Original article

Effectiveness of a pay-for-performance intervention to improve maternal and child health services in Afghanistan: a cluster-randomized trial

Cyrus Y. Engineer,[1,2]* Elina Dale,[3] Anubhav Agarwal,[4] Arunika Agarwal,[4] Olakunle Alonge,[2] Anbrasi Edward,[2] Shivam Gupta,[2] Holly B. Schuh,[2] Gilbert Burnham[2] and David H. Peters[2]

[1]Towson University, Department of Interprofessional Health Studies, Towson, MD, USA, [2]Johns Hopkins University, Department of International Health, Baltimore, MD, USA, [3]World Health Organization, Health Systems Governance and Financing, Geneva, Switzerland, [4]Harvard T.H. Chan School of Public Health, Department of Global Health and Population, Boston, MA, USA

*Corresponding author. 8000 York Road, Towson, MD, 21252, USA. E-mail: cengine1@jhu.edu, cengineer@towson.edu

Abstract

Background: A cluster randomized trial of a pay-for-performance (P4P) scheme was implemented in Afghanistan to test whether P4P could improve maternal and child (MCH) services.

Methods: All 442 primary care facilities in 11 provinces were matched by type of facility and outpatient volume, and randomly assigned to the P4P or comparison arm. P4P facilities were given bonus payments based on the MCH services provided. An endline household sample survey was conducted in 72 randomly selected matched pair catchment areas (3421 P4P households; 3427 comparison).The quality of services was assessed in 81 randomly sampled matched pairs of facilities. Data collectors and households were blinded to the intervention assignment. MCH outcomes were assessed at the cluster level.

Results: There were no substantial differences in any of the five MCH coverage indicators (P4P vs comparison): modern contraception(10.7% vs 11.2% ($P = 0.90$); antenatal care: 56.2% vs 55.6% ($P = 0.94$); skilled birth attendance (33.9% vs 28.5%, $P = 0.17$); postnatal care (31.2% vs 30.3%, $P = 0.98$); and childhood pentavalent3 vaccination (49.6 vs 52.3%, $P = 0.41$), or in the equity measures. There were substantial increases in the quality of history and physical examinations index ($P = 0.01$); client counselling index ($P = 0.01$); and time spent with patients ($P = 0.05$). Health workers reported limited understanding about the bonuses.

Conclusions: The intervention had minimal effect, possibly due to difficulties communicating with health workers and inattention to demand-side factors. P4P interventions need to consider management and community demand issues.

Key words: Pay-for-performance, Cluster-randomized trial, Maternal and child health, Performance based financing, Incentives, Health worker motivation

---

Key Messages

- Despite high expectations, pay for performance (P4P) incentives to improve maternal child health services do not always work as intended at the population level, as demonstrated in the Afghanistan P4P intervention.
- P4P is intended to improve health worker motivation and satisfaction, but this did not occur in the Afghanistan study. Despite this, the P4P still had a positive effect on health worker behaviour related to improvements in three measures of technical quality of care at outpatient facilities, although the intervention did not have any impact on any of the other 17 measures of quality at health facilities that were less directly under health worker control.
- The inattention to demand-side factors and difficulty in communicating to health workers about the intervention may have undermined the potential effects of the P4P intervention. More attention needs to be given to these factors in the design, management and implementation of P4P programmes.
- More work is needed to understand the relationship between P4P incentives and its effects on health worker motivation and behaviour. A more sophisticated understanding of organizational culture, leadership, management and psychology are needed rather than a simple expectation that more money will result in better performance.

Introduction

Performance-based financing for health services has become popular with international donors as an effort to achieve the Millennium Development Goals (MDGs).[1] Pilot projects using a pay-for-performance (P4P) approach, where health workers are paid based on the volume and/or quality of services provided, have rapidly expanded around the world, especially in low- and middle-income countries, yet without a similar growth in the empirical evidence base.[2,3] Afghanistan started its own P4P programme with support from the World Bank (WB) in 2010 as a means of increasing coverage and quality of priority maternal and child health (MCH) services, covering about one-third of the country. This paper reports on the main outcomes of a large-scale cluster-randomized trial of P4P that ran between September 2010 and December 2012.

Afghanistan has made impressive gains in the delivery of a Basic Package of Health Services (BPHS) since 2003, though concerns remain about the low level of coverage and quality of MCH services.[4] For example, use of modern contraceptives in rural Afghanistan had increased over 3-fold, from 5% in 2003 to 16% in 2006, antenatal care (ANC) use increased from 5% to 32% and the proportion of women who delivered with a skilled practitioner rose from 14% to 19%.[5–7]

Encouraged by the P4P experiences in Cambodia and Rwanda,[8,9] the Ministry of Public Health (MOPH) decided to test the approach in Afghanistan's growing health system,

where non-governmental organizations (NGOs) and the MOPH each provide the BPHS in entire provinces. Afghanistan had previously tested performance-based contracts and other contracting mechanisms to deliver BPHS during 2004-08. Contracting methods gained greater improvements in health services compared with non-contracted areas, but the bonuses were small and not given to health providers.[10] Contracting and incentives addressing equity concerns have shown that providers have autonomy to innovate and stimulate demands in various ways, including recruiting more CHWs.[11] The P4P intervention's specific objectives were to increase key MCH service coverage (by addressing low motivation of providers and poor quality of patient-provider interactions), make services more equitable, improve the motivation of health workers, and raise patient satisfaction and the technical quality of BPHS services. The P4P was applied at the health facility (HF) level, making cluster-randomized trial an appropriate design, with HF defined as the cluster, and population estimates of MCH services measured in each facility's catchment area.

Methods

Intervention design

P4P bonuses were provided quarterly to health workers, based on the volume of nine health services at each facility reported through the Health Management Information

System (HMIS) (Table 1), with additional annual payments also made based on two measures of equity of service provision, a balanced scorecard that addresses quality of services, and contraceptive prevalence rates (CPR) in HF catchment areas. Funds to the health workers were channelled through the NGOs managing those facilities, and paid on top of their regular budgets. The NGOs' central offices retained 10% of the performance payment. The total amounts paid were adjusted by a quality score based on a National Monitoring Checklist (NMC), which was assessed quarterly by an independent team of provincial MOPH officers and consisted of items related to equipment functionality, drug availability, quality of medical charts and number of households visited by CHWs. Each NGO negotiated with the MOPH to adjust their payment based on their baseline conditions and expected improvements, with an intention to account for the differences in insecurity and geographical inaccessibility that vary by facility. Health facilities submitted monthly reports on the volume of services provided, which were verified quarterly by independent monitors, record-matching and random home visits of patients reported as service users. Systematic audits of 1100 HF visits verified over 95% of the medical records used for payments, and random sampling of over 29 000 household visits based on medical records verified 89% of the reported services. The NMC reports were submitted directly to the MOPH. The MOPH's independent unit authorized quarterly payments after result verification, usually with a 1-month lag at the end of each quarter. HF managers distributed the performance incentives in their own way, which included giving individual bonuses proportional to the health worker's salary, giving them in equal amounts to all staff, or giving them based on their determination of an individual's contribution to the

performance indicators. The bonus amounts paid were about 6-11% above their base salary in 2011, and increased to about 14-28% in 2011, depending on the health worker's cadre. The increase was made in response to complaints about the level of the bonuses. Payments were incorporated with the regular pay checques, and communication of the bonus amounts to health workers was left up to the NGO and facility managers.

Randomization and masking

Following an initial 3-month pre-pilot phase in June 2010 in two provinces, the P4P was rolled out sequentially between September and December 2010 to 11 of 34 provinces chosen by the MOPH in consultation with the WB and representing different areas of Afghanistan (see Figure 1). All 442 BPHS health facilities were stratified within each province by facility level, matched according to the average number of outpatient visits per month, and then randomly assigned to the P4P or comparison arm. The comparison arm's facilities provided care as usual without bonuses. NGOs managing facilities were contracted by the MOPH to provide services throughout a province, thereby

Table 1. Performance indicators and payments

Maternal and Child Health Services	Amount paid per unit per quarter/unit cost	
	Initial rate	Revised rate
1. First antenatal care visit (ANC1)	USD 1·30	USD 2·67
2. Second antenatal care visit (ANC2)	USD 1·30	USD 2·67
3. Third antenatal care visit (ANC3)	USD 1·30	USD 2·67
4. Fourth antenatal care visit (ANC4)	USD 1·30	USD 2·67
5.Skilled birth attendance cases (SBA)	USD 10·37	USD 35·63
6. First postnatal care visit (PNC1)	USD 1·30	USD 2·67
7. Second postnatal care visit (PNC2)	USD 1·30	USD 2·67
8. Pentavalent3 vaccination	USD 3·00	USD 3·00
9. Tuberculosis (TB) case detection	USD 5·00	USD 5·00

Rates were revised in October 2011.

Source: Results Based Financing Operations Manual, Ministry of Public Health, Afghanistan
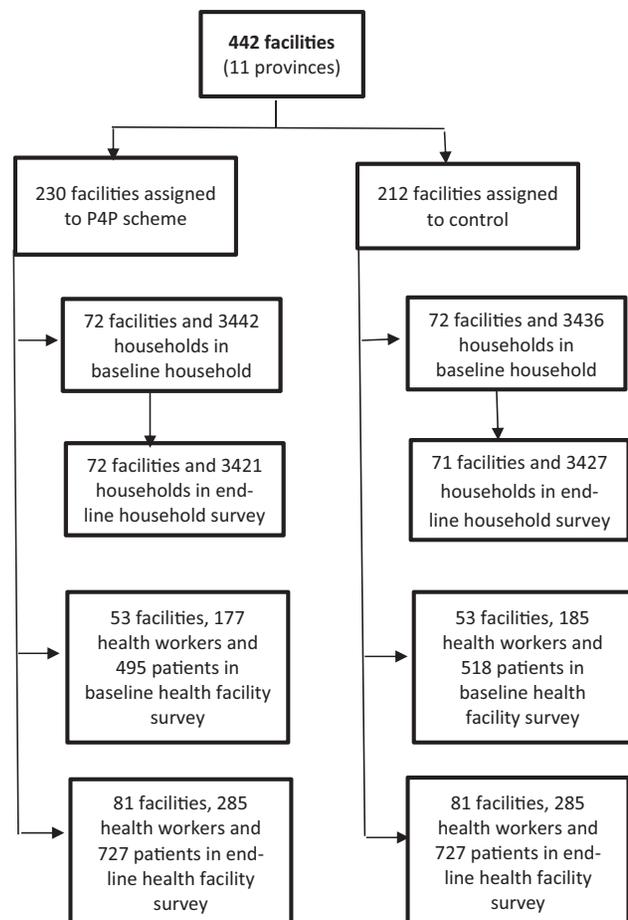

Figure 1. Pay for Performance Trial Scheme.

managing both intervention and comparison sites, potentially preventing 'contamination' of P4P to comparison sites. The MOPH, in consultation with the WB, did the randomization. There was no masking of the health workers or managers due to the nature of the intervention, though the intervention was not advertised to the public.

Population level outcomes were assessed through household surveys, with a baseline survey conducted in late 2010, and an endline survey in late 2012, 23-25 months after the initial rollout of the P4P scheme, varying by province and at the planned end of the study. The baseline surveys provided information to the MOPH about health conditions in the study area, and demonstrated that the P4P and comparison areas were similar with respect to study outcomes and demographic characteristics at the beginning of the trial (results in the Appendix, Tables A1 and A2, available as Supplementary data at *IJE* online). This paper uses the endline data. Both surveys used the same multi-stage probability-sampling scheme, with the same HF pairs as the primary sampling unit. In all, 72 matched HF pairs were randomly selected from the nine provinces that were safe enough for the data collection team to conduct the household survey in 2010. Two villages were randomly selected from all villages in each HF catchment area (catchment areas are contiguous and cover the entire population); a household listing and map were created prior to the survey, from which 24 households were randomly selected per village, with respondents comprising the head of household and each woman aged 12-49 years. Equal cluster sizes were assumed. On average, each HF provides services to about 10 villages, with an average size of 250 households per village, and an average household size of seven people. A total of 6908 households were interviewed in the endline survey, including 8162 ever-married women aged 12–49 years and 7821 children under the age of 5 years. Women who delivered in the previous 2 years were used for assessing ANC, delivery and postnatal care (PNC) services, and children aged 12-23 months were used for immunization coverage. The recall period for outpatient services was 2 weeks. After the baseline survey, it was found that for the most conservative indicator, pentavalent 3 coverage among children aged 12-23 months, the sample size had 98% power to detect the desired effect size of 0.2 at significance level of 0.05, compared with an *ex ante* estimate of 84% power.

HF level effects were assessed through an HF survey in late 2012, implemented in the same 72 randomly selected matched pairs of health facilities used for the household survey and an additional 9 pairs of health facilities from the same provinces randomly selected as part of the annual Balanced Scorecard assessment of BPHS facilities across the country.[12–14] The Balanced Scorecard assessments included structured observation of the HF, interviews with 560 randomly selected health workers and observations and exit interviews with a random sample of 1468 patients.

Trained interviewers, who were masked to type of site, completed both surveys that were pre-tested, translated and back-translated for consistency. Verbal informed consent was taken from all participants, and institutional review board approval was obtained through the Afghanistan MOPH and Johns Hopkins University.

Outcome measures

The primary population-level outcome measures were identified before the trial, to represent important health services related to the MDGs 4 and 5. The measures were linked to the performance payments: contraceptive prevalence; proportion of deliveries with at least one antenatal care visit; proportion of deliveries with a skilled birth attendant; proportion of births with at least one postnatal care visit in the first 6 weeks; and proportion of children aged 12–23 years with pentavalent3 vaccination. Two measures of equity of service utilization: concentration index of institutional delivery; and concentration index of outpatient visits of children under 5; were calculated by standard methods[15] using a standardized wealth index for each province following the Filmer and Pritchett method.[16]

Other outcome measurements were made at the HF level using indicators from the Balanced Scorecard for the BPHS delivery, details of which are described elsewhere.[12–14] In 2012, the Balanced Scorecard had 20 indicators covering five domains of quality of care: Client and community perspectives, including an index of overall client satisfaction and perceived quality of care; Human resources perspectives, including a health worker satisfaction index and health worker motivation index; Physical capacity of HF inputs (drugs, equipment, infrastructure); Quality of service provision, measuring four processes of care; and Management systems (see the 2012 National Balanced Scorecard Report, available as Supplementary data at *IJE* online).

Statistical analysis

Analysis was by intention-to-treat at the cluster (HF) level using Wilcoxon matched-pairs signed-ranks testing for non-normally distributed continuous or binary outcomes and paired t-tests for normal continuous outcomes. *P*-value was set at the 0.05 level before analysis. Household survey data from two villages of the same facility formed a single cluster. The HF (our primary sampling unit) was treated as a single cluster, with analysis performed in STATA™ (version 12).[17] Because there were no differences between the

intervention and comparison populations at baseline with respect to the main health services outcomes or key socioeconomic and demographic characteristics, we used the endline results to compare intervention and comparison arms.[18] Cluster-level analysis was done by collapsing data in each cluster and constructing a relevant cluster level statistic which was used to estimate the Wilcoxon matched-pairs signed-ranks test to assess differences between the P4P and comparison samples.[18] Randomization assures that the resulting summary measures are statistically independent, thereby eliminating the need for adjustment for clustering effects. For randomized controlled trials (RCTs), conducting analysis at the same unit as the unit of allocation is considered to be the 'gold standard'.[19]

To further account for unmeasured factors that could influence the outcome during implementation, we conducted a multiple regression analysis to assess difference-in-differences between baseline and endline in the intervention and comparison sites (see Supplementary data, available at *IJE* online).

Role of the funding source

The trial was implemented by the MOPH. Johns Hopkins University was a third-party evaluator appointed by the MOPH. MOPH staff participated in field supervision of data collection activities, but had no other role in data collection analysis, manuscript preparation or submission.

All authors had full access to the study data and take responsibility for data integrity and accuracy of data analysis. The corresponding author had final responsibility to submit for publication.

Results

Population level impact

The key characteristics of the households sampled are shown in Table 2, demonstrating similar demographic and economic characteristics between intervention and comparison areas.

Results of the five key MCH service coverage indicators and two equity indicators are shown in Table 3. The P4P had no substantial impact on increasing the coverage or equity of the targeted MCH services at the population level. Although contraceptive prevalence, skilled birth attendance and postnatal care coverage increased from baseline levels, the levels were still low, with no difference between the P4P and comparison areas (see Table A3 and Figures A1 and A2, available as Supplementary data at *IJE* online). The difference-in-differences analyses were also similar, though there was a decline in pentavalent3 coverage in the pre-post period compared with the comparison areas (−5.7%, *P* 0.01). P4P did not have an effect on the equity of care use. The concentration index for equity of institutional delivery was above 0.10 in both arms of the

Table 2. Household and resident characteristics in endline survey (2012)

	P4P Intervention	Comparison	Difference	*P*-value
Number of facility catchment areas	72	71	–	–
Total households interviewed, (*n*)	3421	3427	–	–
Households per facility catchment [median (range)]	48 (43–49)	48 (42–50)	–	0.046-
Residents per household	7·4 (0.9)	7·3 (0.9)	0.2	0.153
Total number of children aged 5 years or more and adults (*n*)	21457	20848	–	–
Ever-married women of reproductive age (12–49 years) interviewed, (n)	4035	4128	–	–
Women of reproductive age with live births < 24 months prior to survey	39·9 (10.6)	39 (9.6)	0.9	0.818
Children 0–4 years of age, (*n*)	3969	3962	–	–
Children 12–23 months of age, (**n**)	734	692		
Wealth quintiles (%)[a]			–	0.070
Lowest	17·7 (1.9)	18·6 (2.1)	–	–
Second	20·5 (1.7)	25·3 (1.6)	–	–
Middle	21·5 (1.5)	24·8 (1.3)	–	–
Fourth	20 (1.6)	17·1 (2.0)	–	–
Highest	20·3 (2.4)	14·2 (2.0)	–	–
Women's age (mean years)	31·8 (1.8)	31·6 (1.7)	0.2	0.469
Currently married women of reproductive age (%)	94·8 (3.8)	94·5 (3.8)	3.7	0.503
Literate women of reproductive age (%)[b]	4·6 (5.8)	4·0 (4.9)	0.6	0.751
Child's sex [male, 12–23 months (%)]	51·2 (17.2)	53 (19.3)	1.8	0.569

Data are cluster mean or % unless otherwise stated.

[a]*P*-value is based on design-based chi-square.

[b]*P*-value is based on log-transformed proportions.

Table 3. Population-level maternal and child health service indicators

Indicator	P4P intervention		Comparison		ICC	*P*-value
	Mean	95% CI	Mean	95% CI		
Current use of modern family planning methods (%)	10.7	(7.9, 12.2)	11.2	(8.3, 12.9)	0.0951	0.9
At least one antenatal checkup from a skilled provider (%)	56.2	(50.1, 62.3)	55.6	(49.5, 61.8)	0.276	0.9
Skilled birth attendant present at latest delivery (%)	33.9	(28.1, 39.7)	28.5	(24.1, 33.0)	0.206	0.2
Postnatal checkup within 42 days of delivery by a skilled provider (%)	31.2	(25.8, 36.6)	30.3	(25.7, 34.9)	0.192	0.98
Children received pentavalent3 vaccination (%)	49.6	(43.5, 55.6)	52.3	(46.0, 58.6)	0.0266	0.4
Equity of institutional deliveries (Concentration Index)	0.1758	(0.105, 0.247)	0.1000	(0.026, 0.175)	–	0.3
Equity of children's utilization of outpatient services (Concentration Index)	0.0025	(−0.023, 0.028)	0.0047	(−0.018, 0.027)	–	0.98

There were 72 matched pairs. Statistical significance tested using Wilcoxon matched pair signed-ranks test. 95% confidence intervals (CIs) were calculated using paired t-tests.

ICC is the intracluster correlation coefficient indicating the degree of clustering of observations at the cluster/village level.

trial, indicating that wealthier women were more likely to get institutional deliveries than poorer women.

HF level impact

The P4P intervention had a positive impact on three measures of quality of service provision at BPHS facilities, but did not have an impact on any of the other 17 Balanced Scorecard indicators (Table 4). P4P providers spent more time with patients, conducted more complete histories and physical examinations and provided more counselling to patients. The P4P was intended to increase health worker motivation, yet indices of motivation and job satisfaction were same in both groups. P4P had no impact on client satisfaction, perceptions of quality of care or level of community involvement.

Discussion

An emerging body of evidence indicates that use of performance-based incentives can be an effective way to improve health services. especially in low- and middle-income countries.[3,8–10,20,21] However, the entity targeted (individuals vs organizations), size of incentive payments (absolute amount and relative to existing salary), frequency of payments, the number and type of included performance measures, definition of performance (absolute, relative, improvement), intra-facility distribution procedures, verification process and other aspects of these programmes vary widely from country to country.[3] Thus, the more relevant question is not whether P4P programmes work, but what circumstances and design characteristics may impact on their effectiveness. The minimal effects found in our study point to some of the pitfalls in the design and implementation of the P4P programmes, consistent with the emerging evidence from other studies.[22]

Problems with implementation likely dampened any potential effect. The scheme was rolled out in phases, but there were some delays, particularly with the initial payments. Based on the health workers surveyed, only 37.9% of health workers in the P4P sites recognized that they had received any payment from the P4P intervention, although 86.7% reported that the P4P HF had received performance payments (see Table A4, available as Supplementary data at *IJE* online). In fact, payments were made with their regular salary payments, but the amounts of the bonuses were not separately identified. Payments were directly deposited into people's bank accounts for the first time (previously staff were paid in cash), and levels of performance were not communicated to them on a consistent basis. However, there was communication from health workers to the MOPH to prompt an increase in payment levels in late 2011, suggesting that the MOPH was concerned that the level of incentives was not enough to produce the desired changes.

The community and HMIS verification analysis suggested that there was no major manipulation of the payment triggers by the health facilities, suggesting that the reporting of results for payments was likely to be largely genuine (see Supplementary data available at *IJE* online). It is still possible that there was insufficient time for the P4P intervention to have an impact at the population level, as there was only a 23-25 month period between the end-line household and facility surveys and when the intervention began depending on the rollout in the respective provinces. However, previous experience with contracting in Afghanistan showed that substantial changes in Balanced Scorecard indicators of the delivery of the BPHS could occur within 1 year.[10] The matched pair cluster randomized trial design also minimizes the potential for other factors to affect the P4P intervention group differently from their comparison groups.

Table 4. Balanced Scorecard indicators at P4P and comparison facilities

		P4P intervention		Comparison		ICC[b]	*P*-value
Domain	**Client and Community**	Mean (%)	95% CI	Mean (%)	95% CI		
1	Overall Client Satisfaction and Perceived Quality of Care Index	76.5	74.3; 78.6	75.1	72.8; 77.4	0.530	0.2
2	Community Involvement and Decision Making Index	86.9	82.3; 91.6	87.6	83.1; 92.1	–	0.9
Domain	**Human Resources**						
3	Health Worker Satisfaction Index	63.8	62.0; 65.6	63.4	61.8; 65.0	0.508	0.9
4	Health Worker Motivation Index	72.7	71.4; 74.1	72.0	70.4; 73.5	0.274	0.4
5	Salary Payment Current	59.6	50.1; 69.0	61.7	52.1; 71.2	0.631	0.6
6	Minimum Staffing Index	19.8	10.9; 28.6	17.3	8.9; 25.7	–	0.5
7	Provider Knowledge Score	65.9	64.0; 67.8	65.7	63.6; 67.8	0.295	0.8
8	Staff Received Training (in past year)	6.3	4.9; 7.8	6.1	4.9; 7.3	0.039	0.6
Domain	**Physical Capacity**						
9	Equipment Functionality Index	79.3	75.5; 83.1	77.9	74.6; 81.3	–	0.1
10	Pharmaceuticals and Vaccines Availability Index	80.4	76.8; 84.1	79.8	76.4; 83.1	–	0.4
11[a]	Laboratory Functionality Index (CHCs only, 18 pairs)	68.4	58.3; 78.5	74.0	68.5; 79.4	–	0.5
12	Clinical Guidelines Index	81.2	76.2; 86.2	80.6	76.0; 85.2	–	0.5
13	Functional Infrastructure Index	60.2	54.5; 66.0	55.7	49.0; 62.4	–	0.3
Domain	**Quality of Service Provision**						
14	Client Background and Physical Assessment Index	76.4	74.0; 78.8	72.3	69.3; 75.3	0.411	0.01
15	Client Counselling Index	35.3	32.1; 38.5	29.3	26.2; 32.4	0.468	0.01
16	Universal Precautions	56.7	51.8; 61.6	54.1	49.6; 58.6	–	0.2
17	Time Spent with Client	14.5	9.7; 19.4	8.6	5.2; 12.0	0.247	0.05
Domain	**Management Systems**						
18	HMIS Use Index	81.5	75.2; 87.8	81.5	76.3; 86.7	–	0.5
19	Financial Systems	6.3	0.8; 11.7	2.5	−1.0; 6.0	–	0.2
20	HF Management Functionality Index	46.5	43.2; 49.8	45.9	42.5; 49.2	–	1.0
	Composite score for the BSC	53.0	51.0; 55.0	51.4	4.6; 53.2		0.1

There are 81 pairs of health facilities. All *P*-values are calculated with Wilcoxon matched pairs signed-rank test for non-normally distributed continuous or binary outcomes and with paired *t*-test for normal continuous outcomes. All 95% confidence interval (CI) estimates are calculated with paired *t*-test.

[a]Laboratory functionality index applies to comprehensive health centre facilities only; the sample size for this indicator is 18 pairs.

[b]ICC is intracluster correlation coefficient. For indicators without facility-level variation, ICC is not estimated.

The other likely explanation lies with the design of the intervention itself. The theory of change for the P4P intervention involved the simple assumption that paying health workers on this basis would make them more motivated to be more productive and provide better quality of care. It was then expected that better quality of services would lead to increased client satisfaction and demand for services, which would result in an increased use of critical child and maternal health services. The rationale for why such a change would also occur in a more equitable manner was never clear except that health workers were going to be incentivized to provide care equitably. The results suggest that the linkages between payment and motivation of workers to improve targeted services require more finely-tuned understanding of human motivation, as well as more sophisticated approaches to managing organizations and individuals beyond performance payments (e.g. taking into account organizational culture, leadership, management and psychology, among other things).

In management sciences, there is a greater scepticism about the powers of incentive payments and their less than direct effect on motivation and performance.[23] Deming's theory of 'Pay is not a motivator' suggests that monetary or other incentives are not sufficient to initiate or sustain changes related to service delivery and utilization.[24] Rewards can be considered 'temporary motivators' and sustenance requires that employees are intrinsically motivated, which is further linked to the type of work and the way any job has been designed. Hackman and Oldham's research on job characteristics showed that jobs with higher task identity, skill variety and task significance elevate psychological states (meaningfulness of work) and are often linked with higher performance.[25]

The study findings suggest that mechanisms hypothesized for the P4P intervention to change provider motivation and community demand did not occur, even if there was some improvement in the technical quality of care provided by health workers. Using 20 indicators to assess the

quality of care would suggest that by chance one of the indicators would demonstrate a difference with a *P*-value < 0.05. Finding three indicators with *P* < 0.05 all in the same domain, comprising clinicians adhering to process of care standards, suggests this is a real change in this domain. However, the P4P sites were unable to translate the extra resources to change other aspects of quality of care or increase the coverage of health services.

The lack of more explicit attention to demand-side considerations of the population is the most glaring flaw of the approach. There was no component of the intervention designed to raise awareness or create demand in the communities, nor were there interventions designed to enable health providers to understand the barriers faced by the communities they serve or effectively engage in raising demand for MCH services. Unlike other trials, there was no CHW component. Future research is needed to understand how health care organizations can overcome barriers to utilization of health services and promote healthy behaviour.

The Rwanda P4P evaluation concluded that the largest effects are observed for services over which the provider has greatest control or which require least effort on the part of the provider, and for which facilities receive largest financial incentives.[9] Population-level effects, such as coverage for antenatal services or child immunization, also depend on patients' health-seeking behaviour. Without addressing demand-side constraints, it may be difficult to increase utilization of health care services substantially. A study evaluating a two-phase conditional cash transfer programme in Nicaragua was designed to address poverty by making cash transfers to poor rural households, conditional upon the children in these households attending school and making visits to preventive health care providers. The demand-side incentives (monitoring and enforcing compliance) were complemented with supply-side incentives through a performance-based scheme. The results suggest that a well-targeted strategy of supply-side performance incentives could, on its own, be enough to achieve and maintain high levels of health care service use among poor rural populations, as seen by the improvements in immunizations and growth monitoring and reductions in stunting.[21]

The Afghanistan experience suggests that there is a need to identify where the most important bottlenecks to service use are. Conditional cash transfers directed at households or patients can stimulate demand and could complement supply-side financing, as seen in India, Uganda and Brazil, to promote institutional deliveries and improve antenatal care.[26–28] In Rwanda, P4P included CHWs to promote outreach to communities.[9] Including CHWs should be considered if this approach is to continue in Afghanistan.

For P4P to continue as an approach to improve MCH services, effort is needed to ensure that the beneficiaries both benefit from and understand the reason for their bonuses. P4P may be effective in some places but does not replace the critical need to enhance the broader leadership and management capabilities of health services organizations and to understand and address issues concerning demand for services and barriers faced by communities and households.

Supplementary Data

Supplementary data are available at *IJE* online.

References

1. Meesen B, Soucat A, Sekabaraga C. Performance-based financing: just a donor fad or catalyst towards comprehensive health reform. *Bull World Health Organ* 2011;**89**:153–56.
2. Miller G, Babariaz KS. *Pay-for-Performance Incentives in Low- and Middle-Income Country Health Programs*. NBER Working Paper No. 18932. Cambridge MA: National Bureau of Economic Research, 2013.
3. Eijkenaar F. Pay for performance in health care: an international overview of initiatives. *Med Care ResRev* 2012;**69**:251–76.
4. Belay TA. *Building on Early Gains in Afghanistan's Health, Nutrition, and Population Sector: Challenges and Options*. Washington DC: World Bank, 2010.
5. Johns Hopkins Bloomberg School of Public Health & Indian Institute of Health Management Research. *Afghanistan Multiple Indicator Cluster Survey 2003: A Re-analysis of Critical Health Service Delivery Indicators*. Kabul: Afghanistan Ministry of Public Health, 2005.
6. Johns Hopkins University Bloomberg School of Public Health & Institute of Health Management Research. *Afghanistan Health Survey 2006: Estimates ofPpriority Health Indicators for Rural Afghanistan*. Kabul: General Directorate of Policy and Planning, Ministry of Public Health, 2008.
7. UNICEF. *Afghanistan: Progress of Provinces. Multiple Indicators Clusters Survey 2003*. Kabul: UNICEF, 2004.
8. Schwartz JB, Bhushan I. Improving equity in immunization through a public-private partnership in Cambodia. *Bull World Health Organ* 2004;**82**:661–67.
9. Basinga P, Gertler PJ, Binagwaho A, Soucat ALB, Sturdy J, Vermeersch CMJ. Effect on maternal and child health services in Rwanda of payment to primary health-care providers for performance: an impact evaluation. *Lancet* 2011;**377**:1421–28.
10. Arur A, Peters D, Hansen P, Mashkoor MA, Steinhardt LC, Burnham G. Contracting for health and curative care use in

Afghanistan between 2004 and 2005. *Health Policy Plan* 2010;**25:** 135–44.

11. Alonge O, Gupta S, Engineer C, Salehi A, Peters DH. Assessing the pro-poor effect of different contracting schemes for health services on health facilities in rural Afghanistan Health Policy Plan. *Health Policy Plan* 2015;**30:**1229–42.

12. Peters DH, Noor AA, Singh LP, Kakar FK, Hansen PM, Burnham G. A Balanced Scorecard for Health Services in Afghanistan. *Bull World Health Organ* 2007;**85:**146–51.

13. Edward A, Kumar B, Kakar F, Salehi AS, Burnham G, Peters DH. Configuring Balanced Scorecards for measuring health systems performance: evidence from five years' evaluation in Afghanistan. *PLOS Med* 2011;**8:**e1001066.

14. Johns Hopkins Bloomberg School of Public Health & Indian Institute of Health Management Research. *Afghanistan Health Sector Balanced Scorecard 2012-13*. Kabul: Afghanistan Ministry of Public Health., 2013.

15. Kakwani NC. *Income Inequality and Poverty: Methods of Estimation and Policy Applications*. New York, NY: Oxford University Press, 1980.

16. Filmer D, Pritchett L. Estimating wealth effects without expenditure data - or tears: an application to educational enrollments in states of India. *Demography* 2001;**38:**115–32.

17. Stata Corp. *Stata 12 Survey Data Reference Manual*. College Station, TX: Stata Press, 2011.

18. Hayes RJ, Moulton LH. *Cluster Randomised Trial*. Boca Raton, FL: Chapman & Hall/CRC Press, 2009.

19. Cochrane Library. *Cochrane Handbook for Systematic Reviews of Interventions*. Chapter 16.3. http://handbook.cochrane.org/). Version 5.1.0, March 2011. [Accessed June 6, 2014]

20. Eichler R. *Can 'Pay for Performance' Increase Utilization by the Poor and Improve the Quality of Health Services*. Washington, DC: Center for Global Development, 2004.

21. Regalia F, Castro L. Nicaragua: combining demand-and supply-side incentives. In: Eichler R, Levine R (eds). *Performance Incentives for Global Health: Potential and Pitfalls*. Washington, DC: Center for Global Development, 2009.

22. Eijkenaar F, Emmert M, Scheppach M, Schöffski O. Effects of pay for performance in health care: a systematic review of systematic reviews. *Health Policy* 2013;**110:**115–30.

23. Kohn A. Why incentive plans cannot work. *Harvard Business Review* 1993; 54–63.

24. Deming W.E. *The New Economics for Industry, Government & Education*. Cambridge, MA: Massachusetts Institute of Technology Center for Advanced Engineering Study, 1993.

25. Hackman, J.R, Oldham,GR. Motivation through the design of work: Test of a theory. *Organizational Behavior and Human Performance* 1976;**16:**250–79.

26. Zavier AJ, Santhya KG. How conditional cash transfers to promote institutional delivery can also influence postpartum contraception: evidence from Rajasthan, India. *Int J Gynaecol Obstet* 2013;**123**(**Suppl 1**):e43–46.

27. Ekirapa-Kiracho E, Waiswa PE, Rahman MH *et al*. Increasing access to institutional deliveries using demand and supply side incentives: Early results from a quasi-experimental study. *BMC Int Health Hum Rights* 2011;**11**(**Suppl 1**):S11.

28. Shei A, Costa F, Reis MG, Ko AI. The impact of Brazil's BolsaFamilia conditional cash transfer program on children's health care utilization and health outcomes. *BMC Int Health Hum Rights* 2014;**14:**10.