Tropical Medicine and International Health doi:10.1111/tmi.12809

VOLUME 22 NO 1 PP 92–102 JANUARY 2017

Improving quality of care through payment for performance: examining effects on the availability and stock-out of essential medical commodities in Tanzania

Peter Binyaruka[1,2] and Josephine Borghi[3]

1 *Ifakara Health Institute, Dar es Salaam, Tanzania*
2 *Centre for International Health, University of Bergen, Bergen, Norway*
3 *Department of Global Health and Development, London School of Hygiene & Tropical Medicine, London, UK*

Abstract

OBJECTIVE To evaluate the effects of payment for performance (P4P) on the availability and stock-out rate of reproductive, maternal, newborn and child health (RMNCH) medical commodities in Tanzania and assess the distributional effects.

METHODS The availability of RMNCH commodities (medicines, supplies and equipment) on the day of the survey, and stock-outs for at least one day in the 90 days prior to the survey, was measured in 75 intervention and 75 comparison facilities in January 2012 and 13 months later. Composite scores for each subgroup of commodities were generated. A difference-in-differences linear regression was used to estimate the effect of P4P on outcomes and differential effects by facility location, level of care, ownership and socio-economic status of the catchment population.

RESULTS We estimated a significant increase in the availability of medicines by 8.4 percentage points ($P = 0.002$) and an 8.3 percentage point increase ($P = 0.050$) in the availability of medical supplies. P4P had no effect on the availability of functioning equipment. Most items with a significant increase in availability also showed a significant reduction in stock-outs. Effects were generally equally distributed across facilities, with effects on stock-outs of many medicines being pro-poor, and greater effects in facilities in rural compared to urban districts.

CONCLUSION P4P can improve the availability of medicines and medical supplies, especially in poor, rural areas, when these commodities are incentivised at both facility and district levels, making services more acceptable, effective and affordable, enhancing progress towards universal health coverage.

keywords Policy evaluation, payment for performance, medical commodities, structural quality of care, RMNCH, health financing

Introduction

The availability of essential medical commodities (medicines, medical supplies and equipment) is a key component of effective service delivery required for maintaining population health [1]. Shortages of medical commodities are associated with poor structural quality of care, or poor quality relating to the attributes of the setting in which care delivery occurs [2, 3], low levels of patient satisfaction and preventable deaths [4–9]. Medicine and supply shortages in public facilities are also responsible for a large share of the out-of-pocket payments faced by households in low- and middle-income settings limiting the affordability of care [1, 10]. However, ensuring the availability of essential medical commodities remains a challenge for many low-income country health systems.

According to the United Nations Commission on Life-Saving Commodities, payment for performance (P4P) is a strategy to improve access to life-saving commodities for maternal and child health [11, 12]. P4P provides financial incentives to providers and/or healthcare managers based on the achievement of pre-defined performance targets and is currently being rolled out in many low-income countries [13, 14]. P4P could theoretically affect the availability of medical commodities by, for example, incentivising the provision of intermittent preventive treatment (IPT) for malaria during antenatal care (ANC), through facility-level bonus payments, which can be used to procure commodities, and by incentivising health care managers to reduce drug stock-out rates.

The copyright line for this article was changed on 11 January 2017 after original online publication.

However, empirically, only four studies have reported on the effect of P4P on the availability of medical commodities in low-income countries. The effects are varied with no effects on the availability of drugs and equipment in Afghanistan [15]; no effects on patient perceptions of drug availability in Burundi [16]; an increase in patient perceptions of drug availability in the Democratic Republic of Congo (DRC) [17]; and a reduction in the availability of vaccines and equipment in another study from the DRC [18]. Only one study reports on stock-out rates [18] and none of the studies shed light on the pathways through which such changes occurred. Previous studies have not examined the potential heterogeneity of effects across facilities and effects on commodities related to non-incentivised services (spillover effects). This paper examines the effect of P4P on the availability and stock-outs of medicines, medical supplies and equipment for reproductive, maternal, newborn and child health in Tanzania and assesses whether these effects differed by facility location, level of care, facility ownership and socio-economic status of the facility's catchment population.

Methods

Study setting

Since the 1990s, Tanzania began a process of decentralisation of government functions including health services, involving the transfer of power from central to local government authorities [19]. As a result, district-level managers are responsible for preparing annual health sector plans and budgets to implement health programmes and renovations in facilities and are responsible for generating and managing resources for the district. District managers are supported by a regional health management team, while health facility governing committees oversee the implementation of plans and the management of resources at facility level. Public health facilities order medical commodities on a quarterly basis, based on an estimate of quantity needs; they submit requests to the district who review and send them on to the medical stores department (MSD) and distribute medical commodities to facilities (the 'pull' system) [20–22]. Districts and facilities can also use their own funds (e.g. insurance contributions, user fees and P4P bonus payments) to procure commodities in case of stock-outs [22–24]. Non-public hospitals that are contracted by districts to deliver services on behalf of the Ministry of Health and Social Welfare (MoHSW) also receive medical commodities from the MSD. All other non-public facilities either procure commodities from the MSD, foreign or local manufacturers, privately owned accredited drug dispensing outlets (ADDOs) and pharmacies [25–27]. Some commodities (vaccines, antiretrovirals (ARVs), vitamin A and family planning) are managed through disease-specific vertical programmes, which are financed externally, and distributed via the MSD or directly to facilities [24, 28, 29]. The MSD supply chain suffers from a shortage of commodities, inadequate budget allocations, inadequate tracking mechanisms and late delivery of required commodities [8, 22, 24, 30]. As a result, facilities experience regular shortages of essential drugs and supplies especially in the public sector [22, 24, 30, 31]. For example, out of 1297 facilities surveyed in 2012, only 41% stocked the 14 essential tracer medicines at the time of the survey [31]. An assessment in 2010 found that the MSD fulfilled 68% of hospital orders and 67% of orders from health centres and dispensaries [32].

P4P in Tanzania

In 2011, the MoHSW in Tanzania, with financial support from the Government of Norway, introduced a P4P scheme in Pwani region to improve reproductive, maternal, newborn and child health (RMNCH), which is ongoing. Pwani region is one of 30 regions in the country and has seven districts with more than 209 health facilities and a population of just over a million [33]. Financial incentives are given to health facilities, district and regional managers based on their performance on pre-defined service delivery targets (Table 1) [34, 35]. Most of the targets at facility level pertain to increases in service coverage, with four that involve the provision of medicines such as antiretroviral therapy (ART), IPT during ANC, vaccines and supplies such as partographs. District managers are rewarded for reducing the proportion of facilities in the district reporting stock-outs of essential medicines (Appendix S1a) for at least one week. Districts are required to verify facility performance reports, resulting in more frequent contact between district managers and providers which may also help reduce stock-outs. Facilities are required to open bank accounts to receive performance payments.

Facility and district performance data are verified every six months (one cycle). For dispensaries, the maximum payout, if all targets are fully attained, is USD 820 per cycle, while maximum payouts are USD 3220 and USD 6790 for health centres and hospitals, respectively. Incentive payouts at facility level include bonuses to staff (equivalent to 10% of monthly salary) and funds that can be used to procure drugs and supplies and for facility improvement (10% of the total in hospitals and 25% in lower level facilities). District and regional managers receive bonus payments of up to USD 3000 per cycle based on the performance of facilities in their district or region.

Table 1 Service indicators and performance targets for facilities

Performance indicators	Method	Baseline coverage (previous cycle) 0–20%	21–40%	41–70%	71–85%	85%+
Coverage indicators						
% of facility-based deliveries	Percentage point increase	15	10	5	5	Maintain
% of mothers attending a facility within 7 days of delivery	Percentage point increase	15	10	5	5	Maintain
% of women using long-term contraceptives	Percentage point increase	20	15	10	Maintain above 71	Maintain
% children under 1 year received measles vaccine	Overall result	50	65	75	80+	Maintain
% children under 1 year received Penta 3 vaccine	Overall result	50	65	75	80+	Maintain
% of complete partographs	Overall result	80	80	80	80+	Maintain above 80
HMIS reports submitted to district managers on time and complete	Overall result	100	100	100	100	100
Content of care indicators						
% ANC clients receiving two doses of IPT	Overall result	80	80	80	80+	Maintain above 80%
% HIV+ ANC clients on ART	Overall result	40	60	75	75+	Maintain
% children received polio vaccine (OPV0) at birth	Overall result	60	75	80	80+	Maintain

85%+ = 85% or more; 80%+ = 80% or more; HMIS, Health Management Information System Source: The United Republic of Tanzania, Ministry of Health and Social Welfare, 2011. The Coast Region Pay for Performance (P4P) Pilot: Design Document.

Study design

This study uses data from a controlled before and after study of the P4P scheme in Pwani region, Tanzania, conducted in all seven intervention districts and four comparison districts from Morogoro and Lindi regions [34, 35]. Baseline data were collected in January 2012 and 13 months later.

Data sources

The data on the availability and stock-outs of essential RMNCH commodities within the previous 90 days were collected through a survey of 75 facilities in each study arm. In the intervention arm, we included all 6 hospitals and 16 health centres that were eligible for the P4P scheme and a random sample of 53 eligible dispensaries. A corresponding number of facilities were surveyed in the comparison arm. The facility survey also documented facility characteristics and was administered to the facility incharge. To proxy the socio-economic status (SES) of the facility catchment population, we used data from a survey of 1500 households of women who had delivered in the previous 12 months prior to the baseline survey in each arm and a similar number in the follow-up survey (20 households sampled from the catchment area of each facility). More details on data sources and data collection are provided elsewhere [34, 35].

Outcome measures

Our main outcomes are the availability of RMNCH medicines, medical supplies and functioning equipment, and the stock-outs of medicines and supplies at the facility. If a commodity was available on the day of the survey, the outcome was coded 1 and 0 otherwise; if a commodity was out of stock for at least one day in the 90 days prior to the survey, the outcome was coded 1 and 0 otherwise (Appendix S1a).

Medical commodities were classified in terms of their therapeutic use as antibiotics, antimalarials, antihypertensives, antidiarrhoeal, anti-retrovirals (ARVs), oxytocics, vaccines, family planning, vitamin A, medical supplies and medical equipment (Appendix S1a). We differentiated between items that relate directly to a P4P target and those which do not, to examine eventual spillover effects. Items were also classified according to their beneficiary/recipient group along the RMNCH continuum of care based on the WHO classification of priority medicines [11, 12]. For each of these groupings, we generated composite scores based on an unweighted mean score across items in the group, which can be interpreted as the mean percentage availability/stock-out rate within the grouping across facilities. We measured the proportion of facilities with availability/stock-out of the respective commodity groups. In the generation of scores, we gave equal

weight to each commodity item for ease of interpretation, but we acknowledge some of the items may be more effective than others in enhancing better health outcomes.

Subgroup effects

We examined whether the effects of P4P differed with the wealth of the facility catchment population to see whether benefits were pro-poor, given the greater burden of out-of-pocket payments from stock-outs on poorer groups [1, 10, 30, 36]. We also examined effects by facility ownership (public/non-public) given the differing procurement and supply systems in public and non-public sectors; level of care (dispensary/health centre or hospital) given that dispensaries are typically worse off in drug availability [7, 31, 37]; and whether the facility was in an urban or rural district as facilities in urban districts are better connected by roads facilitating the distribution of commodities relative to those in rural districts.

To generate a wealth score for each household in the catchment area of the facility based on their ownership of 42 household items and characteristics we used principal component analysis (PCA)[38, 39] (Appendix S1c). We then calculated the average wealth score of the 20 households sampled within the facility catchment area. We ranked facilities by these scores from poorest (low score) to least poor and split them into terciles (poorest, middle and least poor).

Statistical analysis

We compared facility characteristics and outcome scores across study arms by using t-tests adjusting for clustering at the facility level. We used a linear difference-in-differences regression model to identify the effects of P4P on the availability and stock-out of medical commodities (1):

$$Y_{it} = \beta_0 + \beta_1(P4P_i \times \delta_t) + \beta_2\delta_t + \gamma_i + \varepsilon_{it} \qquad (1)$$

where Y_{it} is the outcome of facility i at time t. $P4P_i$ is a dummy variable, taking the value 1 if a facility is exposed to P4P and 0 if not. We controlled for time-invariant determinants γ_i with facility fixed effects, and δ_t year fixed effects. The error term is ε_{it}. The effect of P4P on the outcome is given by β_1.

In order to examine subgroup effects, we included a triple interaction term between treatment effect ($P4P_i \times \delta_t$) and subgrouping variable G_i. The associated two-order interaction terms were also included. The coefficient of interest for the differential effect is β_3 (2):

$$Y_{it} = \beta_0 + \beta_1(P4P_i \times \delta_t) + \beta_2\delta_t + \beta_3(P4P_i \times \delta_t \times G_i) + \beta_4(P4P_i \times G_i) + \beta_5(G_i \times \delta_t) + \gamma_i + \varepsilon_{it} \qquad (2)$$

For each of the effects, we report the confidence interval based on standard errors that are clustered at the facility level. As a robustness check, we clustered the standard errors at the district level and used the bootstrapping method to adjust for the small number of clusters [40]. We were unable to test whether the availability and stock-out outcomes were parallel between study arms prior to the intervention. However, we tested and confirmed that trends in facility-level utilisation for all incentivised services were parallel prior to the intervention [35, 41]. All analyses were performed using STATA version 13.

Ethical issues

The evaluation study received ethical approval from the Ifakara Health Institute Institutional Review Board and the Ethics Committee of the London School of Hygiene & Tropical Medicine. Study participants provided written consent to participate in this study, requiring them to sign a consent form that was read out to them by the interviewers. This consent form was reviewed and approved by the ethics committees prior to the start of the research.

Results

Baseline facility characteristics were fairly balanced across study arms (Table 2). However, facilities in the intervention arm were serving poorer populations than those in the comparison arm.

P4P was associated with an 8.4 percentage point increase in the availability of all 37 medicines combined ($P = 0.002$) and an 8.3 percentage point increase in the availability of medical supplies, although this was only borderline significant ($P = 0.050$) (Table 3). P4P had no effect on the availability of functioning equipment. Effects were noted for some medicines associated with P4P targets (antimalarials, antihypertensives and oxytocics used for deliveries) and supplies (partograph), although this effect was only borderline significant. There was no effect on vaccines, family planning and ARVs. Effects were observed for items that were not clearly linked to service targets, but were incentivised for district managers (antibiotics).

P4P was also associated with a reduction in stock-outs of medicines and medical supplies (Table 4). Most of those items where we found a significant increase in

Table 2 Baseline characteristics of health facilities

Facility characteristic	Intervention facilities (n = 75)	Control facilities (n = 75)	Difference (P–value)
Level of care			
Hospital (%)	8.0	8.0	0
Health centre (%)	21.3	21.3	0
Dispensary (%)	70.7	70.7	0
Ownership status			
Government/public facility (%)	84.0	82.7	1.3 (0.828)
Faith-based organisation (FBO) facility (%)	10.7	12.0	−1.3 (0.798)
Military/parastatal /private facility (%)	5.3	5.3	0 (0.652)
Infrastructure			
Electricity available (%)	68.0	66.7	1.3 (0.863)
Clean water available (%)	73.3	78.7	−5.3 (0.448)
Community/area features			
Facility in rural districts (%)	78.7	84.0	−5.3 (0.405)
Distance (km) from district headquarter, mean [SD]	56.9 [38.8]	62.9 [41.8]	−6.0 (0.367)
Poorest SES facilities (%)	40.0	26.7	13.3 (0.084)
Middle SES facilities (%)	34.7	32.0	2.7 (0.731)
Least poor SES facilities (%)	25.3	41.3	−16.0 (0.038)

SD is for standard deviation

availability were also less likely to be out of stock. In addition, there was a borderline significant 10.2 percentage point reduction in vaccine stock-outs (P = 0.073) and a 13.6 percentage point reduction in stock-outs of family planning medicines (P = 0.062) (Table 4). The effects of P4P on IPT and partograph stock-outs were not significant.

P4P reduced the stock-out of medicines across the RMNCH continuum of care and that of medical supplies benefiting mothers and newborns (Appendix S1b). Effects on availability were most pronounced for maternal, newborn and child medicines and reproductive health supplies.

The effect of P4P on the stock-outs of medicines overall was pro-poor, with reduction in facilities in the poorest tercile being 24.5 percentage points greater than that in the least poor tercile (P = 0.019); specifically, the effects on stock-outs of antimalarials, antibiotics and oxytocics were pro-poor; effects on antimalarial availability were also marginally pro-poor (Table 5). P4P had a greater effect on the availability of medicines and medical supplies in facilities in rural districts (by 10.4 percentage points, P = 0.051; and 22 percentage points, P = 0.003, respectively). Similarly, the effect of P4P on availability and stock-outs of antimalarials was greater in facilities in rural than urban districts (23.1 percentage points, P = 0.020; and 23.1 percentage points, P = 0.070, respectively). The effect of P4P on availability and stock-outs of antihypertensives was greater in health centres and hospitals than in dispensaries [by 19.9 percentage points (P = 0.020) and 26.1 percentage points (P = 0.064), respectively]. There were no differential effects by facility ownership.

When standard errors were clustered at the district level, the effects on the availability of antimalarials, oxytocics and delivery care drugs combined and, on stock-outs of oxytocics, vaccines and delivery care drugs combined were maintained (results not shown). However, the effects on composite indices for medicines combined and medical supplies were no longer significant.

Discussion

We examined the effects of P4P on the availability and stock-out rate of medical commodities for RMNCH. P4P was associated with significant improvements in availability and reductions in stock-outs of medicines and medical supplies, but had no effect on the availability of equipment. Among medicines, the main effects were for drugs associated with the delivery of some incentivised services: antimalarials, drugs to induce labour and manage bleeding (oxytocics) or manage hypertension during delivery (antihypertensives). However, there was little or no evidence of effects on medicines linked to other incentivised

Table 3 Effects of P4P on the availability of medical commodities mean score

	Baseline survey			Follow-up survey			Difference in differences, effect			
Category	P4P facilities	Control facilities	Difference	P4P facilities	Control facilities	Difference	*N*	Beta† [95% CI]	*P*–value	%*D**
Medicines combined (%)	60.8	65.7	−4.9**	63.9	60.7	3.2	295	8.4 [3.0 to 13.7]	0.002	13.8
Antimalarials – all (%)	60.3	69.9	−9.6**	69.7	59.3	−10.4**	295	20.5 [11.8 to 29.3]	0.000	33.9
Antimalarials – targeted (%)	74.6	93.2	−18.6***	96.0	90.7	5.3	295	25.2 [11.1 to 39.4]	0.001	33.8
Antibiotics (%)	36.3	39.9	−3.6	43.1	39.8	3.3	295	7.4 [0.8 to 14.1]	0.028	20.4
Antihypertensives (%)	36.2	37.1	−0.9	43.8	36.4	7.4*	295	8.7 [0.4 to 16.9]	0.040	24.0
Antidiarrhoeals (%)	60.6	63.5	−2.9	74.0	75.3	−1.3	295	1.9 [−12.5 to 16.3]	0.795	3.1
Oxytocics (%)	42.7	45.0	−2.3	45.8	32.9	−12.9***	295	15.0 [3.0 to 26.9]	0.014	35.1
Delivery care drugs – targeted (%)	39.5	41.1	−1.6	44.8	34.6	10.2***	295	11.8 [3.8 to 19.8]	0.004	29.9
ARVs – targeted (%)	55.4	50.3	5.1	57.4	60.4	−3.0	210	−7.9 [−20.3 to 4.7]	0.208	14.3
Vaccines – all (%)	94.8	92.9	1.9	96.9	92.9	4.0*	276	5.3 [−2.7 to 13.3]	0.193	5.6
Vaccines – targeted (%)	95.2	92.7	2.5	97.1	94.8	2.3	276	3.1 [−5.4 to 11.5]	0.475	3.3
Vitamin A (%)	91.9	91.8	0.1	92.9	92.9	0.0	276	3.2 [−8.7 to 15.0]	0.597	3.5
Family planning – targeted (%)	91.7	99.5	−7.8**	56.5	59.5	−3.0	255	7.3 [−4.6 to 19.3]	0.227	7.9
Medical supplies (%)	64.4	72.4	−8.0**	66.4	66.4	0.0	299	8.3 [0.01 to 16.5]	0.050	12.9
Partograph – targeted (%)	63.5	75.8	−12.3	77.0	76.0	1.0	274	16.1 [−3.0 to 35.3]	0.098	25.4
Medical equipment (%)	55.0	54.9	0.1	72.8	68.8	4.0	299	3.8 [−4.9 to 12.6]	0.391	6.9

Items included for medicines combined [37], medical supplies [11] and equipment [16]; 'targeted' are commodities linked to services targeted/incentivised by P4P; number of observations (*N*) is small for ARVs, family planning and vaccines because not all facilities stock these commodities; *the % D = (Beta / baseline mean) ×100, where the baseline mean of the dependent variable is for the intervention facilities; †the Beta is the estimated intervention effect controlling for a year dummy and facility fixed effects; *** denotes significance at 1%, ** at 5% and * at 10% level.

services such as vaccines, family planning, ARVs and supplies such as the partograph. P4P improved the availability/reduced stock-outs for some of the drugs that districts were incentivised for, including antibiotics (ampicillin, amoxicillin, gentamycin and flagyl). However, the scheme also reduced stock-outs of antibiotics that were not tied to any incentive (e.g. cotrimoxazole, chloramphenicol and crystapen injection). This suggests that P4P schemes have the potential to improve drug availability beyond those drugs that are directly linked to the delivery of incentivised services. Effects were generally equally distributed across facilities, with effects on medicine stock-outs being pro-poor in many cases, and greater in facilities in rural compared to urban districts. Greater improvements in the availability/stock-out reduction of antihypertensives in higher-level facilities are likely reflective of the greater number of obstetric referral cases at these facilities and associated need.

There are a variety of potential pathways to P4P effects on medicines and supplies in our study. The effect may in part be due to the provision of medicines being a pre-condition for meeting certain performance targets (e.g. IPT during ANC). The financial autonomy resulting from bank accounts enabled facilities to use bonus funds and cost sharing revenue (from user fees and community-based insurance) to procure drugs and supplies, consistent with findings from a process evaluation carried out alongside this study [42]. Incentives to district managers to limit drug stock-outs were also important, given the role of district managers in the procurement and supply process. By providing incentives to facilities and districts, the scheme ensured that stakeholders at all levels were working towards the same goals. The verification system under P4P also meant that district supervision was intensified, providing more opportunities for district managers to identify and address stock-outs of a wider range of drugs.

A number of medicines associated with incentivised services were not affected by P4P (vaccines, ARVs and family planning). The procurement of these items depended on donor funding [24, 28, 29]. The average availability of vaccines was above 94% at baseline (91% for family planning), so there was also little scope for improvement. Tanzania faced a problem with shortages of ARVs during the period of this study due to the introduction of a new treatment regimen, weak procurement

Table 4 Effects of P4P on the stock-out of medical commodities mean score

	Baseline survey			Follow-up survey			Difference in differences, effect			
Category	P4P facilities	Control facilities	Difference	P4P facilities	Control facilities	Difference	*N*	Beta† [95% CI]	*P*–value	%*D**
Medicines combined (%)	43.1	33.5	9.6***	26.7	30.4	−3.7	295	−13.6 [−22.1 to −5.1]	0.002	31.6
Antimalarials – all (%)	41.9	42.6	−0.7	29.8	40.4	−10.6**	295	−10.5 [−21.6 to 0.6]	0.064	25.1
Antimalarials – targeted (%)	27.1	18.9	8.2	6.7	10.7	−4.0	294	−13.3 [−29.8 to 3.2]	0.113	49.1
Antibiotics (%)	59.1	47.9	11.2**	41.0	45.2	−4.2	295	−16.6 [−29.5 to −3.8]	0.012	28.1
Antihypertensives (%)	57.0	46.0	11.0**	34.9	44.0	−9.1*	295	−21.0 [−35.1 to −6.9]	0.004	36.8
Antidiarrhoeals (%)	42.9	36.9	6.0	26.0	27.3	−1.3	294	−5.9 [−22.2 to 10.3]	0.472	13.8
Oxytocics (%)	55.2	39.3	15.9***	36.9	48.9	−12.0**	294	−27.2 [−43.7 to −10.7]	0.001	49.3
Delivery care drugs – targeted (%)	56.1	42.4	13.7***	35.9	46.4	−10.5**	295	−24.7 [−38.4 to −11.0]	0.000	44.0
ARVs – targeted (%)	40.6	32.5	8.1	25.0	25.0	0.0	210	−4.9 [−22.8 to 12.9]	0.585	12.1
Vaccines – all (%)	17.1	12.9	4.2	6.9	9.3	−2.4	276	−10.2 [−21.4 to 0.9]	0.073	59.6
Vaccines – targeted (%)	15.6	11.9	3.7	6.7	7.0	0.3	276	−7.4 [−18.8 to 4.1]	0.206	47.4
Vitamin A (%)	14.5	8.2	6.3	10.0	7.0	3.0	276	−6.6 [−20.3 to 7.0]	0.339	45.5
Family planning – targeted (%)	45.4	38.2	7.2	24.0	25.2	−1.2	255	−13.6 [−27.9 to 0.7]	0.062	29.9
Medical supplies (%)	39.7	29.4	10.3**	20.8	21.8	−1.0	286	−13.1 [−23.1 to −3.2]	0.010	32.9
Partograph – targeted (%)	33.9	18.6	15.3*	13.9	13.0	0.0	262	−12.3 [−31.9 to 7.3]	0.217	36.3

Items included for medicines combined [37], medical supplies [11] and equipment [16]; 'targeted' are commodities linked to services targeted/incentivised by P4P; number of observations (*N*) is small for ARVs, family planning and vaccines because not all facilities stock these commodities; *the % *D* = (Beta/baseline mean) ×100, where the baseline mean of the dependent variable is for the intervention facilities; †the Beta is the estimated intervention effect controlling for a year dummy and facility fixed effects; *** denotes significance at 1%, ** at 5% and * at 10% level.

mechanisms, and shortages of ARVs on the global market that were outside of facilities' control [43]. The lack of effect on equipment availability may be due to the lack of incentives attached to equipment availability at the facility or the district level. The cost of equipment is also higher than that of many drugs and supplies, which may have deterred facilities from such investments.

Our study stands in contrast to a recent review from low- and middle-income countries concluding that P4P is not effective in improving structural quality of care [44]. However, our finding of increased availability of drugs is consistent with that reported from South Kivu Province in the DRC [17], but contrary to the findings from Afghanistan [15], Burundi [16, 45] and Katanga Province in the DRC [18] that showed no effects. The differences in context and variation in programme design likely explain the difference in effects. In Afghanistan, Burundi and the DRC drugs/supplies were incentivised through service targets, and providers had financial autonomy as in Tanzania [15–17], and in Burundi, up to 50% of the bonus could be used to procure drugs; however, this was not clearly the case in the other settings. Unlike the Pwani scheme, many schemes weight bonus payments with structural quality scores, which include the availability of drugs and supplies [15–17]. While facilities could channel a percentage of their bonus to districts in the DRC [46], districts were not directly incentivised, nor were they incentivised in other settings.

Despite the importance of assessing distributional effects within programme evaluation [47, 48], ours is the first study to examine the heterogeneity of the effect of P4P on medical commodities. The pro-poor effects on medicines are encouraging as are the pro-rural effects and these are consistent with universal health coverage (UHC) goals and efforts to meet sustainable development goal (SDG) 3.

There are several limitations to this study. First, we used household data from the facility catchment area to proxy the SES of the facility's location based on a sample of 20 households that may not have accurately reflected the entire catchment population. Second, there was an imbalance in SES across study arms, although our results were reasonably robust when dividing facilities into SES groups in each arm separately. Third, we were unable to control for time-varying confounding factors due to a

Table 5 Heterogeneity of P4P effects on the availability and stock-out of medical commodities mean score

	Average effects			Differential effects by facility characteristics				
	Difference in differences, effect			Facility SES		Facility location	Ownership status	Level of care
Category	*N*	Beta† [95% CI]	*P*-value	(=1 if poorest SES) *β* (*P*–value)	(=1 if middle SES) *β* (*P*–value)	(=1 if in rural district) *β* (*P*–value)	(=1 if public facility) *β* (*P*–value)	(=1 if dispensary facility) *β* (*P*–value)
Panel A: Availability								
Medicines combined (%)	295	8.4 [3.0 to 13.7]	0.002	8.1 (0.173)	1.5 (0.817)	**10.4 (0.051)**	7.4 (0.180)	2.8 (0.578)
Antimalarials – all (%)	295	20.5 [11.8 to 29.3]	0.000	**18.9 (0.071)**	**25.9 (0.022)**	**23.1 (0.020)**	14.3 (0.119)	1.9 (0.844)
Antimalarials – targeted (%)	295	25.2 [11.1 to 39.4]	0.001	19.7 (0.319)	19.0 (0.333)	**24.2 (0.078)**	–23.1 (0.214)	–11.4 (0.522)
Antibiotics (%)	295	7.4 [0.8 to 14.1]	0.028	13.9 (0.113)	6.9 (0.409)	7.6 (0.421)	5.3 (0.617)	–5.6 (0.453)
Antihypertensives (%)	295	8.7 [0.4 to 16.9]	0.040	-9.3 (0.398)	–7.2 (0.485)	–2.6 (0.781)	12.8 (0.216)	**–19.9 (0.020)**
Oxytocics (%)	295	15.0 [3.0 to 26.9]	0.014	–10.8 (0.476)	–5.2 (0.731)	4.0 (0.762)	–3.4 (0.776)	5.4 (0.666)
Delivery care drugs – targeted (%)	295	11.8 [3.8 to 19.8]	0.004	–10.0 (0.322)	–6.2 (0.548)	0.7 (0.935)	4.7 (0.566)	–7.3 (0.365)
Medical supplies (%)	299	8.3 [0.01 to 16.5]	0.050	2.8 (0.815)	1.9 (0.849)	**22.0 (0.003)**	6.2 (0.522)	–4.9 (0.539)
Partograph – targeted (%)	274	16.1 [-3.0 to 35.3]	0.098	**41.5 (0.065)**	25.2 (0.307)	17.5 (0.510)	9.9 (0.652)	3.8 (0.850)
Panel B: Stock-out								
Medicines combined (%)	295	−13.6 [−22.1 to −5.1]	0.002	**−24.5 (0.019)**	−16.5 (0.110)	−8.9 (0.342)	2.5 (0.771)	7.4 (0.383)
Antimalarials – all (%)	295	−10.5 [−21.6 to 0.6]	0.064	**−23.6 (0.098)**	−23.6 (0.111)	**−23.1 (0.070)**	5.7 (0.624)	5.0 (0.696)
Antibiotics (%)	295	−16.6 [−29.5 to −3.8]	0.012	**−32.0 (0.032)**	−6.7 (0.668)	−15.3 (0.322)	−7.7 (0.624)	**23.9 (0.095)**
Antihypertensives (%)	295	−21.0 [−35.1 to −6.9]	0.004	−12.5 (0.465)	−5.9 (0.733)	−4.9 (0.743)	13.9 (0.263)	**26.1 (0.064)**
Oxytocics (%)	294	−27.2 [−43.7 to −10.7]	0.001	**−49.1 (0.017)**	−27.5 (0.171)	−14.4 (0.487)	−1.3 (0.946)	−17.7 (0.311)
Delivery care drugs – targeted (%)	295	−24.7 [−38.4 to −11.0]	0.000	**−30.8 (0.062)**	−18.4 (0.272)	−10.3 (0.512)	5.6 (0.680)	6.4 (0.649)
Vaccines – all (%)	276	−10.2 [−21.4 to 0.9]	0.073	1.0 (0.946)	−4.3 (0.749)	14.5 (0.160)	3.7 (0.749)	**17.9 (0.082)**
Family planning – targeted (%)	255	−13.6 [−27.9 to 0.7]	0.062	−29.6 (0.127)	−21.3 (0.128)	3.7 (0.842)	−11.9 (0.402)	8.2 (0.560)
Medical supplies (%)	286	−13.1 [−23.1 to −3.2]	0.010	−9.6 (0.467)	8.9 (0.471)	−1.9 (0.869)	−5.4 (0.543)	−0.9 (0.927)

Reference category in brackets: for poorest and middle SES (least poor SES), rural (urban), public (non-public) and dispensary (hospital & health centres); †the Beta is the estimated average intervention effect controlling for a year dummy and facility fixed effects; *β* is the estimated differential effects of P4P controlling for a year dummy and facility fixed effects; and statistically significant differential effects in bold (*P*-value < 0.10).

lack of data, but confounding bias due to time-invariant factors were adjusted through fixed effects estimation. Fourth, although we tested and confirmed the assumption of parallel trends in facility utilisation outcomes prior to the intervention, we failed to test that of drug availability and stock-out outcomes due to a lack of historical data on these outcomes. We were also unable to capture seasonal fluctuations in drug availability as this requires time series data which were not available. Finally, potential type I errors due to multiple hypotheses testing are a concern to inference; however, we used subgroups of items to minimise the risk of this error.

Conclusion

Our study has shown that P4P, when introduced with facility and district-level incentives and in a context where facilities and local government authorities have autonomy over the use of funds, can improve the availability of drugs and supplies and enhance good quality of care. This makes services more acceptable, effective and affordable, especially in facilities serving poor, rural populations, enhancing progress towards universal health coverage [1, 10].

Acknowledgements

We thank Kara Hanson, Masuma Mamdani, Gaute Torsvik, Bjarne Robberstad and Iddy Mayumana for reviewing the paper and for valuable comments and suggestions. We acknowledge the process evaluation findings provided by Masuma Mamdani and Iddy Mayumana that supplement this study. We also thank the whole P4P evaluation research team, including data collectors and field coordinators. The Government of Norway funded the data collection for the programme evaluation that was used in this paper. The Research Council of Norway and 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 analysis and writing. This study is part of a PhD thesis at the University of Bergen for Peter Binyaruka, who is financially supported by the Norwegian State Education Loan Fund. The funding bodies had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

References

1. WHO. *Equitable Access to Essential Medicines: A Framework for Collective Action. Volume March*. World Health Organization: Geneva, 2004.
2. Donabedian A. The quality of care. How can it be assessed? *JAMA* 1988: **260**: 1743–1748.
3. Donabedian A. Evaluating the quality of medical care. *Milbank Mem Fund Q* 2005: **83**: 691–729.
4. Quick JD, Boohene N-A, Rankin J, Mbwasi RJ. Medicines supply in Africa. *BMJ* 2005: **331**: 709–710.
5. Uzochukwu BS, Onwujekwe OE, Akpala CO. Effect of the Bamako-Initiative drug revolving fund on availability and rational use of essential drugs in primary health care facilities in south-east Nigeria. *Heal Policy Plan* 2002: **17**: 378–383.
6. Macha J, Harris B, Garshong B *et al.* Factors influencing the burden of health care financing and the distribution of health care benefits in Ghana, Tanzania and South Africa. *Health Policy Plan* 2012: **27**(Suppl 1): 46–54.
7. Penfold S, Shamba D, Hanson C *et al.* Staff experiences of providing maternity services in rural southern Tanzania - a focus on equipment, drug and supply issues. *BMC Health Serv Res* 2013: **13**: 61.
8. Mkoka DA, Goicolea I, Kiwara A, Mwangu M, Hurtig A-K. Availability of drugs and medical supplies for emergency obstetric care: experience of health facility managers in a rural District of Tanzania. *BMC Pregnancy Childbirth* 2014: **14**: 108.
9. UNCoLSC. *Scaling Up Life Saving Commodities for Women, Children, and Newborns - An Advocacy Toolkit*. PATH: Washington, DC, 2015.
10. Cameron A, Ewen M, Ross-Degnan D, Ball D, Laing R. Medicine prices, availability, and affordability in 36 developing and middle-income countries: a secondary analysis. *Lancet* 2009: **373**: 240–249.
11. United Nations. *UN Commission on Life-Saving Commodities for Women and Children: Commissioner's Report*. United Nations: New York, 2012 (September).
12. Pronyk PM, Nemser B, Maliqi B *et al.* The UN Commission on Life Saving Commodities 3 years on: global progress update and results of a multicountry assessment. *Lancet Glob Heal* 2016: **4**: e276–e286.
13. Witter S, Fretheim A, Kessy FL, Lindahl AK. Paying for performance to improve the delivery of health interventions in low- and middle-income countries. *Cochrane Database Syst Rev* 2012, 2:CD007899.
14. Meessen B, Soucat A, Sekabaraga C. Performance-based financing: just a donor fad or a catalyst towards comprehensive health-care reform? *Bull World Health Organ* 2011: **89**: 153–156.
15. Engineer CY, Dale E, Agarwal A *et al.* Effectiveness of a pay-for-performance intervention to improve maternal and child health services in Afghanistan: a cluster-randomized trial. *Int J Epidemiol* 2016: **45**: 1–9.
16. Bonfrer I, Soeters R, Van de Poel E *et al.* Introduction of performance-based financing in burundi was associated with improvements in care and quality. *Health Aff (Millwood)* 2014: **33**: 2179–2187.
17. Soeters R, Peerenboom PB, Mushagalusa P, Kimanuka C. Performance-based financing experiment improved health

care in the Democratic Republic of Congo. *Health Aff (Millwood)* 2011: **30**: 1518–1527.
18. Huillery E, Seban J. *Pay-for-Performance, Motivation and Final Output in the Health Sector: Experimental Evidence from the Democratic Republic of Congo*. Sciences Po Economics Discussion Papers, 2014.
19. Frumence G, Nyamhanga T, Mwangu M, Hurtig AK. Challenges to the implementation of health sector decentralization in Tanzania: Experiences from kongwa district council. *Glob Health Action* 2013: **6**: 1–11.
20. MoHSW. *Mapping of the Medicines Procurement and Supply Management System in Tanzania*. Ministry of Health and Social Welfare (MoHSW): Dar es Salaam, 2008.
21. Euro Health Group. *The United Republic of Tanzania Drug Tracking Study*. Euro Health Group: Denmark, 2007 (August).
22. SIKIKA. *Medicines and Medical Supplies Availability Report. Using Absorbent Gauze Availability Survey as an Entry Point. A Case of 71 Districts and 30 Health Facilities across Mainland Tanzania*. SIKIKA: Dar es Salaam, Tanzania, 2011.
23. MoHSW. *Mid Term Review of the Health Sector Strategic Plan III 2009-2015: Health Care Financing*. Technical Report, Ministry of Health and Social Welfare (MoHSW), United Republic of Tanzania, 2013 (October 2013).
24. MoHSW. *Tanzania Health Sector Strategic Plan 2015 -2020 (HSSP IV)*. Ministry of Health and Social Welfare (MoHSW): Dar es Salaam, 2015 (July).
25. Centre for Pharmaceutical Management. *Accredited Drug Dispensing Outlets in Tanzania: Strategies for Enhancing Access to Medicines Program*. Management Sciences for Health: Arlington, VA, 2008.
26. Zomboko FE, Tripathi SK, Kamuzora FK. Challenges in procurement and use of donated medical-equipments: study of a Selected Referral Hospital in Tanzania. *J Arts Sci Commer* 2012: **4**: 41–48.
27. Rutta E, Liana J, Embrey M *et al.* Accrediting retail drug shops to strengthen Tanzania's public health system: an ADDO case study. *J Pharm policy Pract* 2015: **8**: 23.
28. MoHSW. *Tanzania Mainland Expanded Programme on Immunization (EPI) Review*. Ministry of Health and Social Welfare (MoHSW): Dar es Salaam, 2010 (July).
29. Yadav P, Lega Tata H, Bababley M. *Storage and Supply Chain Management*. The World Medicines Situation 2011, WHO: Geneva, 2011.
30. Wales J, Tobias J, Malangalila E, Swai G, Wild L. *Stock-Outs of Essential Medicines in Tanzania: A Political Economy Approach to Analysing Problems and Identifying Solutions*. TWAWEZA: Twaweza ni sisi, Dar es Salaam, Tanzania, 2014 (March).
31. MoHSW. *Tanzania Service Availability and Readiness Assessment (SARA) 2012*. Ministry of Health and Social Welfare (MoHSW) and Ifakara Health Institute: Dar es Salaam, 2013.
32. USAID. *Tanzania Health System Assessment 2010*. Health Systems 20/20 project. Abt Associates Inc.: Bethesda, MD, 2011 (January).
33. NBS. *2012 Population and Housing Census: Population Distribution by Administrative Areas*. National Bureau of Statistics (NBS): Dar es Salaam, 2013.
34. Borghi J, Mayumana I, Mashasi I *et al.* Protocol for the evaluation of a pay for performance programme in Pwani region in Tanzania: A controlled before and after study. *Implement Sci* 2013: **8**: 80.
35. Binyaruka P, Patouillard E, Powell-Jackson T, Greco G, Maestad O, Borghi J. Effect of paying for performance on utilisation, quality, and user costs of health services in Tanzania: a controlled before and after study. *PLoS One* 2015: **10**: e0135013.
36. Kamuzora P, Gilson L. Factors influencing implementation of the Community Health Fund in Tanzania. *Health Policy Plan* 2007: **22**: 95–102.
37. Choi Y, Ametepi P. Comparison of medicine availability measurements at health facilities: evidence from Service Provision Assessment surveys in five sub-Saharan African countries. *BMC Health Serv Res* 2013: **13**: 266.
38. Filmer D, Pritchett LH. Estimating wealth effects without expenditure data–or tears: an application to educational enrollments in states of India. *Demography* 2001: **38**: 115–132.
39. Vyas S, Kumaranayake L. Constructing socio-economic status indices: how to use principal components analysis. *Health Policy Plan* 2006: **21**: 459–468.
40. Cameron A, Miller D. A Practitioner's Guide to Cluster-Robust Inference. *J Hum Resour* 2015: **50**: 317–372.
41. Anselmi L, Binyaruka P, Borghi J. Understanding causal pathways within health systems policy evaluation through mediation analysis: an application to payment for performance (P4P) in Tanzania. (Forthcoming).
42. Mayumana I, Borghi J, Anselmi L, Mamdani M, Lange S. Effects of Payment for Performance on accountability mechanisms: evidence from Pwani, Tanzania. (Forthcoming).
43. SIKIKA. *Shortage of Antiretrovirals: What Went Wrong?* SIKIKA: Dar es Salaam, Tanzania, 2014.
44. Das A, Gopalan S, Chandramohan D. Effect of pay for performance to improve quality of maternal and child care in low- and middle-income countries: a systematic review. *BMC Public Health* 2016: **16**: 1–11.
45. Rudasingwa M, Soeters R, Bossuyt M. The effect of performance-based financial incentives on improving health care provision in Burundi: a controlled cohort study. *Glob J Health Sci* 2015: **7**: 15–29.
46. Fox S, Witter S, Wylde E, Mafuta E, Lievens T. Paying health workers for performance in a fragmented, fragile state: reflections from Katanga Province, Democratic Republic of Congo. *Health Policy Plan* 2014: **29**: 96–105.
47. Djebbari H, Smith J. Heterogeneous impacts in PROGRESA. *J Econom* 2008: **145**: 64–80.

48. Markovitz AA, Ryan AM. Pay-for-Performance: Disappointing Results or Masked Heterogeneity? *Med Care Res Rev* 2016: doi: 10.1177/1077558715619282.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Appendix S1 (a) List of medical commodities and classification. (b) Effects of P4P on mean score availability and stock-out of medical commodities [By RMNCH classification]. (c) Items used to construct household socio-economic status score.

Corresponding Author Peter Binyaruka, Ifakara Health Institute, PO Box 78373, Dar es Salaam, Tanzania. E-mail: pbinyaruka@ihi.or.tz