Social Science & Medicine 179 (2017) 61–73

Contents lists available at ScienceDirect

Social Science & Medicine

journal homepage: www.elsevier.com/locate/socscimed

Effects of Payment for Performance on accountability mechanisms: Evidence from Pwani, Tanzania

Iddy Mayumana [a], Jo Borghi [b], Laura Anselmi [c], Masuma Mamdani [a], Siri Lange [d, *]

[a] Ifakara Health Institute, P.O. Box 78 373, Dar es Salaam, Tanzania
[b] London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, UK
[c] Manchester Centre for Health Economics, University of Manchester, Oxford Road, Manchester M13 9PL, UK
[d] Chr. Michelsen Institute, Norway, P.O.Box 6033, N-5892 Bergen, Norway

ARTICLE INFO

Article history:
Received 25 June 2016
Received in revised form
13 February 2017
Accepted 13 February 2017
Available online 20 February 2017

Keywords:
Payment for performance
P4P
Performance-based financing
PBF
Results-based financing
RBF
Accountability
Tanzania

ABSTRACT

Payment for Performance (P4P) aims to improve provider motivation to perform better, but little is known about the effects of P4P on accountability mechanisms. We examined the effect of P4P in Tanzania on internal and external accountability mechanisms. We carried out 93 individual in-depth interviews, 9 group interviews and 19 Focus Group Discussions in five intervention districts in three rounds of data collection between 2011 and 2013. We carried out surveys in 150 health facilities across Pwani region and four control districts, and interviewed 200 health workers, before the scheme was introduced and 13 months later. We examined the effects of P4P on internal accountability mechanisms including management changes, supervision, and priority setting, and external accountability mechanisms including provider responsiveness to patients, and engagement with Health Facility Governing Committees. P4P had some positive effects on internal accountability, with increased timeliness of supervision and the provision of feedback during supervision, but a lack of effect on supervision intensity. P4P reduced the interruption of service delivery due to broken equipment as well as drug stock-outs due to increased financial autonomy and responsiveness from managers. Management practices became less hierarchical, with less emphasis on bureaucratic procedures. Effects on external accountability were mixed, health workers treated pregnant women more kindly, but outreach activities did not increase. Facilities were more likely to have committees but their role was largely limited. P4P resulted in improvements in internal accountability measures through improved relations and communication between stakeholders that were incentivised at different levels of the system and enhanced provider autonomy over funds. P4P had more limited effects on external accountability, though attitudes towards patients appeared to improve, community engagement through health facility governing committees remained limited. Implementers should examine the lines of accountability when setting incentives and deciding who to incentivise in P4P schemes.

1. Introduction

Since the 1990s, a variety of accountability mechanisms like user committees, suggestion boxes, performance appraisal of health workers and maternal death audits have been introduced in low income countries to enhance health services, but these initiatives often do not function adequately (Fox, 2015; McCoy et al., 2012).

Payment for Performance (P4P), also called Performance-based financing (PBF), has in recent years been widely promoted in low income countries to improve providers' motivation and accountability to deliver better care (Meessen et al., 2011; Witter et al., 2013) by paying bonuses based on the achievement of pre-specified performance targets (Ireland et al., 2011; Meessen et al., 2011; Njuki et al., 2012).

While there is a growing body of evidence evaluating the impact of P4P, the focus has been primarily on health service outcomes (Basinga et al., 2011; Bonfrer et al., 2014). Recent studies have paid more attention to context and the processes by which these outcomes are or are not achieved, and the effects of P4P on people

* Corresponding author.
E-mail addresses: imayumana@ihi.or.tz (I. Mayumana), Josephine.Borghi@lshtm.ac.uk (J. Borghi), laura.anselmi@manchester.ac.uk (L. Anselmi), siri.lange@cmi.no (S. Lange).

http://dx.doi.org/10.1016/j.socscimed.2017.02.022

within the health system, their relationships and the work environment (Bertone et al., 2016; Bertone and Meessen, 2013; Bhatnagar and George, 2016; Huillery and Seban, 2014; Lohmann et al., 2016; Paul et al., 2014; Renmans et al., 2016; Ssengooba et al., 2012). However, a review of P4P studies concludes that the findings are often contradictory, that context and design matter, and that the exact mechanisms that P4P trigger remain unknown (Renmans et al., 2016). This paper seeks to contribute to this emerging field by assessing whether and to what degree accountability processes were enhanced by the Tanzanian P4P scheme.

1.1. Study setting

In January 2011, the Government of Tanzania in collaboration with Clinton Health Access Initiative (CHAI) introduced a P4P scheme in Pwani Region, funded by the Government of Norway. The scheme provided incentive payments in six monthly payment cycles to all health facilities in the region offering maternal and child health services based on their achievement of pre-defined maternal and child health performance targets (Binyaruka et al., 2015; MoHSW, 2012). 70–75% of the bonus payments went to staff, approximately 10 percent of their salaries (Binyaruka et al., 2015:3–4). The rest was earmarked for facility improvement (MoHSW, 2012). The decision for how facility funds were to be spent was to be made by health workers and health facility governing committees (HFGC), comprised of facility in-charge and community members (URT, 2001) though the community members were not eligible for bonus payments. In order to receive bonus payments facilities had to open bank accounts.

Managers at Council and Regional levels received payments based on the achievement of facilities in their district/region and additional targets linked to drug stock-outs in their district/region.

To keep track of facility and district/regional performance, the Pilot Management Team (PMT), comprised of MOHSW and CHAI staff, issued score cards indicating their achievement per indicator, bonus earned, bonus distribution between facility and health workers, the number of health workers eligible, and the next targets. The implementation of P4P was accompanied by the introduction of an electronic District Health Information System (DHIS) used to track performance indicators. In each cycle the PMT and district managers organized two day performance feedback meetings with providers.

The P4P programme had a positive effects on two of the eight service delivery indicators: an eight percent increase in institutional deliveries and a ten percent increase in the provision of anti-malarials during pregnancy (Binyaruka et al., 2015).

1.2. Accountability measures and assumed pathways of change

In this study we differentiate between *internal accountability*, mechanisms that are aimed at relations within and between different levels of the health system; and *external accountability*, aimed at relations between health providers and clients (Cleary et al., 2013). P4P may improve *internal* accountability through more supportive supervision linked to the verification of performance data, by strengthening relations between managers and providers through joint incentives, and encouraging providers to place demands on higher levels (Meessen et al., 2011). P4P may affect *external* accountability by encouraging provider responsiveness to users (Meessen et al., 2011) to attract clients to meet targets (Meessen et al., 2007), and increasing outreach, and the financial autonomy linked to P4P may stimulate health facility governing committees that were otherwise inactive (Falisse et al., 2012). A complete overview of accountability mechanisms, and

Table 1
Accountability mechanisms and P4P.

Accountability mechanisms (adapted from Cleary et al., 2013)	Themes identified	Indicators	Assumed pathway of change
Internal accountability Human resource management	Content and frequency of supervision	Number of supervision visits in past 9 days % of staff who were supervised within 90 days	P4P involves frequent verification of performance data by district, regional and national managers. This serves to stimulate supervision visits, making them more frequent and focused (Bhatnagar and George, 2016) As district managers receive incentives based on facility performance they may also be motivated to visit facilities more often (Janssen et al., 2015)
	Organizational culture	Administrative/management meeting frequency Training selection based on need	P4P improves collaboration between managers and health workers as they work together towards a common goal (Janssen et al., 2015) P4P may foster team spirit and collaboration (Kalk et al., 2010)
Budgeting, planning, priority setting, target setting	Financial autonomy Prioritisation of resources for results	Facility has a Community Health Fund Number of Community Health Fund members Availability of drugs, supplies and equipment	P4P may encourage providers to enroll households in the community insurance scheme in order to increase service uptake and increase revenue available to the facility (premium contributions). The facility funds from P4P together with the local management of funds in bank accounts, enable providers to invest in commodities and equipment.
External accountability Facility Committees	Response to citizen voice	Existence of a facility governing committee Whether committee met in the last 90 days	P4P may stimulate the formation of and frequency of meetings of governing committees at facility level by: providing a budget for them to manage (the facility-level bonus payment); and by paying them incentives (Falisse et al., 2012)
Responsiveness to patients		Outreach activities carried out Provider kindness during delivery	In order to meet service coverage targets, providers may undertake more outreach activities (Huillery and Seban, 2014) In order to encourage patients to attend facilities, they may change their behaviour and become kinder

specific examples identified within the Tanzanian context, together with the assumed pathway of change is provided in Table 1. Proposed indicators to measure each mechanism are also described here and presented in the Data Collection section.

1.3. Data collection

This study used a mix methods design. Qualitative data was collected in five of the seven intervention districts selected to represent peri-urban (Kibaha town and Bagamoyo), rural (Mkuranga and Kisarawe) and remote settings (Mafia island). Fifteen health facilities were purposively selected to represent variations in level of care and ownership: thirteen were public, one was private, and one faith-based. The data collection took place in three rounds over the period December 2011–March 2013, covering various programme stages (Fig. 1). Although the programme started in 2011, the first cycle involved compiling baseline data with training on performance indicators occurring in the second half of 2011, and the initial round of performance payment being made in 2012.

In-depth interviews were carried out with health workers, managers at council level and national level and stakeholders. Group interviews were conducted with regional managers and health facility committee members from three government facilities. A total of 93 individual in-depth interviews (IDIs), 9 group interviews and 18 Focus Group Discussions (FGDs) were conducted by four social scientists working in pairs. The interviews were recorded digitally and subsequently transcribed and translated into English. Observations of performance feedback meetings and data verification activities were also done.

Quantitative data collection was done in January 2012 and thirteen months later. Health facility and health worker surveys were carried out before and after the implementation of P4P in all seven intervention districts in Pwani region and four comparison districts (Kilwa, Mvomero, Morogoro town and Morogoro rural). A total of 150 facilities, 75 in the intervention and 75 in the comparison group were sampled, representing 46% of all eligible facilities in Pwani and 34% of all facilities in the comparison districts. In each facility one or two health workers delivering reproductive and child health services picked at random from those on duty were also interviewed. Facilities were randomly sampled amongst those where P4P was implemented and matching comparison facilities were selected based on provider type, ownership, and case load (Borghi et al., 2013).

1.4. Data analysis

We used the Cleary et al. framework to define internal and external accountability. We then identified relevant themes within the qualitative data, and indicators within the surveys (Table 1). Verbatim transcriptions of qualitative data were first read to get an overall impression. A coding system was then developed and the data was managed and coded using NVivo 10 software.

The quantitative indicators measured are summarised in Table 1. We used a difference-in-difference linear regression model to isolate the effect of P4P on the outcomes of interest, as shown in Equation (1).

Equation 1

$$Y_{ijt} = \beta_0 + \beta_1 (P4P_j \times \delta_t) + \beta_2 \delta_t + \gamma_j + \varepsilon_{ijt} \quad (1)$$

In all models, we included facility fixed effects (γ_j) to control for facility-level unobserved time invariant characteristics, and year fixed effects (δ_t) a dummy variable taking the value of 0 at baseline and 1 at endline, with health worker outcomes clustered at the facility level. The effect of P4P on outcomes is estimated as β_1. We confirmed the robustness and precision of our results to: removing the facility fixed effects from the model; using non-linear (logit) models for binary outcomes; and, clustering standard errors at the district level (Cameron and Miller, 2015). To adjust for multiple outcome testing, we applied a Bonferroni correction which accounts for possible correlation between outcomes (Bonfrer et al., 2014a,b). The lagged dependent variable approach has been proposed as an alternative approach to difference-in-differences. It maximises statistical power and, when trends are not parallel, produces unbiased results. As we were unable to test whether the pre-intervention trends in the specific outcomes considered in this paper were parallel, we applied a lagged dependent variable approach as a further robustness check (McKenzie, 2012a,b; Ozler, 2015).

All analyses were carried out at the health facility level. To generate health facility values for indictors collected at the health worker level, the maximum value reported at a given facility was selected for supervision outcomes and mean scores were estimated for indicators of satisfaction with community relations across health workers in the same facility (McKenzie, 2012a,b).

2. Results

We present the P4P scheme's effects on internal and external accountability mechanisms. The findings are summarised in Table 4.

2.1. Internal accountability

2.1.1. Content and frequency of supervision

There was no effect of P4P on the number of supervision visits by managers. However, there was a reduction of 17% (SE: 7.1) in the

Fig. 1. Qualitative data collection in relation to programme implementation.

number of facilities reporting that supervision happened less than once per quarter (see Table 1) - the recommended frequency for supervision visits. Managers indicated that they could not increase the intensity of visits due to a lack of funds for fuel and allowances, and the council cars were often not available:

Supportive supervision is not done as it is supposed to be, because we have transport problems. We have one car and it is being used for many activities. You may plan to go for supervision, but in the end you realize that the only car has been assigned for a different activity (FGD with council managers, November 2012).

However, the interviews with managers revealed that they were very concerned about reaching targets, and health workers reported that managers were keen to supervise health workers, help facilities achieve their targets, and ensure that they provide correct and timely data. In all districts, health workers and managers worked together after the official working time, something that had rarely happened before:

During the first visit, they came here around 9 pm […] we worked with the regional and district managers until 12 midnight. If you wanted to leave they became very aggressive (FGD with health workers, October 2012)

As for the content of the supervision visits, some health workers felt that the supervision visits simply focused on collecting data/reports in the early phase. From the second quarter of 2012 onwards, however, health workers felt their managers were more pro-active in solving problems. The survey data revealed P4P had a positive and significant effect on the provision of both positive and negative feedback during supervision (+24.8% SE: 11.4 and + 28.3% SE 10.9 respectively), but there were no other effects noted (Table 2).

2.1.2. Financial accountability and autonomy

Health workers expressed satisfaction with the transparent processes surrounding the allocation of the bonus payments. This was facilitated by the score cards which indicated the total amount of bonus earned by the facility, and its distribution between health workers and the facility:

Everything is done in a transparent way. We know how much the facility gets, how much for staff … To be honest it is difficult to get such information for other (funding) sources (FGD with health workers, October 2012).

The fact that bonus funds were channelled directly to the facilities' bank accounts minimized the risk of misuse of funds by district level managers. The direct and transparent payment of funds, as well as more frequent contact with their managers was reported to enhance trust and improve the relationship between health workers and their managers. However, health workers complained that score cards were not updated every payment cycle to reflect the actual number of bonus beneficiaries. If new staff had been hired, the bonus for each individual would be lower than what the scorecard indicated.

Health workers at public facilities appreciated the autonomy they had in deciding how to use P4P funds for facility improvement. Two out of three facilities reported that they had used the funds to make the facilities more attractive and thus increase the chance of reaching targets.

While P4P had no effect on coverage of community based health insurance (the Community Health Fund) at the facility (see Table 3), health workers were able to retain some of the premiums collected in their bank accounts, increasing resource availability.

2.1.3. Resource prioritisation to meet targets

We found evidence of district and facility managers reallocating staff in order to meet targets as shortages of skilled and efficient staff were identified as the cause of poor performance among certain facilities. In one case the council health management team (CHMT) decided to transfer a clinical officer to a dispensary that was run by a nurse and had failed to submit the performance data. In another case, a facility manager requested additional staffing support from the district:

We needed someone to assist our nurse on RCH activities, so what I did was to request one nurse [from the CHMT] and they gave us one (Health worker, July 2012)

In other cases districts upgraded dispensaries to enable them to provide delivery care services and meet targets:

Some of our facilities were not providing delivery services […]. They [the PMT] decided to stop paying us [the bonus] until our dispensaries provide delivery services. We decided to make sure that every dispensary set aside a room for deliveries (District manager, July 2012)

In addition to a lack of qualified staff, many health facilities were struggling to ensure they had the drugs, supplies and equipment needed to meet targets. District managers reported that health workers were more likely to report stock outs and to expect support from district managers to address this:

Now, if they don't have vaccines or gas they perceive it as an emergency, and they will communicate it to you as an emergency. In the past, they did not care when they were out of gas, but now they know that if they don't have gas they will not be able to achieve their targets (District manager, July 2012)

Health workers also reported that managers were more responsive than before in addressing drug and supply constraints at the facility level:

Nowadays if there's a shortage of medicines you only have to make a call to the DRCHCo [District Reproductive and Child Health Coordinator] and immediately without a delay they are brought, since if s/he delays then s/he will also lose out [smiled] (Health worker, January 2012)

The survey data indicated P4P significantly reduced the stock out rate of drugs and medical supplies by 16.9% (SE: 5.8) and 15.2% (5.1) respectively and the interruption of service delivery due to broken equipment by 14.9% (7.3) (Table 3).

In most cases, the use of facility-level P4P funds was linked directly to the targets, like buying anti-malarials for pregnant women and paying traditional birth attendants (TBAs) to bring women for deliveries, installing solar power in maternity wards, and buying oxytocin. However, some health workers, particularly at hospitals, questioned their ability to reach targets and provide quality services due to the constraints in the system. They argued that basic medical supplies and equipment should have been in place before the introduction of P4P:

The work environment remained the same, and we have the same resources. We have been informed about the targets and we have to struggle to achieve the goals, but P4P did not bring any new

Table 2
Effect of P4P on human resource management.

	Baseline			Difference in difference	
	Intervention Mean (SD)	Comparison Mean (SD)	Difference (P-Value)	N	Fixed Effects Beta (SE)
Frequency of supervisions					
Number of district/regional supervision carried out	1.7	1.5	0.2	272	−0.1
	(2.8)	(1.8)	(0.64)		(0.5)
Last supervision received in the last 30 days (0–1) (%)	57.7	49.3	8.4	266	15.1
	(49.7)	(50.3)	(0.31)		(11.9)
Last supervision received in the last 31–90 days (0–1) (%)	31.0	41.3	−10.0	266	2.0
	(46.6)	(49.6)	(0.19)		(12.0)
Last supervision received more than 90 days (0–1) (%)	11.3	9.3	1.9	266	−17.1**
	(31.8)	(29.3)	(0.70)		(7.1)
Content of supervision from facility survey					
Check records/reports (0–1) (%)	51.7	50.8	0.9	271	−6.9
	(50.4)	(50.4)	(0.92)		(13.0)
Check drug supply (0–1) (%)	16.7**	33.8**	−17.0**	264	8.8
	(37.6)	(47.7)	(0.023)		(12.1)
Check service delivery (0–1) (%)	21.7*	35.4*	−14.0*	261	9.2
	(41.5)	(48.2)	(0.09)		(13.7)
Provide positive feed-back (0–1) (%)	10.0***	29.2***	−19***	265	24.8**
	(30.3)	(45.8)	(0.00)		(11.4)
Provide negative feed-back (0–1) (%)	8.3***	27.7***	−19***	268	28.3**
	(27.9)	(45.1)	(0.00)		(10.9)
Provide updates (0–1) (%)	18.3	21.5	−3.2	263	0.5
	(39.0)	(41.4)	(0.66)		(12.3)
Discuss problems (0–1) (%)	23.3	26.2	−2.8	264	−0.4
	(42.7)	(44.3)	(0.71)		(12.2)
Deliver supplies (0–1) (%)	8.3	3.1	5.3	260	−18.2**
	(27.9)	(17.4)	(0.21)		(8.2)
Content of supervision from health worker survey					
Bring drugs/supplies (0–1) (%)	17.1	17.3	−0.2	265	−14.9
	(38.0)	(38.1)	(0.97)		(10.9)
Check records (0–1) (%)	48.6	56.0	−7.4	265	−1.1
	(50.3)	(50.0)	(0.37)		(12.8)
Check finances (0–1) (%)	2.9	5.3	−2.5	265	8.8
	(16.8)	(22.6)	(0.45)		(6.2)
Observe consultation (0–1) (%)	4.3	10.7	−6.4	265	3.7
	(20.4)	(31.1)	(0.15)		(7.7)
Check knowledge (0–1) (%)	11.4	21.3	−9.9	265	10.5
	(32.0)	(41.2)	(0.11)		(10.4)
Instruct on service delivery (0–1) (%)	30.0	25.3	4.7	265	−13.9
	(0.462)	(43.8)	(0.53)		(10.9)
Instruct on filling HMIS (0–1) (%)	14.3	14.7	−0.4	265	−2.0
	(35.2)	(35.6)	(0.95)		(9.7)
Discuss performance (0–1) (%)	24.3	22.7	−1.6	265	1.5
	(43.2)	(42.1)	(0.82)		(10.9)
Inspect facility (0–1) (%)	15.7***	37.3***	−22.0***	265	12.1
	(36.7)	(48.7)	(0.00)		(11.4)
Do nothing (0–1) (%)	11.4*	4.0*	7.4*	265	−5.6
	(32.0)	(197)	(0.09)		(5.6)
Other					
No. of admin./managerial meetings in past 90 days	2.0	1.8	0.2	271	−0.1
	(1.1)	(1.2)	(0.30)		(0.4)
Health workers reporting selection for training based on need (0–1) (%)	23.5	21.2	2.4	276	−3.9
	(41.0)	(37.7)	(0.73)		(9.5)

* significant at 10%; ** significant at 5%; *** significant at 1%.
Sample: 150 health facility in two time periods.
Beta (SE) are coefficients for continuous variables and percentage changes for binary indicators and their means.

equipment. … I am still doing surgery in a room with no AC. (…) If there is no medicine I can't be blamed for not having played my role (Health worker, February 2012).

In two of the districts, health workers and managers expressed concern about the lack of adequate support from some of the district managers who were not eligible for P4P bonuses. In one case this was the District Executive Director (DED), in the other case it was a councillor. In both cases, these authority figures were criticized for not prioritising the health department - which eventually affected P4P implementation:

Our cars are under the control of the District Executive Director. This department (health) might not be able to implement its activities because the cars are being used by other departments at the council level. […] The transport problem affects us especially on the issue of data validation; we fail to do data validation on time (District manager, October 2012)

In response to this situation, the Regional Administrative Secretary (RAS) wrote a warning letter to both the DED and the district medical officer (DMO), instructing them to ensure availability of a car for data verification. Soon after, the DED reportedly released a

Table 3
Effect of P4P on financial accountability and resource prioritisation.

	Baseline			Difference in difference	
	Intervention Mean (SD)	Comparison Mean (SD)	Difference (P-value)	N	Fixed Effects Beta (SE)
Facility with functioning community health fund (CHF) (%)	79.5*** (40.7)	55.6*** (50.0)	24.0*** (0.00)	295	−2.9 (7.3)
Number of CHF members	22.7 (40.0)	14.7 (31.9)	8.01 (0.24)	217	8.5 (12.9)
Equipment functioning index [a] (0–1) (%)	56.7 (18.2)	54.8 (17.2)	1.9 (0.53)	295	3.2 (4.3)
Service delivery disruption due to broken equipment in last 90 days (%)	25.4** (43.8)	12.2** (32.9)	13.0** (0.04)	292	−14.9** (7.3)
Vaccines stock-out index (0–1) [b] (%)	17.1 (30.7)	12.9 (28.0)	4.2 (0.41)	276	−10.2* (5.6)
Drug stock-out index (0–1) [c] (%)	54.4* (23.5)	46.0* (27.8)	8.4* (0.05)	295	−16.9*** (5.8)
Medical supplies stock-out index [d] (0–1) (%)	39.4*** (25.3)	26.1*** (23.5)	0.13.0*** (0.00)	275	−15.2*** (5.1)

* significant at 10%; ** significant at 5%; *** significant at 1%.
Sample: 150 health facility in two time periods.
(a) Equipment includes: BP apparatus available at least one, Stethoscope apparatus, Time/watch, Infant/child weighing scale, MUAC measuring tape, Test kit for hemoglobin, Re-agents for test kit for hemoglobin, Neonatal ambu-bag & mask, Incubator, Autoclave equipment, Cord clamps apparatus, Infant laryngoscope, Mucus suction apparatus, Delivery kits, Delivery table, Vacuum extractor, Thermometer, Examination torch/lamp, Stainless steel bowls.
(b) Vaccine includes vaccine against Tetanus, BCG, OPV, DPT, Measles.
(c) Drugs includes: ALU-Blisters 24, ALU-Blisters 18, ALU-Blisters 12, ALU-Blisters 6, Quinine tablets, Quinine syrup, Quinine injection, SP [IPTp], Anti-malarial availability index, ALU-Blisters 24, ALU-Blisters 18, ALU-Blisters 12, ALU-Blisters 6, Quinine tablets, Quinine syrup, Quinine injection, SP [IPTp], Anti-malarial index,,Cotrimo-xazola tablets, Cotrimo-xazola syrup, Flagly tablets, Flagly injection, Gentamycin injection 20 mg, Gentamycin injection 80 mg, Ampiciline tablets, Ampiciline injection, Ampiciline syrup, Chloramphenical tablets, Chloramphenical injection, Chloramphenical syrup, X-pen injection, Antibiotics availability index,Cotrimo-xazola tablets, Cotrimo-xazola syrup, Flagly tablets, Flagly injection, Gentamycin injection 20 mg, Gentamycin injection 80 mg, Ampiciline tablets, Ampiciline injection, Ampiciline syrup, Chloramphenical tablets, Chloramphenical injection, Chloramphenical syrup, X-pen injection, Antibiotics index, Aldomet tablets, Hydralazine tablets, Hydralazine injection, Nifedipine tablets, Anti-hypertensive drugs availability index.
(d) Medical supplies include: Sterile latex gloves, Disinfection, Cotton wool, Malaria RDT, Glass slide malaria test, Partograph, Sutures, Urine catheters, Suction catheters, Oxygen supply, Gas supply.

car to CHMT.

2.1.4. Organizational culture and cooperation

The qualitative data suggested that health staff felt that relations with managers had improved, with the latter becoming more accessible and less hierarchical in their dealings with providers:

> *There are changes. It is not like in the previous days where they (CHMT members) used to be the real bosses; they were not listening, but rather directing you on what to do. … nowadays when they come you discuss with them, and they may even ask if there is any staff member who has a problem. (…) The DMO may even give you his contact details (Health facility in-charge, April 2012)*

Increased interactions between providers and their managers helped to improve trust between the two parties, facilitated communication, and contributed to establishing a good working relationship. In short, there was a sense of common goal. On the managers side they appreciated the efforts made by providers to meet targets:

> *P4P has created a good relationship between the CHMT and health facility staff in the district, to the extent that they (providers) do respond positively once we visit or tell them anything about data - they are ready and they do understand us. Now they are doing their best and the situation is different from before. (District manager, February 2013)*

District level managers also felt that the PMT was supportive to their needs, finding ways of solving the problems together, and overcoming bottlenecks rather than giving instructions and then leaving, as was the case with national managers in the past:

> *Before, something could take six months to reach (…) the Ministry. But now, (…) if it is something to do with MSD [Medical Store Department] for example - (…) they can call them straight away. In the end the problem that you discussed today has been solved the next day! So it has really helped to minimize bureaucracies (District manager, July 2012)*

At the dispensary level, health workers reported that P4P had entailed more collaboration, with tasks being shared between workers rather than having people assigned to specific activities, like vaccinations and filling in forms:

> *Before (P4P) the exercise of filling forms was done by the RCH nurse alone, but since P4P implementation started we work together. If she (the nurse) is not present at the RCH department the other staff will take care of it; we don't want to lose mothers who seek RCH services (FGD with health workers, 2012).*

At hospitals, on the other hand, the unequal distribution of bonus payments between RCH staff (60%) and non-RCH staff (30%) created tensions:

> *We depend on each other. If there is no doctor at RCH any doctor can support RCH work. There was a time we asked why RCH are paid more (FGD with CHMT, November 2012).*

There was no effect of P4P on the quantitative indicators considered (the number of administrative/managerial meetings held during the last 90 days; the allocation of training opportunities according to need) (Table 2).

Table 4
Overview of P4P effects on accountability.

Expected effects of P4P (theory of change)	Evidence of positive effects	No, limited, or negative effects and identified obstacles
Internal Accountability		
Frequency of supervision	Less facilities report last supervision received more than 90 days ago	No effect on number of district/regional supervision visits carried out Lack of resources (vehicles and funds for allowances)
Content of supportive supervision	Provide positive and negative feedback	No effects on other indicators for content of supervision Emphasis on data collection, not on quality of services
Financial accountability and autonomy	Score cards enhanced trust Used P4P funds to increase chances of reaching targets	Limited financial autonomy over other funds No effect on number of facilities with functioning Community Health Fund (CHF) No effect on number of CHF members No effect on equipment functioning
Resource prioritisation to meet targets	Staff reallocated within the district Reduced stock-out rates Reduced service delivery disruption due to broken equipment	Poor performing staff transferred, not fired District managers outside the health sector may not prioritise using resources on P4P related activities
Organizational culture and cooperation	Less hierarchical Less bureaucratic Teamwork spirit enhanced HW exert pressure on management Management more responsive to facility requests	No effect on number of administrative/managerial meetings No effect on health workers reporting selection for training based on need At hospital level RCH staff was prioritised, and this was seen as unfair by other staff
External accountability		
Enhancing provider responsiveness to users and improve relationship with local community	Kindness to women during deliveries enhanced	No effect on patient experience of interpersonal care for the other targeted services No effect on HW satisfaction with relationship with local leaders No effect on number of facilities having outreach services No effect on number of outreach visits Lack funds for fuel and allowances
Role of Health Facility Governing Committees enhanced	Higher probability of holding meetings	No effect on number of facilities with committee No effect on record keeping Role limited to approval of plans, lack knowledge/confidence, not invited No funds for transport/meeting allowance

2.2. External accountability measures

2.2.1. Responsiveness to clients and community relations

In order to achieve targets, health workers reported changing their attitudes towards clients:

> *Currently pregnant women are enticed to come to deliver at the health facility to the extent that they are surprised. This is different from the situation in the past when they were given harsh words. Now health workers use polite language and this is a result of P4P […] (Health worker, January 2012)*

The survey data confirmed that there was a 0.38 point increase in the mean provider kindness score during delivery (95% CI: −0.06 to 0.80), although this was not significant at $p < 0.05$ level (Binyaruka et al., 2015). There was no P4P effect on provider relations with community leaders. Outreach services were not targeted by P4P, but health workers identified outreach activities as a mechanism for increasing utilization and therefore something that could be indirectly affected by P4P. However, our informants explained that outreach activities could not be performed due to a lack of funds:

> *Another challenge is that (…) staff fails to do mobile outreach services to offer vaccines. … Due to transport problems staff do not go for outreach (CHMT member, December 2011).*

The quantitative analysis also found no effect on the number of facilities having outreach services or on the number of outreach services performed (Table 5).

2.2.2. Involvement of health facility governing committees

Although Health Facility Governing Committees (HFGCs) were to be involved in the planning of the use of the health facility bonus, community members within the HFGCs had not been trained and were not incentivised, and at most facilities the procedure was not followed:

> *Usually the staff meeting decides how much money we want to use and what we want to buy. This is the procedure used […]. [Then we] leave it to the committee to approve it (Health worker, October 2012).*

Community members on the committee argued that they failed to participate in the discussion of committee issues because they perceived them as being scholarly/technical and decisions were left to the providers. Committee members were sometimes called to confirm the receipt of supplies purchased by P4P money, but they did not feel well informed about the purchases:

> *We check the drugs, but what I usually ask myself is this: 'What about the remaining drugs, where are they?' (FGD with HFGC members, March 2013).*

However, anecdotal evidence shows that in 2015, some health workers used their own P4P bonus to pay committee members an allowance for attending meetings and to travel to the bank for withdrawal of P4P funds. Survey data showed that P4P was associated with a positive and significant increase in the probability of having held a HFGC meeting in the last 90 days (+18%, SE: 9.1) (though this is not robust to sensitivity analysis (Annex Table 3A),

Table 5
Effect of P4P on external accountability indicators.

	Baseline			Difference in difference	
	Intervention Mean	Comparison Mean	T-test	N	Fixed Effects Beta
	(SD)	(SD)	(P-value)		(SE)
Governing committee					
Facility with governing committee (0–1) (%)	73.4	70.0	3.4	283	2.0
	(44.5)	(46.2)	(0.66)		(10.5)
Governing committee met in the past 90 days (0–1) (%)	94.4	93.2	1.3	291	18.2**
	(23.1)	(25.4)	(0.75)		(9.1)
Records of governing committee meeting available (0–1) (%)	92.5**	80.3**	12.2**	250	−6.0
	(26.5)	(40.1)	(0.04)		(8.3)
Outreach services					
Facility has outreach services (0–1) (%)	60.3	58.3	2.0	295	12.3
	(49.3)	(49.6)	(0.81)		(9.0)
Number of outreach visits in past 90 days	2.0	2.0	−0.1	295	52.1
	(2.8)	(2.3)	(0.87)		(61.7)
HW satisfaction local relationship					
Mean HW satisfaction with safety in community (0–1) (%)	59.2	57.8	1.4	291	17.3*
	(45.8)	(45.9)	(0.86)		(9.8)
Mean HW satisfaction with relationship with local leaders (0–1) (%)	69.0*	57.1*	11.9*	291	−11.9
	(40.9)	(45.6)	(0.10)		(9.7)

* significant at 10%; ** significant at 5%; *** significant at 1%.
Sample: 150 health facility in two time periods.
Beta (SE) are coefficients for continuous variables and percentage changes for binary indicators and their means.

but not on their record keeping. While the overall role of the committees was limited in relation to P4P, there was one case where the committee members were able to track the misuse of TSh. 614,000 bonus funds for facility improvement and action against the responsible health worker was taken. In this case, a ward councillor led the process, a factor that may explain the committee's success.

3. Discussion

The theory of change of P4P suggests that such schemes will have a positive effect on supervision, and that this is particularly important in contexts where most primary facilities in rural areas are staffed by lower grade staff (Meessen et al., 2006). P4P was indeed found to increase the timeliness of supervision and had a positive effect on the provision of feedback, particularly in relation to data verification. These changes are clearly linked to the teamwork spirit that P4P enhanced. The increased supervision by district managers seemed to be in part linked to their role in data verification.

Recent WHO recommendations state that the autonomy of providers "is a critical prerequisite" for P4P programs to be successful (WHO, 2016). We found evidence that health workers prioritised the use of P4P bonus payments for strategies that would help them to meet targets, and that this was facilitated by greater financial autonomy linked to P4P, a finding reported elsewhere (Meessen et al., 2011). Indeed, there was a significant reduction in the stock out rate of drugs and medical supplies, and reduction in service disruption due to broken equipment. In the Democratic Republic of Congo (DRC), in contrast, P4P had a negative effect on the availability of equipment. This was because the facilities reduced user fees in order to attract more clients, but did not succeed in this. Their income was thus reduced (Huillery and Seban, 2014).

As for organizational culture, we found evidence from the qualitative data that P4P helped improve communication and interpersonal relations between health workers and their managers; though we had less evidence of this from the limited quantitative data. Knowing that their own bonus and their district's rating compared to other districts depended on the performance of health workers, managers at the district level treated health workers in a less authoritarian way. This stands in contrast to a study from Benin where the authors found that PBF "does not seem to foster collaboration and teamwork between levels of the health system under the World Bank model, probably because it relies mostly on external actors" (Paul et al., 2014: 212).

In addition to supervision, the improved collaboration also helped in the reduction of stock-outs of drugs and medical supplies and health workers reportedly pushed their managers to deliver. The effects on the stock out of medical supplies and drugs were similar to that reported previously (Binyaruka and Borghi, 2017; Anselmi et al. 2017). Differences in the size of the coefficient are due to a difference in the classification of drugs and supplies in the former study, and a difference in the analytical approach in the latter which estimated mediators at the household level and included household level covariates in the regression analysis. District managers reported more responsiveness from the national P4P team compared to what they were used to from national level managers. District managers on their side were also pro-active in addressing issues they identified as barriers to meeting targets, including the reallocation of staff to meet facility needs, and supporting facilities to provide delivery care services where these services were not available. Unlike in Rwanda, where facilities had greater autonomy than in Tanzania (Meessen et al., 2011), providers did not have the power to hire and fire staff, but they were found to engage with district managers about human resource issues. Moreover, in contrast to the DRC (Huillery and Seban, 2014), where the facility head could decide on the payment distribution among health workers, and Nigeria (Bhatnagar and George, 2016), where the bonus payments were individualized based on performance, the bonus was equal for all health staff at primary facilities in Tanzania, something that

contributed to a feeling of fairness.

Improved cooperation between different levels within the system was clearly driven by having a shared goal: reaching P4P targets. In two of the five districts, managers at the district level who were neither eligible for bonuses nor had specific performance indicators (a District Executive Director and a district councillor), did not prioritise spending resources on P4P, despite the fact that the DEDs were supposed to sign the P4P performance agreement (MoHSW, 2012). This demonstrates that there is a need to examine the lines of accountability within the local context when setting incentives and deciding who to incentivise. In this case, incentivising DEDs may have resulted in greater cooperation in sharing needed resources for the implementation of P4P. However, one would also risk that incentives to DEDs would make them prioritise P4P activities at the cost of activities in other sectors.

P4P is expected to have a positive effect on external accountability through services that are more responsive to patient needs. The reported increase in provider kindness during deliveries (Binyaruka et al., 2015), provides some evidence of this, and is likely to reduce the number of home births (Kruk et al., 2014). As in Rwanda (Kalk et al., 2010), health workers started to see their patients as clients that should be treated well.

In contrast to the DRC and Rwanda (Huillery and Seban, 2014; Renmans et al., 2016), there was no effect on outreach services, which are important for utilization and access for the poor in remote communities. Health workers and managers argued that although they wanted to conduct outreach, there was a lack of resources and the facility bonus was not large enough to facilitate such services. This finding confirms the findings of other studies which have shown that in the design of P4P schemes, there are often unrealistic expectations of what institutions can actually do, and an underestimation of constraints (Ireland et al., 2011:695; Ssengooba et al., 2012).

The P4P design was based on the assumption that giving HFGCs a role in how the facility bonus was to be spent, and in the withdrawal of the funds, would encourage them to be active. We found that P4P enhanced committees' potential of holding meetings, but the role of the HFGC members was generally limited to approval of decisions and budgets that had been made by the health workers. HFGCs limited involvement may also have been due to a lack of explicit incentive to community members within the committee. This is an important difference between the Tanzanian P4P scheme design and the Burundian one (Falisse et al., 2012). Committees also lacked funds to organize meetings. Last but not least, the power imbalance between the educated health workers and the committee members is very high, and HFGCs in Tanzania have been found to have great problems challenging health workers (Wales et al., 2014). As Fox et al. have pointed out, enhancing the level of information is not enough for social accountability measures to be successful - grassroots stakeholders also need to have 'teeth' (Fox, 2015). In the one case where a HFGC had taken disciplinary actions against a clinical officer in-charge who had misspent the P4P funds, the committee received support from a ward councillor.

Our study suffers from some limitations. Our measure of external accountability is limited to the two indicators that can be feasibly measured with the data available. Although we did conduct interviews with HFGCs that included community members, we did not conduct interviews with a wider set of community members to explore their perceptions of accountability. Hence our assessment of external accountability is inevitably narrower than that of internal accountability. Neither do we have data showing the relative size of P4P bonuses compared to other sources of income. Moreover, the surveys included a large number of questions, which may have resulted in respondent fatigue. For the quantitative indicators of accountability, there was variation in the number of observations available for different indicators, with data incompleteness being greatest for the number of CHF members. This limits the generalisability of some of the indicators across the sample. The difference-in −difference design used for the quantitative analysis relies on the assumption of parallel trends in outcomes in intervention and comparison areas. Although trends for the outcome variables considered in this analysis could not be tested due to a lack of data on pre-intervention trends, the pre-intervention trends in facility level outpatient visits and other services were parallel. The results obtained using the lagged dependent variable approach were generally similar to those from the difference in differences analysis. However, the effects of P4P on five of the outcomes were no longer statistically significant (discussion of problems and delivery of supplies during the supervision, service delivery disruption due to broken equipment in the last 90 days and vaccine and medical supplies stock out) (Appendix Table 4A).

4. Conclusion

The P4P pilot in Tanzania contributed to some improvement in *internal accountability* measures such as timeliness of supervision and provision of positive and negative feedback. The active involvement of the PMT, and the presence of shared goals between managers and providers, appears to have played a central role for these improvements. The improved relations between managers and providers, and greater teamwork, coupled with enhanced provider autonomy over funds, entailed an improved handling of systemic challenges like staff availability and lack of medicines and supplies. P4P had more limited effect on *external accountability*. Though attitudes towards patients appeared to improve, in general community engagement through health facility governing committees remained limited. Implementers should examine the lines of accountability within the local context when setting incentives and deciding who to incentivise in P4P schemes.

Acknowledgments

The Government of Norway funded the data collection for the programme evaluation that was used in this paper. The UK Department for International Development as part of the Consortium for Research on Resilient and Responsive Health Systems supported the funding of the authors' time undertaking data re-analysis and writing. The Research Council of Norway also supported the time of IM, SL, MM and JB. The funding bodies had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript. Thanks are due to Irene Mashasi, Ikunda Njau, and Albert Majura who participated in the qualitative data collection, as well as the surveyors taking part in quantitative data collection.

Appendices

Table 1A
Effect of P4P on human resource management indicators, sensitivity analysis.

	OLS (Facility SE clustering)		OLS (Facility fixed effects)		OLS (Facility fixed effects and standard errors clustered at the district level)		Logit marginal effects (Facility fixed effects)	
	N	Beta (P-value)	N	Beta (P-value)	N	Beta (P-value)	N	Beta (P-value)
Frequency of supervision								
Number of district/regional supervision carried out	272	−0.097 (0.847)	272	−0.1 (0.807)	272	−0.1 (0.775)		
Last supervision received in the last 30 days (0–1) (%)	266	4.92 (0.666)	266	15.1 (0.209)	266	15.1 (0.230)	98	16.7 (0.233)
Last supervision received in the last 31–90 days (0–1) (%)	266	10.3 (0.365)	266	2.0 (0.867)	266	2.0 (0.844)	98	3.6 (0.803)
Last supervision received more than 90 days (0–1) (%)	266	−15.3** (0.030)	266	−17.1** (0.018)	266	−17.1 (0.514)	36	−369 (0.995)
Content of supervision from facility surveys								
Check records/reports (0–1) (%)	271	0.90 (0.942)	271	−6.9 (0.600)	271	−6.9 (0.550)	126	−4.8 (0.708)
Check drug supply (0–1) (%)	264	13.3 (0.242)	264	8.8 (0.469)	264	8.8 (0.741)	122	−2.0 (0.878)
Check service delivery (0–1) (%)	261	17.0 (0.177)	261	9.2 (0.504)	261	9.2 (0.494)	130	8.3 (0.505)
Provide positive feed-back (0–1) (%)	265	29.4*** (0.006)	265	24.8** (0.031)	265	24.8 (0.185)	88	33.6** (0.016)
Provide negative feed-back (0–1) (%)	268	31.2*** (0.002)	268	28.3** (0.010)	268	28.3* (0.080)	88	38.6*** (0.005)
Provide updates (0–1) (%)	263	3.82 (0.737)	263	0.5 (0.967)	263	0.5 (0.984)	108	3.3 (0.811)
Discuss problems (0–1) (%)	264	3.59 (0.754)	264	−0.4 (0.976)	264	−0.4 (0.984)	110	−7.4 (0.588)
Deliver supplies (0–1) (%)	260	−12.7* (0.093)	260	−18.2** (0.028)	260	−18.2 (0.497)	48	−40.3* (0.058)
Content of supervision from health worker survey								
Bring drugs/supplies (0–1) (%)	265	−11.5 (0.264)	265	−14.9 (0.175)	265	−14.9 (0.572)	82	−18.2 (0.230)
Check records (0–1) (%)	265	4.10 (0.733)	265	−1.1 (0.933)	265	−1.1 (0.968)	110	0.0 (1.000)
Check finances (0–1) (%)	265	5.81 (0.316)	265	8.8 (0.158)	265	8.8 (0.738)	26	42.3* (0.096)
Observe consultation (0–1) (%)	265	3.05 (0.669)	265	3.7 (0.635)	265	3.7 (0.515)	40	12.4 (0.578)
Check knowledge (0–1) (%)	265	9.90 (0.312)	265	10.5 (0.315)	265	10.5 (0.385)	72	17.5 (0.284)
Instruct on service delivery (0–1) (%)	265	−16.3 (0.114)	265	−13.9 (0.205)	265	−13.9 (0.165)	80	−21.2 (16.8)
Instruct on filling HMIS (0–1) (%)	265	−1.29 (0.882)	265	−2.0 (0.834)	265	−2.0 (0.939)	62	−0.061 (0.738)
Discuss performance (0–1) (%)	265	0.048 (0.996)	265	1.5 (0.887)	265	1.5 (0.953)	78	2.4 (0.882)
Inspect facility (0–1) (%)	265	21.6** (0.043)	265	12.1 (0.287)	265	12.1 (0.644)	86	17.1 (0.257)
Do nothing (0–1) (%)	265	−5.76 (0.249)	265	−5.6 (0.323)	265	−5.6 (0.832)	22	−13.7 (0.658)
Other								
No. of admin./managerial meetings in past 90 days	271	−0.017 (0.957)	271	−0.1 (0.701)	271	−0.1 (0.755)		
Health workers reporting selection for training based on need (0–1) (%)	276	−7.03 (0.451)	276	−3.9 (0.683)	276	−3.9 (0.620)		

* significant at 10%; ** significant at 5%; *** significant at 1%.
Beta are coefficients for continuous variables and percentage changes for binary indicators and their means.
P-values in parentheses.
Bonferroni adjustment accounting for intra-outcomes correlation: p-value threshold for joint significance of the family of outcomes at 5%: 0.002474.
Calculates based on 22 internal accountability non aggregated indicators, 274 pairwise correlations excluding diagonal, Average: 0.054079487.
Marginal effects reported for logit.
Sample: 150 health facility in two time periods.

Table 2A
Effect of P4P on financial accountability and resource prioritisation, sensitivity analysis.

OLS (Facility SE clustering)		OLS (Facility fixed effects)		OLS (Facility fixed effects and standard errors clustered at the district level)		Logit – Marginal effects (Facility fixed effects)	
N	Beta (P-value)	N	Beta (P-value)	N	Beta (P-value)	N	Beta (P-value)
Facility with functioning CHF (0–1) (%)							
295	−2.6 (0.729)	295	−2.9 (0.695)	295	−2.9 (0.903)	56	3.6 (0.856)
Number of CHF members							
217	−16.8 (0.194)	217	8.5 (0.514)	217	8.5*** (0.000)		
Equipment functioning index (0–1) (%)							
295	2.83 (0.512)	295	3.2 (0.459)	295	3.2 (0.325)		
Service delivery disruption due to broken equipment in last 90 days (%)							
292	−14.3* (0.051)	292	−14.9** (0.044)	292	−14.9 (0.130)	60	−18.0 (0.331)
Vaccines stock-out index (0–1) (%)							
276	−6.66 (0.257)	276	−10.2* (0.069)	276	−10.2 (0.685)		
Drug stock-out index (0–1) (%)							
295	−16.2*** (0.006)	295	−16.9*** (0.004)	295	−16.9 (0.150)		
Medical supplies stock-out index (0–1) (%)							
275	−14.6*** (0.005)	275	−15.2*** (0.004)	275	−15.2** (0.015)		

* significant at 10%; ** significant at 5%; *** significant at 1%.
Beta are coefficients for continuous variables and percentage changes for binary indicators and their means.
P-values in parentheses.
Bonferroni adjustment accounting for intra-outcomes correlation: p-value threshold for joint significance of the family of outcomes at 5%: 0.0074331.
Calculates based on 7 financing non aggregated indicators, 21 pairwise correlations excluding diagonal, Average: 0.020471429.
Marginal effects reported for logit.
Sample: 150 health facility in two time periods.

Table 3A
Effect of P4P on external accountability indicators, sensitivity analysis.

External accountability indicators	OLS (Facility SE clustering)		OLS (Facility fixed effects)		OLS (Facility fixed effects and standard errors clustered at the district level)		Logit (Facility fixed effects)	
	N	Beta (P-value)	N	Beta (P-value)	N	Beta (P-value)	N	Beta (P-value)
Facility with governing committee (0–1) (%)	283	3.57 (0.727)	283	2.0 (0.852)	283	2.0 (0.936)	98	12.0 (0.934)
Governing committee met in the past 90 days (0–1) (%)	291	18.0** (0.048)	291	18.2** (0.048)	291	18.2 (0.446)	128	7.4 (0.646)
Records of governing committee meeting available (0–1) (%)	250	−14.6* (0.075)	250	−6.0 (0.470)	250	−6.0 (0.828)	38	−13.7 (0.546)
Facility has outreach services (0–1) (%)	295	11.4 (0.204)	295	12.3 (0.172)	295	12.3 (0.600)	86	18.1 (0.222)
Number of outreach visits in past 90 days	295	49.6 (0.413)	295	52.1 (0.400)	295	52.1 (0.027)		
Mean HW satisfaction with safety in community (0–1) (%)	291	14.9 (0.124)	291	17.3* (0.079)	291	17.3 (0.470)		
Mean HW satisfaction with relationship with local leaders (0–1) (%)	291	−11.5 (0.226)	291	−11.9 (0.219)	291	−11.9 (0.330)		

* significant at 10%; ** significant at 5%; *** significant at 1%.
Beta are coefficients for continuous variables and percentage changes for binary indicators and their means.
P-values in parentheses.
Bonferroni adjustment accounting for intra-outcomes correlation: p-value threshold for joint significance of the family of outcomes at 5%: 0.0081706.
Calculates based on 7 external accountability non aggregated indicators, 21 pairwise correlations excluding diagonal, Average: 0.069085714.
Sample: 150 health facility in two time periods.
Marginal effects reported for logit.
(a) number of outreach visits not included in mean calculation).

Table 4A
Effect of P4P using a lagged dependent variable approach, sensitivity analysis.

	LDV Beta (SE)	N
Human resource management		
Frequency of supervision		
Number of district/regional supervision carried out	0.091	122
	(0.291)	
Last supervision received in the last 30 days (0–1) (%)	13.191	117
	(9.161)	
Last supervision received in the last 31–90 days (0–1) (%)	0.639	117
	(9.175)	
Last supervision received more than 90 days (0–1) (%)	−13.371***	117
	(5.099)	
Content of supervision from facility surveys		
Check records/reports (0–1) (%)	−3.866	121
	(8.759)	
Check drug supply (0–1) (%)	−7.601	116
	(9.389)	
Check service delivery (0–1) (%)	−5.191	114
	(9.506)	
Provide positive feed-back (0–1) (%)	5.177	116
	(8.146)	
Provide negative feed-back (0–1) (%)	9.448	119
	(8.542)	
Provide updates (0–1) (%)	−2.636	115
	(9.356)	
Discuss problems (0–1) (%)	−1.835	116
	(9.445)	
Deliver supplies (0–1) (%)	−11.438	113
	(7.633)	
Content of supervision from health worker survey		
Bring drugs/supplies (0–1) (%)	−10.180	116
	(8.290)	
Check records (0–1) (%)	−4.181	116
	(9.083)	
Check finances (0–1) (%)	3.629	116
	(4.785)	
Observe consultation (0–1) (%)	−2.098	116
	(6.151)	
Check knowledge (0–1) (%)	−1.501	116
	(7.300)	
Instruct on service delivery (0–1) (%)	−12.216	116
	(8.237)	
Instruct on filling HMIS (0–1) (%)	−0.791	116
	(6.312)	
Discuss performance (0–1) (%)	−0.157	116
	(7.734)	
Inspect facility (0–1) (%)	0.663	116
	(8.402)	
Do nothing (0–1) (%)	1.320	116
	(2.981)	
Other		
No. of admin./managerial meetings in past 90 days	0.076	122
	(0.288)	
Health workers reporting selection for training based on need (0–1) (%)	−2.316	128
	(7.810)	
Facility with functioning CHF (0–1) (%)	9.628	145
	(6.313)	
Number of CHF members	10.663	76
	(11.272)	
Equipment functioning index (0–1) (%)	4.504	145
	(3.861)	
Service delivery disruption due to broken equipment in last 90 days (%)	−1.845	143
	(4.152)	
Vaccines stock-out index (0–1) (%)	−4.493	127
	(3.024)	
Drug stock-out index (0–1) (%)	−9.954**	145
	(4.645)	
Medical supplies stock-out index (0–1) (%)	−2.935	125
	(3.294)	
External accountability indicators		
Facility with governing committee (0–1) (%)	5.257	134
	(7.359)	
Governing committee met in the past 90 days (0–1) (%)	17.947**	141
	(8.337)	
Records of governing committee meeting available (0–1) (%)	1.453	106
	(5.220)	

Table 4A (*continued*)

	LDV Beta (SE)	N
Facility has outreach services (0–1) (%)	13.573*	145
	(7.313)	
Number of outreach visits in past 90 days	46.842	145
	(52.885)	
Mean HW satisfaction with safety in community (0–1) (%)	17.139**	141
	(7.250)	
Mean HW satisfaction with relationship with local leaders (0–1) (%)	−1.747	141
	(7.205)	

* significant at 10%; ** significant at 5%; *** significant at 1%.
Beta are coefficients for continuous variables and percentage changes for binary indicators and their means.
Standard Errors-values in parentheses.

References

Anselmi, L., Binyaruka, P., Borghi, J., 2017. Understanding causal pathways within health systems policy evaluation through mediation analysis: an application to payment for performance (P4P) in Tanzania. Implement Sci. 12 (1), 10.

Basinga, P., Gertler, P.J., Binagwaho, A., Soucat, A.L.B., Sturdy, J., Vermeersch, C.M.J., 2011. Effect on maternal and child health services in Rwanda of payment to primary health-care providers for performance: an impact evaluation. Lancet 377 (9775), 1421–1428.

Bertone, M.P., Lagarde, M., Witter, S., 2016. Performance-based financing in the context of the complex remuneration of health workers: findings from a mixed-method study in rural Sierra Leone. Bmc Health Serv. Res. 16 (1), 1–10. http://dx.doi.org/10.1186/s12913-016-1546-8.

Bertone, M.P., Meessen, B., 2013. Studying the link between institutions and health system performance: a framework and an illustration with the analysis of two performance-based financing schemes in Burundi. Health Policy Plan. 28 (8), 847–857. http://dx.doi.org/10.1093/heapol/czs124.

Bhatnagar, A., George, A.S., 2016. Motivating health workers up to a limit: partial effects of performance-based financing on working environments in Nigeria. Health Policy Plan. 31 (7), 868–877. http://dx.doi.org/10.1093/heapol/czw002.

Binyaruka, P., Borghi, J., 2017. Improving quality of care through payment for performance: examining effects on the availability and stock-out of essential medical commodities in Tanzania. Trop. Med. Int. Health 22 (1), 92–102.

Binyaruka, P., Patouillard, E., Powell-Jackson, T., Greco, G., Maestad, O., Borghi, J., 2015. Effect of paying for performance on utilisation, quality, and user costs of health services in Tanzania: a controlled before and after study. Plos One 1–16.

Bonfrer, I., Soeters, R., Van de Poel, E., Basenya, O., Longin, G., van de Looij, F., van Doorslaer, E., 2014a. Introduction of performance-based financing in Burundi was associated with improvements in care and quality. Health Aff. 33 (12), 2179–2187.

Bonfrer, I., Van de Poel, E., Van Doorslaer, E., 2014b. The effects of performance incentives on the utilization and quality of maternal and child care in Burundi. Soc. Sci. Med. 123 (0), 96–104. http://dx.doi.org/10.1016/j.socscimed.2014.11.004.

Borghi, J., Mayumana, I., Mashasi, I., Binyaruka, P., Patouillard, E., Njau, I., Mamdani, M., 2013. Protocol for the evaluation of a pay for performance programme in Pwani region in Tanzania: a controlled before and after study. Implement. Sci. 8 (80).

Cameron, A.C., Miller, D.L., 2015. A Practitioner's guide to cluster-robust inference. J. Hum. Resour. 50 (2), 317–372.

Cleary, S.M., Molyneux, S., Gilson, L., 2013. Resources, attitudes and culture: an understanding of the factors that influence the functioning of accountability mechanisms in primary health care settings. Bmc Health Serv. Res. 13 (1), 1–11. http://dx.doi.org/10.1186/1472-6963-13-320.

Falisse, J.-B., Meessen, B., Ndayishimiye, J., Bossuyt, M., 2012. Community participation and voice mechanisms under performance-based financing schemes in Burundi. Trop. Med. Int. Health 17 (5), 674–682. http://dx.doi.org/10.1111/j.1365-3156.2012.02973.x.

Fox, J.A., 2015. Social accountability: what does the evidence really say? World Dev. 72, 346–361. http://dx.doi.org/10.1016/j.worlddev.2015.03.011.

Huillery, E., Seban, J., 2014. Pay-for-Performance, Motivation and Final Output in the Health Sector: Experimental Evidence from the Democratic Republic of Congo.

Ireland, M., Paulb, E., Dujardina, B., 2011. Can performance-based financing be used to reform health systems in developing countries? Bull. World Health Organ 89, 695–698.

Janssen, W., de Dieu Ngirabega, J., Matungwa, M., Van Bastelaere, S., 2015. Improving quality through performance-based financing in district hospitals in Rwanda between 2006 and 2010: a 5-year experience. Trop. Dr. 45 (1), 27–35. http://dx.doi.org/10.1177/0049475514554481.

Kalk, A., Paul, F.A., Grabosch, E., 2010. 'Paying for performance' in Rwanda: does it pay off? Trop. Med. Int. Health 15 (2), 182–190. http://dx.doi.org/10.1111/j.1365-3156.2009.02430.x.

Kruk, M.E., Kujawski, S., Mbaruku, G., Ramsey, K., Moyo, W., Freedman, L.P., 2014. Disrespectful and abusive treatment during facility delivery in Tanzania: a facility and community survey. Health Policy Plan. 1–8. http://dx.doi.org/10.1093/heapol/czu079.

Lohmann, J., Houlfort, N., De Allegri, M., 2016. Crowding out or no crowding out? A Self-Determination Theory approach to health worker motivation in performance-based financing. Soc. Sci. Med. 169, 1–8. http://dx.doi.org/10.1016/j.socscimed.2016.09.006.

McCoy, D.C., Hall, J.A., Ridge, M., 2012. A systematic review of the literature for evidence on health facility committees in low- and middle-income countries. Health Policy Plan. 27 (6), 449–466. http://dx.doi.org/10.1093/heapol/czr077.

McKenzie, D., 2012a. Beyond baseline and follow-up: the case for more T in experiments. J. Dev. Econ. 99 (2), 210–221.

McKenzie, D., 2012b. Tools of the Trade: a Quick Adjustment for Multiple Hypothesis Testing. http://blogs.worldbank.org/impactevaluations/tools-of-the-trade-a-quick-adjustment-for-multiple-hypothesis-testing. Retrieved 14 Oct 2015.

Meessen, B., Kashala, J.-P.I., Musango, L., 2007. Output-based payment to boost staff productivity in public health centres: contracting in Kabutare district, Rwanda. Bull. World Health Organ. 85, 108–115.

Meessen, B., Musango, L., Kashala, J.-P.I., Lemlin, J., 2006. Reviewing institutions of rural health centres: the Performance Initiative in Butare, Rwanda. Trop. Med. Int. Health 11 (8), 1303–1317. http://dx.doi.org/10.1111/j.1365-3156.2006.01680.x.

Meessen, B., Soucat, A., Sekabaraga, C., 2011. Performance-based financing: just a donor fad or a catalyst towards comprehensive health-care reform? Bull. World Health Organ. 89, 153–156.

MoHSW, 2012. Pwani P4P Pilot Design Document. Dar Es Salaam. Ministry of Health and Social Welfare, United Republic of Tanzania.

Njuki, R., Okal, J., Warren, C.E., Obare, F., Abuya, T., Kanya, L., Askew, I., 2012. Exploring the effectiveness of the output-based aid voucher program to increase uptake of gender-based violence recovery services in Kenya: a qualitative evaluation. [journal article]. Bmc Public Health 12 (1), 1–8. http://dx.doi.org/10.1186/1471-2458-12-426.

Ozler, B., 2015. Why is difference-in-difference estimation still so popular in experimental analysis? World bank Blog at. http://blogs.worldbank.org/impactevaluations/why-difference-difference-estimation-still-so-popular-experimental-analysis.

Paul, E., Sossouhounto, N., Eclou, D.S., 2014. Local stakeholders' perceptions about the introduction of performance-based financing in Benin: a case study in two health districts. Int. J. Health Policy Manag. 3 (4), 207–214. Available at: SSRN. https://ssrn.com/abstract=2502257.

Renmans, D., Holvoet, N., Orach, C.G., Criel, B., 2016. Opening the 'black box' of performance-based financing in low- and lower middle-income countries: a review of the literature. Health Policy Plan. 31 (9), 1297–1309. http://dx.doi.org/10.1093/heapol/czw045.

Ssengooba, F., McPake, B., Palmer, N., 2012. Why performance-based contracting failed in Uganda – an "open-box" evaluation of a complex health system intervention. Soc. Sci. Med. 75 (2), 377–383. http://dx.doi.org/10.1016/j.socscimed.2012.02.050.

URT, 2001. Council Health Service Board. Dar Es Salaam. Ministry of Regional Administration and Local Government, United Republic of Tanzania.

Wales, J., Tobias, J., Malangalila, E., Swai, G., Wild, L., 2014. Stock-outs of Essential Medicines in Tanzania. A Political Economy Approach to Analysing Problems and Identifying Solutions. Overseas Development Institute, London.

WHO, 2016. Main Lessons on Performance-based Financing (PBF) Programmes to Date. http://who.int/health_financing/topics/performance-based-financing/lessons/en/. Retrieved 24 February, 2016.

Witter, S., Toonen, J., Meessen, B., Kagubare, J., Fritsche, G., Vaughan, K., 2013. Performance-based financing as a health system reform: mapping the key dimensions for monitoring and evaluation. Bmc Health Serv. Res. 13 (1), 367.