean outcome for treated:						
Baseline (2002–2003)	0.123	0.132	0.538	0.573	0.596	0.638
Counterfactual (2004–2008 if had been no PBF)	0.201	0.292	0.460	0.895	0.331	0.633

PBF, performance-based financing; OD, Operational Districts.

Least squares estimates obtained from sample of births from 2000 to 2008. For antenatal care and vaccinations, all interactions with the treatment indicator were found to be insignificant and are omitted. Hence, estimates in the first and second panels are identical for these outcomes. Global Alliance for Vaccines and Immunization Health System Strengthening not evaluated because there is little exposure to it by 2008. Bold indicates that effects of *IN* and *Internal contracting* schemes are significantly different at the 10% level. Other details as in notes in Table III.

specified an OD target for institutional deliveries and gave the OD discretion in specifying any incentive payment at the facility level. This scheme had no net impact on the probability of an institutional birth (see Table A3 in Supporting Information), while the estimated effect of *internal contracting* on delivery in any facility is three-quarters of that on delivery in a public facility indicating only partial crowding-out of private clinics.

As with the pilot, the relative effects of the two schemes on antenatal care are quite different from those on institutional delivery. Only the *IN* programme has a robust and highly significant estimated positive effect on the probability of a woman having at least two antenatal care visits, which is around one-fifth of the baseline rate. The absence of any effect of *internal contracting* may be because, unlike for institutional delivery, provision of ANC was not incentivized through a per case payment but only through a bonus paid conditional on reaching a target rate. Neither scheme has a significant, robust estimated effect on the vaccination rate.¹⁹

Contrary to the findings of Matsuoka *et al.* (2014) from a before-after comparison in one OD based on administrative data that are potentially contaminated by over-reporting in response to the financial incentives (Lim *et al.*, 2008), the *GAVI-HSS* scheme is estimated to have no significant effect on either of the two incentivized services observable in the data.²⁰ In part, this may be because of relatively high baseline rates of

¹⁹Heterogeneity of the effects of the two schemes operating in the second phase on the rate of public facility deliveries is not presented because they are similar to those given in Table IV for the effect of any PBF. Differences are that there is a significant effect of *internal contracting* even among poor women and the estimated effect of this scheme is larger, but not significantly so, in rural areas.

²⁰Results are available on request. To estimate the effect of the *GAVI-HSS* scheme introduced between 2007 and 2009 on vaccinations and antenatal care, we restrict the sample to children born in the 2005–2010 period using data only from the 2010 survey. We drop ODs that had previous exposure to any kind of PBF and two ODs in which Special Operating Agencies came into operation in this period.

Table VII. Estimated effects of PBF on neonatal mortality

	Any PBF (1999–2010)	Extension phase (2003–2008)	
		IN	Internal contracting
	–0.003 (0.007)	0.0026 (0.0049)	0.0023 (0.0080)
Sample size	45 489	18 477	
Number of births in districts with PBF	4524	1217	751
Mean neonatal mortality for treated			
Baseline	0.044	0.040	0.043
Counterfactual	0.043	0.036	0.015

PBF, performance-based financing.

Neonatal mortality is number of deaths within a month of birth per 100 births. Least squares estimates of the effect on the probability of death within a month of birth. Models estimated using all control Operational Districts. Specification is model (1), that is, with full set of covariates but, given limited sample size, without allowing interactions between PBF and any covariate.

utilization. Even before the scheme was implemented, 80% of mothers in the treatment areas made at least two antenatal care visits, and two-thirds of children were fully vaccinated. Further increases from these levels may be difficult to realize, particularly for a scheme with only modest per capita funding.

5.3. Effects of PBF on neonatal mortality

The ultimate aim of PBF is to improve health outcomes. Institutional delivery is highly incentivized because it is considered instrumental to improving maternal and infant health (Bhutta *et al.*, 2014). Given that part of the positive effect we find on delivery in public facilities appears to be due to a switch from private clinics, any health impact may be muted. We investigate this by estimating the effect of PBF on neonatal mortality – death within a month of birth.²¹ To increase sample size and so power, we identify mortality for all births the mother reports over the 10 years preceding the survey interview. Even with this extended recall period, there are too few observations to make a reliable estimate of the mortality rate in the three ODs in which the Pilot programme operated from 1999 to 2003. We do not attempt to estimate the impact of either the Pilot-IN or Pilot-OUT schemes on neonatal mortality. The estimated effects of any PBF over the 1999–2008 period and of the two schemes that operated in the extension period are presented in Table VII. In all cases, there is a reasonably precisely estimated zero effect. Even the *internal contracting* programme, which we estimate raised the probability of an institutional delivery in any facility by 10 pp (Table A3 in Supporting Information), appears to have had no impact on neonatal mortality.

6. CONCLUSION

Much faith is being placed in the potential of PBF to increase the quantity and quality of medical care delivered by poorly resourced health systems. This study provides evidence that suggests that this hope is only partially justified. PBF schemes implemented over a decade in Cambodia are estimated to have raised the proportion of babies born in a public clinic by around one-quarter, on average. However, there is no consistent, significant effect on two other incentivized services – antenatal care and child vaccination. The absence of an effect on antenatal care is consistent with findings from Rwanda (Basinga *et al.*, 2011; Gertler and Vermeersch, 2012) and Burundi (Bonfrer, Van de Poel *et al.*, 2014) and is likely to be due to the marginal cost of finding and convincing pregnant women to come for regular check-ups that is high compared with the small monetary incentive that may only be a bonus tied to a provision target. The financial incentives for institutional delivery is

²¹We do not attempt to identify impact of PBF on child health outcomes measured at the time of survey because these are less likely to be a direct consequence of the care around pregnancy/birth that was incentivized by the PBF programmes.

higher – particularly when implemented as a per case payment – and it is less costly for providers to encourage women who have already come into contact with the facility to give birth in it (Gertler and Vermeersch, 2012). The absence of an effect on the vaccination rate may be due to the intensive efforts made by the Extended Programme on Immunization to raise nationwide coverage through twice yearly outreach activities and campaigns that left less room for PBF to have a marginal impact.

There is no significant effect on the rate at which the poorest women give birth in a public facility. This suggests that the effectiveness of PBF is muted when there is a lack of effective demand for the services that providers are incentivized to deliver. Further support for this conclusion is given by the finding that the effect on the rate of delivery in public clinics is almost quadrupled when PBF is accompanied by the distribution of vouchers to cover user fees. Attention needs to be paid to achieving an appropriate balance between supply-side incentives and demand-side coverage.

Over the full period considered, the effect of any form of PBF on institutional births was only half of that on the rate of delivery in public facilities, and there was no effect on skilled birth attendance. This suggests that PBF caused a shift from delivery in private clinics and under the supervision of trained midwives at home to delivery in public health centres. The health effects of such a shift are not obvious, and indeed we find no significant effect on neonatal mortality. This holds even when we restrict attention to one form of PBF (*internal contracting*) that did succeed in raising the rate of institutional births by 10%. Chari and Okeke (2014), who estimate that PBF raised institutional deliveries by a similar magnitude in Rwanda, also find no impact on neonatal mortality. This is consistent with growing scepticism (Das, 2014), backed by evidence (Mazumdar *et al.*, 2011), about the potential for incentivized institutional delivery to improve neonatal health outcomes when the quality of care is seriously deficient because of a lack of equipment and trained personnel. In that context, incentivizing individuals to receive services or providers to deliver them may do little to improve health.

Evidence on the effectiveness of PBF is also lacking in respect of the design of contractual arrangements and the structure of incentives. The Cambodian experience suggests that halfway houses are best avoided. The pilot phase revealed that if the management of a health district is to be contracted to an agency, such as an NGO, then the contractor should not be constrained to operate within public sector employment and procurement rules. However, granting a contractor full autonomy to manage public health facilities and manpower is likely to be politically infeasible in many contexts, as it proved to be in Cambodia. The experience of the extension phase of PBF in that country provides some support for the hypothesis that public sector management operating through internal contracting is more effective than a partially autonomous external management agency.

There are limitations to this study. First, and most important, PBF was not randomly allocated across districts. Our estimates can only be interpreted as causal under the assumption that, conditional upon covariates, districts which did not receive PBF in a particular period provide a good counterfactual of what would have happened to those exposed to PBF in its absence. This assumption cannot formally be tested, but the similarity in pre-treatment outcome trends and the robustness of results to the selection and weighting of controls gives confidence that results are not merely driven by selection bias. Second, the introduction of PBF schemes in Cambodia, as in most other contexts, was accompanied by an increase in budgets. The estimated effects cannot be attributed not only to the change in incentives, but are also driven by the increase in resources. Given the administrative and operational costs that come with PBF, establishing its cost-effectiveness relative to the more traditional input based financing is a priority for future prospective studies. Third, the retrospective nature of the study increases the potential for measurement error. The absence of significant estimated effects on two of the three incentivized services examined may be partly attributable to the difficulty of linking the timing of vaccinations to the starting date of each PBF scheme and to errors in retrospective reporting of antenatal care, which is likely to be substantially greater than for recall of the location of birth.

Notwithstanding these limitations, this paper suggests that PBF can raise utilization of at least some health services. However, with heavy reliance on user fees, which is the case in most PBF schemes now being implemented in sub-Saharan Africa, there is a risk that the poorest are left behind when average rates of utilization are raised. Combining PBF with demand-side mechanisms that waive the cost of user fees could mitigate this

risk. This study also suggests that research needs to move beyond the question of whether PBF is effective in raising utilization to the more pertinent one of how exactly to make finance contingent on performance in order to best improve the medical care delivered in low-resource settings and achieve gains in health outcomes.

ACKNOWLEDGEMENTS

The study was funded through EU-FP7 research grant HEALTH-F2-2009-223166-HEFPA on “Health Equity and Financial Protection in Asia”. Ellen Van de Poel was supported by the Netherlands Organization for Scientific Research (Veni project 451-11-031). We thank Timothy Johnston and Emre Ozaltin, previously Phnom Penh office of the World Bank, Joan Bastide and individuals who commented on our work at the HEFPA workshop in Yogyakarta, the Health System Reform in Asia Conference in Singapore and the doctoral workshop on Health Economics and Policy in Grindelwald.

REFERENCES

- Angrist JD, Pischke JS. 2009. *Mostly Harmless Econometrics: An empiricist's Companion*, Princeton University Press: Princeton.
- Annear PL, Bigdeli M, Chhun Eang R, Jacobs B. 2008. Providing access to health services for the poor: health equity in Cambodia. *Studies in Health Services Organisation and Policy* **23**: 189–226.
- Basinga P, Gertler PJ, Binagwaho A, Soucat ALB, Sturdy J, Vermeersch CMJ. 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.
- Bonfrer I, Van de Poel E, Van Doorslaer E. 2014. The effects of performance incentives on the utilization and quality of maternal and child care in Burundi. *Social Science and Medicine* **123**: 96–104.
- Bonfrer I, Soeters R, Van de Poel E, Basenya O, Longin G, van de Looij F, Van Doorslaer E. 2014. Introduction of performance-based financing was associated with improvements in care and quality. *Health Affairs* **12**: 2179–2187.
- Bhushan I, Bloom E, Clingingsmith D, Hong R, King E, Kremer M, Loevinsohn B, Schwartz JB. 2007. Contracting for Health: Evidence from Cambodia, : Cambridge, MA. Available at <http://faculty.weatherhead.case.edu/clingingsmith/cambodia13JUN07.pdf>
- Bhutta ZA, Das JK, Bahl L, Lawn JE, Salam RA, Paul VK, *et al.* 2014. Can available interventions end preventable deaths in mothers, newborn babies, and stillbirths, and at what cost? *The Lancet* **384**(9940): 347–360.
- Braye A. 2007. Cambodia Health Services Contracting Review Final Report, Agence Francaise de Developpement: Clichy France.
- Chari AN, Okeke EN. 2014. Can institutional deliveries reduce neonatal mortality? Rand Labor and Population, Working Paper -1072. Santa Monica, CA: Rand Cooperation
- Das J. 2014. Are institutional birth institutionalizing deaths? World Bank Future Development Blog, available at <http://blogs.worldbank.org/futuredevelopment/are-institutional-births-institutionalizing-deaths>
- DiNardo J, Fortin NM, Lemieux T. 1996. Labor market institutions and the distribution of wages, 1973–1992: a semiparametric approach. *Econometrica* **64**: 1001–1044.
- Filmer D, Pritchett LH. 2001. Estimating wealth effects without expenditure data—or tears: an application to educational enrollments in states of India. *Demography* **38**: 115–32.
- Flores G, Ir P, Men CR, O'Donnell O, Van Doorslaer E. 2013. Protection of patients through compensation of providers: the impact of Health Equity Funds in Cambodia. *Journal of Health Economics* **32**: 1180–93.
- Fritsche GB, Soeters R, Meessen B (eds). 2014. Performance-Based Financing Toolkit. World Bank: Washington, DC.
- Fujita N, Abe K, Rotem A, Tung R, Keat P, Robins A, Zwi AB. 2013. Addressing the human resources crisis: a case study of Cambodia's efforts to reduce maternal mortality (1980–2012). *BMJ Open* **3**(5): doi: 10.1136/bmjopen-2013-002685.
- Gertler P, Vermeersch C. 2012. Using performance incentives to improve health outcomes. In: Policy Research Working Paper, 6100, pp. 1e39.
- Ho DE, Imai K, King G, Stuart EA. 2007. Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Political Analysis* **15**: 199–236.
- Keller S, Schwartz JB. 2001. Final evaluation report: contracting for health services pilot project. ADB Loan No. 1447-CAM. Phnom Penh: Asian Development Bank.

- Lim S, Stein D, Charrow A, Murray C. 2008. Tracking progress towards universal childhood immunisation and the impact of global initiatives: a systematic analysis of three-dose diphtheria, tetanus and pertussis immunisation coverage. *The Lancet* **372**: 2031–46.
- Liljestrand J, Sambath MR. 2012. Socioeconomic improvements and health system strengthening of maternity care are contributing to maternal mortality reduction in Cambodia. *Reproductive Health Matters* **20**: 62–72.
- Loevinsohn B, Harding A. 2005. Buying results? Contracting for health service delivery in developing countries. *The Lancet* **366**: 676–681.
- Matsuoka S, Obara H, Nagai M, Murakami H, Chan Lon R. 2014. Performance-based financing with GAVI health system strengthening funding in rural Cambodia: a brief assessment of the impact. *Health Policy and Planning* **29**(4): 456–465.
- Mazumdar S, Mills A, Powell-Jackson T. 2011. Financial incentives in health: New evidence from India's Janani Suraksha Yojana. *Social Science Research Network*. Available at http://www.researchgate.net/publication/255707981_Financial_Incentives_in_Health_New_Evidence_from_India's_Janani_Suraksha_Yojana?ev=srch_pub Accessed: 28 February 2012.
- Meesen B, Soucat A, Sekabaraga C. 2011. Performance-based financing: just a donor fad or a catalyst towards comprehensive healthcare reform. *Bulletin of the World Health Organization* **89**: 153–156.
- Ministry of Health Cambodia. 2006. Health Sector Support Project Mid Term Review Report, Ministry of Health: Phnom Penh.
- National Institute of Statistics, Directorate General for Health, ICF Macro. 2011. Cambodia Demographic and Health Survey 2010, National Institute of Statistics, Directorate General for Health, and ICF Macro: Phnom Penh, Cambodia and Calverton, Maryland, USA.
- Pearson M, Pun S. 2009. GAVI Health System Strengthening Support Evaluation: Cambodia Case Study. RFP 0006-08. London: hlsp.
- Puhani P. 2012. The treatment effect, the cross difference, and the interaction term in nonlinear 'difference-in-differences' models. *Economics Letters* **115**: 85–7.
- Soeters R. 2011. Why there is so much enthusiasm for performance-based financing, particularly in developing countries. *Bulletin of the World Health Organization* **89**(9): 621–700.
- Soeters R, Griffiths F. 2003. Improving government health services through contract management: a case from Cambodia. *Health Policy and Planning* **18**: 74–83.
- Van de Poel E, Flores G, Ir P, O'Donnell O, Van Doorslaer E. 2014. Can vouchers deliver? An evaluation of subsidies for maternal health care in Cambodia. *Bulletin of the World Health Organization* **92**: 331–339.
- Witter S, Fretheim A, Kessy F, Lindahl A. 2012. Paying for performance to improve the delivery of health interventions in low- and middle-income countries. *Cochrane Database Syst Rev* 2012 (2): (CD007899).

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article at publisher's web site.