



Pay-for-performance reduces bypassing of health facilities: Evidence from Tanzania

Sosina Bezu^{a,b,*}, Peter Binyaruka^{a,c}, Ottar Mæstad^a, Vincent Somville^{a,d}

^a Chr.Michelsen Institute, P.O.Box 6033, 5892, Bergen, Norway

^b Diversity Institute, Ryerson University, Canada

^c Ifakara Health Institute, Dar es Salaam, Tanzania

^d Norwegian School of Economics, 5045, Bergen, Norway

ARTICLE INFO

Keywords:

bypassing
Maternal care
Health service use
Pay for performance
Health financing
Health governance
Tanzania
Africa

ABSTRACT

Many patients and expectant mothers in low-income countries bypass local health facilities in search of better-quality services. This study examines the impact of a payment-for-performance (P4P) scheme on bypassing practices among expectant women in Tanzania. We expect the P4P intervention to reduce incidences of bypassing by improving the quality of services in local health facilities, thereby reducing the incentive to migrate. We used a difference-in-difference regression model to assess the impact of P4P on bypassing after one year and after three years. In addition, we implemented a machine learning approach to identify factors that predict bypassing. Overall, 38% of women bypassed their local health service provider to deliver in another facility. Our analysis shows that the P4P scheme significantly reduced bypassing. On average, P4P reduced bypassing in the study area by 17% (8 percentage points) over three years. We also identified two main predictors of bypassing - facility type and the distance to the closest hospital. Women are more likely to bypass if their local facility is a dispensary instead of a hospital or a health center. Women are less likely to bypass if they live close to a hospital.

1. Introduction

Bypassing local health facilities is a common practice among patients in low-income countries, including among women bypassing their local service provider to deliver in their preferred facility (Fleming et al., 2016; Gauthier and Wane, 2011; Kruk et al., 2009a,b; Salazar et al., 2016; Shah, 2016). The rate of bypassing among women who delivered in a health facility is reported to be 44–75% in Tanzania (Kanté et al., 2016; Kruk et al., 2009a,b), 55–70% in Nepal (Karkee et al., 2015; Shah, 2016) and 38–68% in India (Rao and Sheffel, 2018; Sabde et al., 2018; Salazar et al., 2016). Bypassing differs across place and facilities. Bypassing is more common in urban than rural areas (Gauthier and Wane, 2011), perhaps because users have more accessible choices. In Ghana, for example, 54% of patients in the capital bypassed at least one modern facility while the rate is only 14% for rural residents.

Empirical studies suggest that women bypass their local facility to obtain better services; the quality of local health facilities plays a leading role in bypassing decisions (Kanté et al., 2016; Kruk et al., 2009a,b; Sabde et al., 2018; Salazar et al., 2016; Shah, 2016). Among the quality

indicators, availability of equipment and the number of qualified staff were found to play a significant role on bypassing decisions (Kanté et al., 2016; Leonard et al., 2002). Often patients bypass lower level health facilities to seek treatment in higher level facilities (Akin and Hutchinson, 1999; Salazar et al., 2016; Shah, 2016). Prices are also important, particularly for poorer users. Low-income users bypass high quality health service providers in search of less expensive treatment in alternative facilities (Kanté et al., 2016; Kruk et al., 2009a,b).

Bypassing is however costly for patients, both financially and in terms of travel time. The financial cost may include additional out-of-pocket fees for services, transportation fees and other incidentals (Rao and Sheffel, 2018). In Chad, bypassing patients spend on average 2.5 times more on consultation costs and twice the travel time of patients who use health centers closer to their home (Gauthier and Wane, 2011).

Bypassing behavior of women is influenced by their health status and socio-economic background. Generally, women are more likely to bypass if they are wealthier (Shah, 2016), first-time mothers (Kruk et al., 2009a,b; Shah, 2016), or experience peripartum complications (Karkee et al., 2015; Sabde et al., 2018). Other individual background factors are

* Corresponding author. Diversity Institute, Ryerson University, 350 Victoria Street, ON M5B 2K3, Toronto, Canada.

E-mail addresses: sosinac@yahoo.com, sosina.bezu@ryerson.ca (S. Bezu).

context specific. For example, in Nepal upper cast women are more likely to bypass local facilities (Sabde et al., 2018), and in India the bypassing probability increases with the woman's level of education (Sabde et al., 2018; Salazar et al., 2016).

While it may be possible to reduce bypassing by instituting a better referral system and restricting the practice of 'jumping' lower level facilities to go directly to hospitals (Akin and Hutchinson, 1999), a significant portion of bypassing also involve bypassing among lower level health facilities (Kanté et al., 2016; Sabde et al., 2018). Most of the findings from studies on bypassing suggest that at the center of bypassing decisions is the search for a certain standard of quality service at an affordable price. We therefore hypothesize that health interventions that improve the quality of service across health facilities will reduce bypassing as the quality standard is more likely to be met by the local facility.

In this study, we test this hypothesis by evaluating the impact of a health system intervention that has the explicit aim of improving quality of health service across all facilities. Specifically, this study assesses the impact of a performance-based financing scheme, also called pay-for-performance (P4P), introduced in the Pwani region of Tanzania with the aim of improving maternal and child health (Binyaruka et al., 2018a). We focus on the effects of P4P on the rate of bypassing of the nearest facility. Using difference-in-difference estimates, we show that P4P significantly reduces bypassing for delivery services. To our knowledge, this is the first study to systematically assess whether and to what extent pay-for-performance influences bypassing.

2. Background: the P4P program in Tanzania

Payment for performance (P4P) is a mechanism by which health providers are funded, at least partially, based on their performance as measured by the quality and quantity of services provided (Meessen et al., 2011). The financing approach may include a fee-for-service payment for the minimum package of services delivered or use of a balanced score card that targets quality (Fritsche et al., 2014). Many governments in Africa are currently piloting or implementing performance-based financing schemes to increase the quality of services. The Health Results Innovation Trust Fund (HRITF) of the World Bank is currently supporting the implementation of P4P in more than 21 countries in Africa (WB, 2018).

P4P was introduced in Tanzania by the Ministry of Health and Social Welfare with the aim of improving maternal and child health (MCH). The scheme was piloted in Pwani region to inform the implementation of a national level P4P program (Binyaruka et al., 2015a). All health facilities, irrespective of ownership, providing MCH services in Pwani participated in the P4P scheme: they didn't have a choice to opt-in or not and we therefore don't need to worry about self-selection into the

program (Binyaruka et al., 2018a).

P4P provided financial bonuses to health facilities and public health managers based on achievement of eight maternal and child health (MCH) performance targets (Table 1). In addition, there were two indicators related to monitoring and reporting. Five of the MCH performance indicators focused on service coverage (e.g., the proportion of women who delivered at a health facility) and three focused on content of care (e.g., provision of preventive treatment for malaria during antenatal care). The targets were either defined in terms of a performance level or in terms of performance improvements since the previous cycle of measurement. Targets were differentiated based on past performance, but in a way that gave all facilities incentives to achieve high performance over time (Binyaruka et al., 2018a).

There were also incentives for health managers. They were rewarded based on the overall performance of facilities in their district/region, as well as an additional set of indicators; maternal and newborn deaths audited properly and timely; reducing stock-out rates of essential drugs; timely reporting the facility data from district to regional level, and from regional to national level.

Performance was assessed, and payments made, every six months. The maximum payout per cycle was USD 820 per dispensary, USD 3220 per health center, and USD 6790 per hospital. To achieve the maximum payout, the facility had to reach all targets. No payments were made if less than three quarters of the targets were met, while 50% was paid if the facility met more than three quarters (but not all) of the targets. At primary health care facilities, three quarters of the bonus were distributed among health workers, while the remaining 25% went to the health facility and could be invested in drugs, supplies, or facility improvements (Anselmi et al., 2017). These payments are additional to the funding facilities receive to cover operational costs and the salaries of health workers. At hospitals, 90% of the bonus went to health workers. The maximal bonus for health workers amounted to around 10% of their salary (Binyaruka et al., 2015a, 2015b).

The Tanzanian health system has a referral structure. Patients should first seek care in a nearby facility, and going to higher level hospital normally requires a referral. This is however weakly imposed and that is why bypassing is possible and so commonly observed. As we discuss next, there however good reason to expect P4P to reduce bypassing.

3. Conceptual framework

Conceptually, we can frame the facility selection as a constrained optimization decision where the client selects a facility that satisfies her preferences, given income and time constraints. For each level of quality that a facility provides, the higher the price (fees plus transportation costs), the lower is the demand for that facility's service and vice versa. Within this general relationship, the health status, previous birth

Table 1
Performance indicators and targets for health facilities.

P4P performance indicators	Type of target	Past performance (previous cycle)				
		0–20%	21–40%	41–70%	71–85%	85%+
Service coverage indicators						
% of institutional/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	Overall result	50%	65%	75%	80%+	Maintain
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
% of children receiving polio vaccine at birth in health facilities	Overall result	60%	75%	80%	80%+	Maintain

Notes: 85%+ = 85% or more; 80%+ = 80% or more; HMIS=Health Management Information System; ANC = Antenatal care . IPT= Intermittent preventive treatment for malaria; ART = Antiretroviral treatment. Source: The United Republic of Tanzania, Ministry of Health and Social Welfare. 2011. The Coast Region Pay for Performance (P4P) Pilot: Design Document.

experience and socioeconomic background of the woman influence her individual valuation of quality.

The quality of services provided by health facilities is not always observable. Hence, clients use structural quality indicators as proxy; for example, users are often attracted by facilities with more doctors, equipment and emergency treatment capabilities (Kruk et al., 2009; Leonard et al., 2002; Sabde et al., 2018; Salazar et al., 2016). Users may also use other indirect quality indicators. For example, facilities that are recently renovated signal improvement in service quality and attract patients (Kruk et al., 2009a,b).

We hypothesize that P4P influences bypassing decision by changing the quality of care health facilities provide and possibly the prices they charge. The P4P payments to health care facilities and managers is based on achievement of measurable maternal and child health performance

Targets (Table 1). Thus, healthcare facilities have an incentive to improve the quality of their facility and the services they provide in order to attract more women to deliver at their facility and use other maternal and child health services. The additional resources provided to facilities through the P4P scheme increase their capacity to raise quality, and in some cases charge lower fees. Existing studies on the P4P scheme in Tanzania suggest that the program indeed improved quality, particularly their stock of drugs and other medical supplies (Anselmi et al., 2017; Binyaruka and Borghi, 2017; Binyaruka et al., 2015a).

Other aspects of quality indicators are also important for bypassing decisions. Rao and Sheffel's (2018) study on bypassing behavior in rural India shows that the clinical competence of the health care providers had a greater effect on reducing bypassing compared to structural quality of the primary healthcare facility, such as the building condition and drug stock-outs (Rao and Sheffel 2018). While the P4P scheme may not be significant enough to motivate or enable lower level facilities in Tanzania to recruit better trained workers in the short run, there is evidence suggesting that P4P scheme influenced behavior of current employees. An impact study of P4P found that health workers in P4P participating facilities displayed more kindness during deliveries and health workers received more supervision visits from superiors (Anselmi et al., 2017). This is consistent with the incentive structure in the P4P scheme as workers and managers receive 75% of the bonus achieved when facilities meet their target. In addition to improved quality, the P4P scheme also led to a 5% reduction in the share of clients who paid out-of-pocket for deliveries (Binyaruka et al., 2015a).

4. Methods

4.1. Data sources

The analysis is based on two sources of data: our own survey data combined with national administrative data on the location of health facilities. We collected the survey data in 2012, 2013 and 2015 (baseline, midline, and endline). The surveys covered all districts in Pwani as well as four control districts in two neighboring regions (Morogoro and Lindi) that did not implement P4P. The control districts were selected for their similarity with intervention districts in relation to relevant attributes: poverty, literacy rates, institutional delivery rates, infant mortality, population per health facility, and the number of children under one year of age per capita. Another selection criterion was that the control districts should not be exposed to other interventions targeting maternal and child health.

The survey data includes data at the health facility level as well as the household level. In the intervention districts, all eligible hospitals (n = 6) and health centers (n = 16) were included in the surveys. In addition, all private dispensaries (n = 11) and a random sample of public dispensaries (n = 42) were included. An equivalent number of matching facilities were selected from the control districts (Kilwa, Mvomero, Morogoro town and Morogoro rural). In total, 150 health facilities (75 treatment, 75 control) were sampled, 132 that are publicly owned and 18 private. The sample covered 46% of facilities implementing P4P in

Pwani region and 34% of the facilities in the comparison districts.

At the household level, we sampled women who delivered during a one year period preceding the surveys. In each of the three survey rounds, twenty women were randomly selected from the catchment area of each facility, implying a total number of 1500 women from the treatment areas and the same number for the control areas. The facility which is associated with each woman through this sampling procedure will be the one that is closest to her home. In the paper we use "local facility" to refer to these facilities. The sampled women thus always belong to the catchment area of their local facility. Further details about sampling and the surveys are presented in Anselmi et al. (2017) and Borghi et al. (2013).

We asked the women where they delivered their latest born child. When the delivery happened in a health facility, we collected the name and location of the facility. We then combined the list of facilities where our sampled women gave birth with the official list of all facilities (private and public) obtained from the Ministry of Health. The administrative data includes the location (usually the GPS coordinates) of the facility. Merging the two datasets allowed us to check whether the women delivered in their local facility, and if not, how far away they went. As this study focuses on bypassing, the sample used for analysis includes only women who delivered in a facility and the women who were living in the same home at the time of the survey as when they gave birth (Some women were living in a different place when they gave birth, and later moved to the village where we did the survey. They are not included in this analysis, to avoid counting them as "bypassers"). The final dataset includes 144 facilities and 6229 women (for six facilities, we couldn't merge the wards with the wards from the Ministry's list and we therefore cannot include them in this analysis. Three rounds of data collection at 144 facilities implies a potential sample of 8640 women. Around 1000 women are excluded because they did not deliver at a facility. Another 892 are excluded because they moved after they delivered. Finally, around 500 observations are excluded because we were not able to merge the declared place of delivery with the official list.) The difference in the proportion of missing observations between treated of control regions is small (2.9 percentage points) and not statistically significantly different from zero (p-value = 0.166).

4.2. Empirical framework

Our empirical analysis compares pre-post differences in bypassing between treated and comparison areas using a difference-in-differences regression model. The empirical model we estimate is

$$Y_{it} = \alpha + \beta_{ST} P4P_{it} ST_t + \beta_{LT} P4P_{it} LT_t + \rho P4P_{it} + \gamma ST_t + \delta LT_t + \varepsilon_{it} \quad (1)$$

where the term Y_{it} is an indicator variable that takes the value one if woman i delivered in a facility other than her local facility at time t (except if the woman was referred to the other facility from her local one: it happens that a woman is referred from her nearby facility and sent to another one for delivery (typically from a dispensary to a hospital); we don't count those cases as "bypassing" and we kept those women in the analysis, the "bypass" variable just takes a value of zero for them). $P4P_{it}$ denotes the participation status of the local health facility of individual i in the P4P scheme, and ST_t and LT_t are indicators for the 1 year (short-term) and 3 years (longer-term) follow-ups (ST_t is equal to one for observations from the 2013 follow-up and LT_t is equal to one for the 2015 follow-up). ε_{it} is a random error term which is assumed to be identically and independently distributed. This equation is estimated using OLS. The average effect of the P4P scheme after a year is given by the coefficient β_{ST} . Similarly, β_{LT} is interpreted as the effect of the P4P scheme after three years.

In addition to the basic model, we also estimate the model after including other control variables. The controls include the socioeconomic background of the woman and facility level covariates. Control variables are listed in Table 2.

Table 2
Descriptive statistics at baseline.

	No P4P		P4P		
	Mean	Std. Dev.	Mean	Std. Dev.	
Covariates - woman level					
Education					
No education	0.17	0.37	0.19	0.39	
Some primary	0.07	0.25	0.08	0.27	
Primary or some secondary	0.67	0.47	0.67	0.47	
Secondary or above	0.10	0.30	0.07	0.25	^a
Main occupation					
Formal sector job	0.02	0.15	0.02	0.14	
Farmer	0.53	0.50	0.42	0.49	^a
Self employed business	0.20	0.40	0.24	0.43	
Take care of child at home	0.23	0.42	0.31	0.46	^a
Other or no job	0.01	0.12	0.01	0.11	
Age	26.19	6.53	26.45	6.70	
Married	0.63	0.48	0.71	0.45	^a
Has health insurance	0.08	0.28	0.09	0.29	
Religion					
Catholic	0.23	0.42	0.07	0.26	^a
Protestant	0.08	0.28	0.06	0.24	
Muslim	0.68	0.47	0.86	0.34	^a
Observations	1104		980		
Covariates - household level					
Education of household head					
No education	0.01	0.11	0.02	0.13	
Some primary	0.19	0.39	0.21	0.41	
Primary or some secondary	0.68	0.47	0.69	0.46	
Secondary or above	0.12	0.33	0.08	0.27	^a
Household members	4.85	1.85	4.67	1.79	
Household asset index	0.53	3.39	-0.22	2.87	
Km to nearest Dispensary [‡]	1.19	2.87	1	2.28	
Km to nearest Hospital	24.82	20.63	26.33	18.91	
Km to nearest Health Center	11.62	12.09	13.33	17.14	
Owens motorcycle	0.07	0.26	0.07	0.25	
Owens car	0.01	0.11	0.01	0.09	
Observations	1104		980		
Covariates - facility level					
Type					
Dispensary	0.67	0.47	0.68	0.47	
Health center	0.25	0.44	0.24	0.43	
Hospital	0.08	0.28	0.08	0.28	
Ownership					
Public	0.88	0.33	0.89	0.32	
Private	0.13	0.33	0.11	0.32	
Observations	72		72		

[‡] indicates variables not included as a covariate in the estimations.

^a Indicates that the difference in means between P4P and No P4P groups is significant at the 5% level.

In our main specifications, we cluster the standard errors at the facility level, given that we use a panel of repeated observations by facility. An alternative could be to cluster at the district level, since we randomized the sampling at that level, but the number of districts is small (11), and clustering at that level could therefore be misleading. In the context of difference-in-difference estimators with few cluster, it has been shown that even the wild cluster bootstrap methods can fail and can lead to both overrejection or underrejection of the null (Mackinnon and Webb 2017, 2018; Roodman 2019). In order to provide the reader with complete information, we nonetheless add p-values calculated with a wild cluster bootstrap procedure, with clustering at the district level, to our main table (Table 4). The p-values are calculated with the algorithm developed in Roodman et al., (2019).

A crucial assumption for using the difference-in-difference estimate for causal inference is that bypassing would have evolved in the same manner in the treated and control regions without the introduction of P4P. It is of course not possible to test this assumption, but the standard practice is to test for differences in trends before the policy change. We

Table 3
Bypassing rate and place of delivery.

	No P4P	P4P
	Baseline (2012)	
Bypassing	0.37	0.48
Where the woman delivered:		
Other ward	0.32	0.38
Other district	0.13	0.2
Other region	0.07	0.14
Observations	1104	980
Midline (2013)		
Bypassing	0.38	0.42
Where the woman delivered:		
Other ward	0.34	0.34
Other district	0.12	0.15
Other region	0.07	0.11
Observations	854	861
Endline (2015)		
Bypassing	0.32	0.35
Where the woman delivered:		
Other ward	0.29	0.27
Other district	0.09	0.12
Other region	0.04	0.08
Observations	1138	1255

Table 4
Bypassing by facility type and distance to the nearest hospital.

	Proportion who bypass	Standard error	Observations
Hospital	0.21	0.02	523
Health center	0.33	0.01	1611
Hospital <10 kms away	0.73	0.02	393
Hospital >10 kms away	0.20	0.01	1218
Dispensary	0.42	0.01	4095
Hospital <10 kms away	0.60	0.02	731
Hospital >10 kms away	0.38	0.01	3364

can do this in our case with the baseline survey. Indeed, using the baseline data we can check month by month if the trends in bypassing differed between our control and treated areas during the year that preceded the introduction of P4P (see Fig. 1). At baseline, we surveyed mothers of children born 3–14 months before the survey, and the

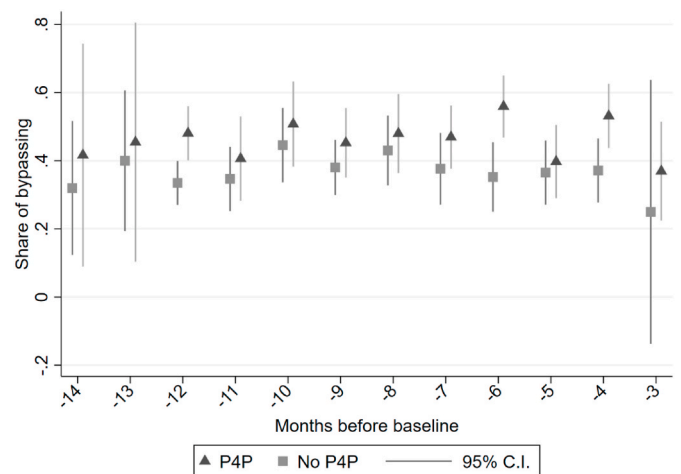


Fig. 1. Bypassing by quarter before the introduction of P4P in treatment and control areas.

horizontal axis therefore goes from -14 (14 months before baseline) to -3 (3 months before baseline).

A visual inspection of Fig. 1 doesn't indicate any important difference in trends. This is confirmed by formal statistical testing. In Appendix 3, we report the estimates of the difference in trends in our main outcome before the introduction of P4P. The time variable is the month of delivery, and it is interacted with a treatment indicator equal to one if the local facility will be among the treated facilities. We test six different specifications, as in Table 5 and we find that the pre-trends are never statistically significantly different.

Finally, in Appendix 1, we use a machine learning approach to identify the variables that best predict the decision to bypass by estimating a classification tree (Breiman et al., 1984). The advantage of this approach is that we can enter all of our variables in the calculation, instead of relying on subjective assessment of which should be in a regression, and follow clear classification rules to retain only the most important variables. By recursively partitioning the data along all the dimensions considered, the algorithm can detect which variables strongly associate with bypassing and can dismiss the variables considered unimportant.

5. Results

In this section, we start by presenting socio-demographic characteristics of the sample. We compare in particular the characteristics of the group that will get P4P to the group that will not, prior to the introduction of P4P. We also describe the evolution of bypassing by group and discuss the strongest correlates of bypassing.

In the second part we turn to the estimation of the P4P impacts on bypassing using a difference-in-difference estimator.

5.1. Descriptive statistics

Table 2 provides summary statistics of the variables used in the econometric analysis. It also serves to describe the sample prior to the introduction of P4P and to compare the group that will get P4P to the group that will not.

Most of the women completed primary education (75%), while few completed secondary education (9%). Half of the women are farmers, and a quarter mostly stay home to take care of the children. Almost a quarter are self-employed business women. They are 26 years old on average. Most of the women are married (66%), only 11% have a health insurance, and they are predominantly Muslim (76%).

The women live in households of 5 members on average, 25 km away from the nearest hospital and 12 km away from the nearest health centers (airline distances). Most do not own a motorized vehicle; only 9% have a motorbike and 1% have a car.

When it comes to local facility type, 67% are dispensaries, 24% are health centers, and the rest are hospitals, 88% are public and 12% are private.

There are some significant differences between groups. In Pwani, women are slightly more educated, less likely to work in a farm and more likely to stay home. They are also more likely to be married. The proportion of Muslims is larger in Pwani and the proportion of Catholics is lower. We will control for these variables (and the other variables shown in the table) when we estimate the effects of P4P.

Table 3 shows the rates of bypassing and where the women went to deliver. Overall, in the 2012–2015 period, 38% of the women bypassed their local facility and delivered in another facility instead. 32% went outside her ward to deliver, 13% went outside her district, and 8% went outside her region.

We observe that bypassing declined over time in both the treatment and control groups, but the reduction is significantly larger in the treatment group. Indeed, over the three years, bypassing declined by 13 percentage points in the treated areas while the reduction was only 5 percentage points in the control areas.

The extent and patterns of bypassing are indicated by the map in Fig. 2. The sampled women live in the catchment areas of 144 health facilities, each represented by a dot in the figure. The facilities are plotted over the map of Tanzanian regions. These women had delivered in 387 different facilities. Every delivery that did not take place at the local facility is represented by a grey arrow indicating where the woman actually delivered (arrows going to facilities far away have been cut to preserve the clarity of the figure).

The map illustrates the extent of bypassing, as there would be no arrows if no one bypassed. It also shows that deliveries tend to concentrate. Dar es Salaam in particular is attracting many women. The distribution of deliveries per facility is highly skewed with the top 5 facilities having 22% of the deliveries in the sample.

Among the available covariates, the strongest predictors of bypassing are type of local facility and distance to the nearest hospital. This is the conclusion from the machine learning estimation reported in Appendix 1 where we estimate a classification tree (Breiman et al., 1984) to identify the variables that allow us to best predict the decision to bypass. The result is also illustrated in Table 4, which reports the proportion of people who bypass by facility type and distance to hospital. When the local facility is a hospital, only 21% bypass. The bypassing rate is 33% for health centers and 42% for dispensaries. The distance to the nearest hospital is also strongly linked to bypassing. When the nearest hospital is less than 10 km away, the bypassing rate goes up to 73% for health centers and 60% for dispensaries.

Finally, bypassing is a costly decision for the women in our sample. In the survey we asked the women if they had to pay for the delivery, and how much, and if they also had to purchase medical supplies for their delivery, and how much.

Twenty percent of the women who bypass pay for their delivery, compared to 18 percent among those who don't bypass (p -value = 0.016) and they pay more on average (USD 5 instead of USD 3, p -value < 0.01). They are also more likely to pay for medical supplies (66.4% instead of 63.8%, p -value = 0.036), and they pay on average USD 4 instead of 3.55 (p -value < 0.01). Overall, women who bypass spend 30% more on their delivery (counting both the price of delivery and the purchase of supplies), without taking the transport cost into account.

5.2. Impact of P4P on bypassing

The difference-in-difference estimates of the impact of P4P on bypassing are shown in Table 5. The coefficient $P4P_{it} * ST_i$ is the effect after one year and the coefficient $P4P_{it} * LT_i$ is the effect after three years. The columns differ by whether we include covariates and facility fixed effects in the regressions.

The statistical analysis suggests that P4P reduces bypassing significantly. Three years after its introduction, P4P had reduced bypassing by 8 percentage points (17%). The estimate is very stable across specifications. In the short run, the estimated impact is somewhat smaller and not statistically significant (except in the first model with no controls, and then only at a 10% level). This could be because more time is needed until facilities make significant enough changes in the quality of their services to convince users to not bypass.

The standard errors are always clustered at the facility level in the upper part of the Table. The last two lines show the p -values calculated with wild-bootstrapped clustering at the district level. The p -values are lower when the standard errors are clustered at the facility level (and always lower than 0.05), when we cluster at the district level the p -values oscillate between 0.05 and 0.12, depending on the specification.

Appendix 2 includes the results of further analysis of the impact of P4P on the likelihood of delivering in one's own ward, district, and region. The results are consistent with those of Table 4: P4P has a strong positive impact on all those variables. Women in the treatment areas are more likely to deliver in facilities close to their place of residence after the implementation of P4P. The results also show that the impacts are larger on intra-ward bypassing, but that P4P also reduced bypassing to

Table 5
Difference-in-difference estimates of the impact of P4P on bypassing (percentage points).

Bypass	(1)	(2)	(3)	(4)	(5)	(6)
$P4P_{it}$	0.11** (0.05)		0.11** (0.05)		0.12*** (0.04)	
$P4P_{it} * ST_t$	-0.07* (0.04)	-0.06 (0.04)	-0.06 (0.04)	-0.06 (0.04)	-0.05 (0.04)	-0.06 (0.04)
$P4P_{it} * LT_t$	-0.08** (0.04)	-0.08** (0.04)	-0.09** (0.04)	-0.08** (0.04)	-0.08** (0.04)	-0.08** (0.04)
ST_t	0.01 (0.03)	0.01 (0.03)	0.02 (0.03)	0.01 (0.03)	0.02 (0.03)	0.01 (0.03)
LT_t	-0.05* (0.03)	-0.05* (0.03)	-0.05* (0.03)	-0.05* (0.03)	-0.06** (0.03)	-0.05* (0.03)
Observations	6192	6192	6158	6158	6158	6158
Controls 1	No	No	Yes	Yes	Yes	Yes
Controls 2	No	No	No	No	Yes	Yes
Facility fixed effects	No	Yes	No	Yes	No	Yes
Wild-cluster bootstrap p-value for:						
$P4P_{it} * ST_t = 0$	0.138	0.221	0.194	0.275	0.319	0.271
$P4P_{it} * LT_t = 0$	0.088	0.114	0.053	0.118	0.104	0.117

Note: Standard errors clustered at the facility level in parenthesis, p-value <0.1 *, <0.05 **, <0.01 ***. Controls 2 include all the variables listed in Table 1, the type of facility (dispensary, health center, hospital) and the facility ownership (public or private). Controls 1 include all the variables listed in Table 1, except for the distance to the nearest hospital and health center and whether the household owns a motorcycle or a car.

other regions.

6. Discussion

Our findings suggest that in addition to any impacts of P4P on performance indicators, reduced bypassing should be considered an additional positive impact of pay-for performance schemes. However, in assessing the generalizability of this finding to other settings with a P4P scheme it must be acknowledged that performance-based financing mechanisms are heterogeneous, and that the P4P scheme in Tanzania also has its peculiarities: First, it does not include incentives for structural quality of care (e.g., availability of equipment and drugs) at facility level except for health managers (Binyaruka and Borghi 2017), while other performance-based schemes in low- and middle-income countries typically include a large number of such performance indicators (Gergen et al., 2017; Kova c et al., 2020). Second, many performance-based schemes are built on a fee-for-service model, which implies that there is always an incentive to further improve performance. However, P4P in Tanzania has performance targets (with single or multiple thresholds) (Binyaruka et al., 2018b) which might have limited further improvement in performance beyond those targets in a given period. And finally, the proportion of performance payments going directly to health workers is at the high end (75%), compared to a range of 20–80% in other countries (Gergen et al., 2017). These differences do not however point in any clear direction as to whether we should expect larger or smaller effects on bypassing from P4P in Tanzania than from other performance-based financing mechanisms.

Our estimates may be biased if other interventions have differentially impacted bypassing in Pwani and the control areas during the course of this study. We are aware of one other intervention implemented in Pwani from 2012 onwards to improve the quality of maternal and newborn health services (Larson et al., 2019). The intervention led to some increase in deliveries in facilities, but since the intervention was implemented in 12 facilities only, it is unlikely to have significantly affected our findings. At the same time, a major new road was built in Pwani (between Bagamoyo and Msata). This made travelling easier and might have increased bypassing in Pwani, which would imply that the real impact of P4P on bypassing is bigger than what we have reported above.

7. Conclusion

Bypassing local health facilities is a common practice among patients in developing countries, including among expectant women bypassing their local service provider to deliver in their preferred facility. Women bypass to deliver in facilities with better equipment and more qualified staff. Subsequently, women who bypassed are found to be more likely to be satisfied with the care they received than the non-bypassers, including in Tanzania (Kruk et al., 2009, 2014). However, bypassing may create inefficiency in a country's health system, including overcrowding in hospitals and underused services in lower level health facilities. Bypassing also puts additional burdens on users through higher transportation expenses, more travel time, and often also higher out-of-pocket fees and other inconveniences.

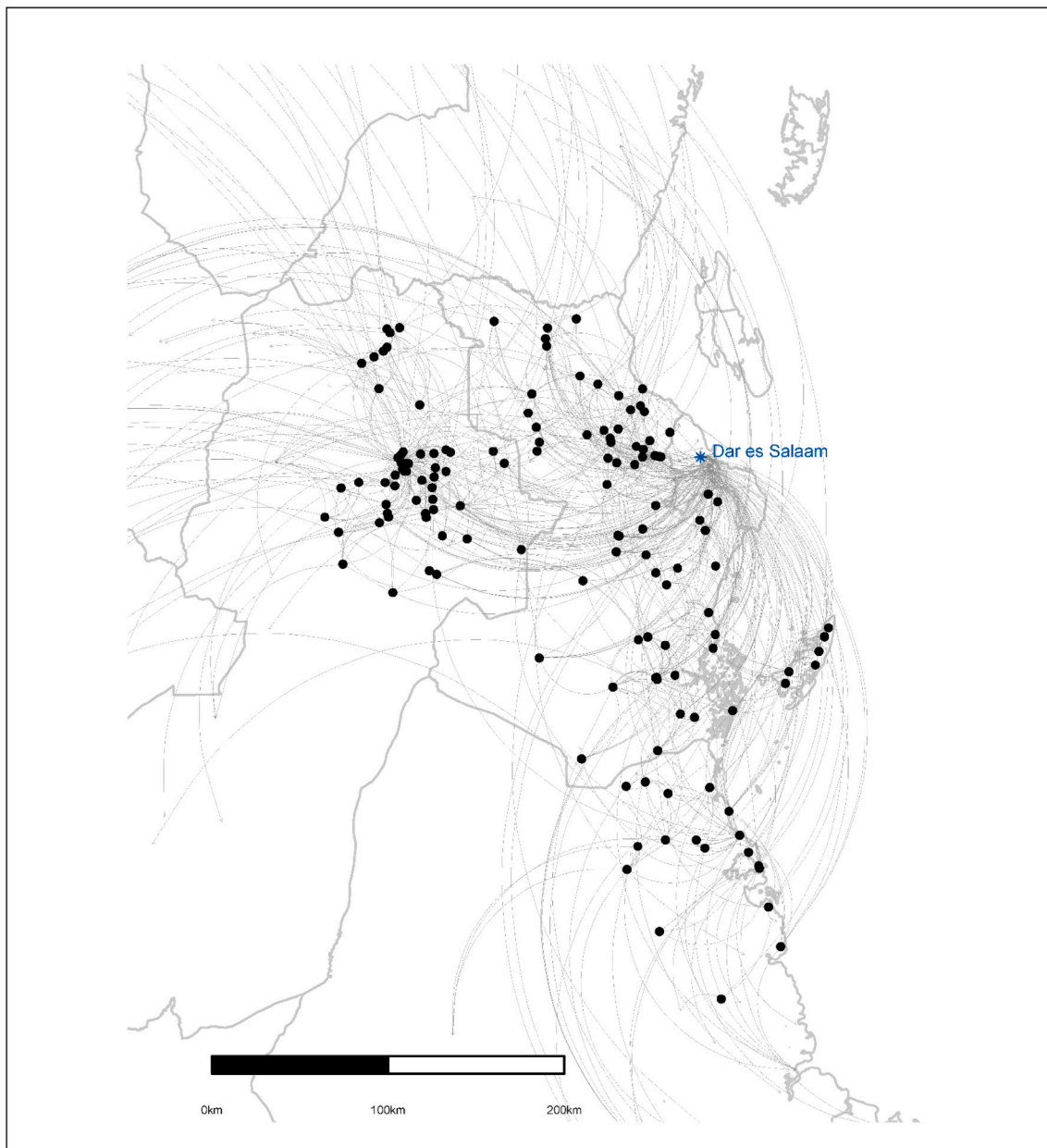
This study examined the impact of a payment-for-performance (P4P) scheme on bypassing in Tanzania. We used three rounds of survey data in the period 2012–2015 and a difference-in-difference regression model to assess the impact of P4P on bypassing. In addition, we implemented a machine learning approach to identify factors that predict bypassing.

We found that bypassing is happening in Tanzania as previously reported (Kanté et al., 2016; Kruk et al., 2009a,b). Overall, 38% of women in our study area bypassed their health service provider to deliver in another facility. We found that 70% of bypassing is predicted by the facility type and the distance to the closest hospital. Women are more likely to bypass if their local facility is a dispensary instead of a hospital or a health center. Women are less likely to bypass if they live close to a hospital. This pattern is probably a reflection of higher quality at higher level facilities.

We found that the P4P scheme significantly reduced bypassing among expectant women. On average, P4P reduced bypassing in the study area by 17% (or 8 percentage points) over three years. The result is robust for different model specifications. Since our study is the only study that reports the effect of P4P on bypassing practice in LMICs, other evaluations of P4P should consider assessing this effect.

This finding has important implication for the assessment of performance-based financing mechanisms, or even broader health policy intervention. Reduced bypassing is an additional benefit of performance-based financing, beyond any positive effects on performance.

It is also possible that not accounting for the effect on bypassing may lead to overestimation of other impacts. For example, several evaluations of pay-for-performance schemes have concluded that the scheme



Note: The map shows regional boundaries. Each dot represents one of the facilities included in the study. Each grey arrow represents a woman delivering outside of her catchment area and going where the arrow ends.

Fig. 2. Map of the health facilities. Note: The map shows regional boundaries. Each dot represents one of the facilities included in the study. Each grey arrow represents a woman delivering outside of her catchment area and going where the arrow ends.

increases facility-based deliveries (Renmans et al., 2016; Witter et al., 2012). To the extent that those evaluations are based on results from geographically limited study areas, the observed increase in local deliveries may partly reflect a reduction in bypassing (to facilities outside of the evaluator's sample) rather than a real increase in aggregate facility-based deliveries.

Credit author statement

Bezu: Writing – original draft, Writing – review & editing Binyaruka: Investigation, Data curation, Writing – review & editing Mæstad:

Writing – original draft, Writing – review & editing, Project administration, Funding acquisition Somville: Formal analysis, Writing – original draft, Writing – review & editing, Visualization

Acknowledgment

We thank the Research Council of Norway for funding this research and the Tanzania P4P program evaluation team for providing the data. We thank the three anonymous reviewers whose comments and feedback helped improve the quality of this paper.

Appendix 1. Factors predicting bypassing - a machine learning approach

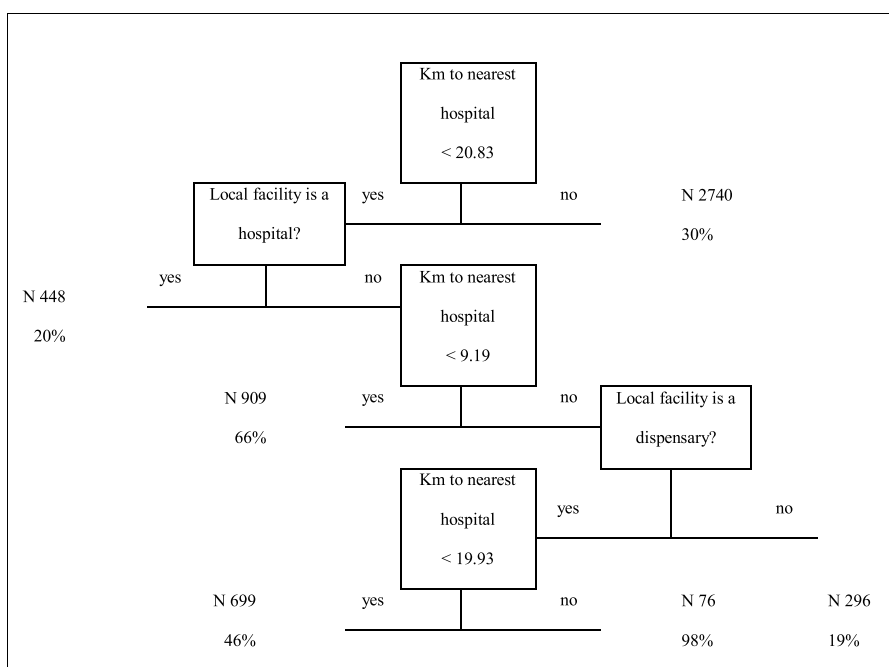
This appendix describes a machine learning approach used to identify factors that predict bypassing. We use this approach to estimate a classification tree (Breiman et al., 1984). The advantage is that we can enter all of our variables in the calculation, instead of relying on subjective assessment of which should be in a regression, and follow clear classification rules to retain only the most important variables.

We include 42 different variables in the estimation of the tree (see list of variables below). We deliberately leave the P4P indicator out of this analysis to focus on our other variables. We use recursive binary splitting to grow the tree and use the classification error rate to make the splits.¹

Only two variables were retained: the distance to the nearest hospital and the type of local facility (dispensary, health center, hospital). The tree is depicted in Fig. 2. It has six terminal nodes, a residual mean deviance of 1.207 and a misclassification error rate of 0.31. This means that we can correctly predict almost 70% of the bypass observation by using only these two variables.

The terminal nodes indicate the predicted outcome (bypass or don't bypass) and the number of observations that fall into that node. It also shows the proportion of observations who bypass in that node. For instance, the upper right terminal node tells us that if the facility is located more than 20.83 km away from a hospital then (i) people are predicted to not bypass, (ii) 2740 observations in our sample are more than 20.83 km away from a hospital and (iii) among those 2740 observations, 30% bypass.

We immediately identify from the tree that bypassing is lowest when the local facility is not a dispensary but a hospital or a health center, and when the closest hospital is not too far away.



Note: N indicates the number of observations in each node and the % refers to the proportion of those observations who bypass.

Fig. 3. Classification tree of bypassing. Note: N indicates the number of observations in each node and the % refers to the proportion of those observations who bypass.

List of variables used in the estimation of the classification tree:

- Highest level of education
- Occupation
- Religion
- Owns electricity
- Owns radio
- Owns TV
- Owns DVD
- Owns mobile phone
- Owns landline phone
- Owns iron
- Owns fridge
- Owns clock
- Owns sewing machine
- Owns table
- Owns sofa
- Owns cupboard

¹ We use R's package tree.s.

- Owns motorcycle
- Owns car
- Has a bank account
- Married
- Insured
- Religion
- Highest level of education
- Occupation
- Birth parity
- Any ANC visit
- Tested for HIV during ANC visit
- Treated for AIDS during ANC visits
- Prescribed malaria drug during ANC
- Taken at least 2 doses of malaria
- IPT drug from health facility
- Received tetanus injection during
- Received iron or folic acid during
- Type of local facility (dispensary, health center, hospital)
- Private or public local facility
- Km to nearest Health Center
- Km to nearest Hospital
- Number of household members
- Age at last birthday
- Number of ANC visits
- Probability of paying for ANC

Appendix 2P4P and bypassing across wards, districts and regions, and between public and private facilities

In this appendix, we test whether P4P affects bypassing not only within wards but also between wards, districts and regions. This is important if the costs of and risks associated with bypassing are increasing in the distance to the facility of delivery. We also test whether P4P affects the likelihood of delivering in a private or public facility. It is important to check that the observed shift is indeed from higher level to lower level public facilities rather than from private to public facilities.

The first three tables below show the difference-in-difference estimates of the impact of P4P on the likelihood of delivering in one’s own ward (Table 5), district (Table 6) and region (Table 7).

The results are consistent with those in Table 4; P4P has a significant negative impact on bypassing. The impact is largest on inter-ward bypassing; the likelihood of delivering in own ward increased by 8 percentage points.

The likelihood of delivering in own district increased by around 5 percentage points, and the likelihood of delivering in own region increased by 4 percentage points.

Table 6
The impact of P4P on the likelihood of delivering in her ward

Same ward	(1)	(2)	(3)	(4)	(5)	(6)
$P4P_{it}$	-0.06 (0.05)		-0.05 (0.05)		-0.06 (0.04)	
$P4P_{it} * ST_t$	0.06 (0.04)	0.06 (0.04)	0.05 (0.04)	0.05 (0.04)	0.05 (0.04)	0.05 (0.04)
$P4P_{it} * LT_t$	0.08** (0.03)	0.08** (0.03)	0.08** (0.03)	0.08** (0.03)	0.07** (0.03)	0.08** (0.03)
ST_t	-0.01 (0.03)	-0.01 (0.03)	-0.02 (0.03)	-0.01 (0.03)	-0.02 (0.03)	-0.01 (0.03)
LT_t	0.03 (0.03)	0.04 (0.03)	0.03 (0.03)	0.04 (0.03)	0.04 (0.03)	0.04 (0.03)
Observations	6192	6192	6158	6158	6158	6158
Controls 1	No	No	Yes	Yes	Yes	Yes
Controls 2	No	No	No	No	Yes	Yes
Facility fixed effects	No	Yes	No	Yes	No	Yes

Note: Standard errors clustered at the facility level in parenthesis, p-value <0.1 *, <0.05 **, <0.01 ***. Controls 2 include all the variables listed in Table 1, the type of facility (dispensary, health center, hospital) and the facility ownership (public or private). Controls 1 include all the variables listed in Table 1, except for the distance to the nearest hospital and health center and whether the household owns a motorcycle or a car.

Table 7
The impact of P4P on the likelihood of delivering in her district.

Same district	(1)	(2)	(3)	(4)	(5)	(6)
$P4P_{it}$	-0.07*** (0.02)		-0.07*** (0.02)		-0.07*** (0.02)	
$P4P_{it} * ST_t$	0.05* (0.03)	0.04 (0.03)	0.04 (0.03)	0.04 (0.03)	0.04 (0.03)	0.04 (0.03)
$P4P_{it} * LT_t$	0.05** (0.02)	0.06** (0.02)	0.05* (0.02)	0.06** (0.02)	0.05* (0.02)	0.06** (0.02)
ST_t	0.00 (0.02)	0.01 (0.02)	-0.00 (0.02)	0.00 (0.02)	-0.00 (0.02)	0.00 (0.02)
LT_t	0.03** (0.01)	0.03* (0.01)	0.04*** (0.01)	0.03** (0.01)	0.04** (0.01)	0.03** (0.01)
Observations	6192	6192	6158	6158	6158	6158
Controls 1	No	No	Yes	Yes	Yes	Yes
Controls 2	No	No	No	No	Yes	Yes
Facility fixed effects	No	Yes	No	Yes	No	Yes

Note: Standard errors clustered at the facility level in parenthesis, p-value <0.1 *, <0.05 **, <0.01 ***. Controls 2 include all the variables listed in Table 1, the type of facility (dispensary, health center, hospital) and the facility ownership (public or private). Controls 1 include all the variables listed in Table 1, except for the distance to the nearest hospital and health center and whether the household owns a motorcycle or a car.

Table 8
The impact of P4P on the likelihood of delivering in her region.

Same region	(1)	(2)	(3)	(4)	(5)	(6)
$P4P_{it}$	-0.07*** (0.02)		-0.07*** (0.02)		-0.07*** (0.02)	
$P4P_{it} * ST_t$	0.03 (0.02)	0.03 (0.02)	0.02 (0.02)	0.03 (0.02)	0.02 (0.02)	0.03 (0.02)
$P4P_{it} * LT_t$	0.04* (0.02)	0.04** (0.02)	0.03* (0.02)	0.04** (0.02)	0.04* (0.02)	0.04** (0.02)
ST_t	-0.00 (0.02)	-0.00 (0.02)	-0.00 (0.02)	-0.00 (0.02)	-0.00 (0.02)	-0.00 (0.02)
LT_t	0.02** (0.01)	0.02** (0.01)	0.03*** (0.01)	0.02** (0.01)	0.03*** (0.01)	0.02** (0.01)
Observations	6192	6192	6158	6158	6158	6158
Controls 1	No	No	Yes	Yes	Yes	Yes
Controls 2	No	No	No	No	Yes	Yes
Facility fixed effects	No	Yes	No	Yes	No	Yes

Note: Standard errors clustered at the facility level in parenthesis, p-value <0.1 *, <0.05 **, <0.01 ***. Controls 2 include all the variables listed in Table 1, the type of facility (dispensary, health center, hospital) and the facility ownership (public or private). Controls 1 include all the variables listed in Table 1, except for the distance to the nearest hospital and health center and whether the household owns a motorcycle or a car.

Table 9 reports the estimates of equation (1) where the dependent variable is an indicator equal to one if the woman delivered in a public facility, and zero if she delivered in a private facility. We find that P4P did not affect the likelihood of delivering in a public rather than a private facility.

Table 9
Difference-in-difference estimates of the impact of P4P on delivering in a public facility (percentage points)

In public	(1)	(2)	(3)	(4)	(5)	(6)
$P4P_{it}$	0.01 (0.04)		-0.00 (0.04)		-0.01 (0.03)	
$P4P_{it} * ST_t$	-0.01 (0.02)	-0.02 (0.02)	-0.01 (0.02)	-0.01 (0.02)	-0.01 (0.02)	-0.01 (0.02)
$P4P_{it} * LT_t$	-0.02 (0.02)	-0.01 (0.02)	-0.01 (0.02)	-0.01 (0.02)	-0.01 (0.02)	-0.01 (0.02)
ST_t	0.00 (0.01)	0.00 (0.01)	-0.00 (0.01)	0.00 (0.01)	0.00 (0.01)	0.00 (0.01)
LT_t	0.02 (0.02)	0.01 (0.01)	0.01 (0.02)	0.01 (0.01)	0.01 (0.02)	0.01 (0.01)
Observations	6192	6192	6158	6158	6158	6158
Controls 1	No	No	Yes	Yes	Yes	Yes
Controls 2	No	No	No	No	Yes	Yes
Facility fixed effects	No	Yes	No	Yes	No	Yes

Note: Standard errors clustered at the facility level in parenthesis, p-value <0.1 *, <0.05 **, <0.01 ***. Controls 2 include all the variables listed in Table 1, the type of facility (dispensary, health center, hospital) and the facility ownership (public or private). Controls 1 include all the variables listed in Table 1, except for the distance to the nearest hospital and health center and whether the household owns a motorcycle or a car.

Appendix 3. Test of parallel pre-trends

In this appendix, we test whether the trends in bypassing differ between treated and control facilities prior to the introduction of P4P. We do so by estimating equation (2) on the sample of women who gave birth before P4P was introduced:

$$Y_{it} = \theta + \mu_T P4P_{it} M_t + \pi P4P_{it} + \tau M_t + \omega_{it} \tag{2}$$

where Y_{it} is an indicator variable that takes the value one if woman i delivered in a facility other than her local facility at time t (except if the woman was referred to the other facility from her local one). $P4P_{it}$ is equal to one if the local health facility of individual i will participate in the P4P scheme, and M_t is a vector of indicators for the months in which i delivered. ω_{it} is a random error term which is assumed here to be identically and independently distributed. This equation is estimated by OLS.

The omitted base category for M_t is “delivered 3 months before”. In addition to the basic model, we also estimate the model after including other control variables. These controls include socio-economic background of the woman and facility level covariates. The control variables are listed in Table 2.

The results are shown in Table 10. The interaction between the time variable (month of delivery) and the future treatment status is never significantly different from zero, indicating that bypassing rates were evolving in parallel in both groups.

Table 10
Test of parallel trends in bypassing prior to the introduction of P4P.

Bypass	(1)	(2)	(3)	(4)	(5)	(6)
$P4P_{it}$	0.12 (0.22)		0.04 (0.20)		0.02 (0.14)	
$P4P_{it}^* - 4\text{ months}$	0.04 (0.23)	-0.00 (0.16)	0.11 (0.21)	0.05 (0.15)	0.15 (0.15)	0.05 (0.15)
$P4P_{it}^* - 5\text{ months}$	-0.09 (0.22)	-0.04 (0.16)	-0.02 (0.20)	-0.02 (0.15)	-0.00 (0.15)	-0.02 (0.15)
$P4P_{it}^* - 6\text{ months}$	0.09 (0.23)	0.04 (0.17)	0.17 (0.21)	0.07 (0.16)	0.14 (0.16)	0.07 (0.16)
$P4P_{it}^* - 7\text{ months}$	-0.03 (0.23)	0.00 (0.17)	0.07 (0.21)	0.05 (0.16)	0.09 (0.16)	0.05 (0.16)
$P4P_{it}^* - 8\text{ months}$	-0.07 (0.23)	0.00 (0.17)	0.04 (0.21)	0.05 (0.15)	0.05 (0.15)	0.05 (0.15)
$P4P_{it}^* - 9\text{ months}$	-0.05 (0.23)	0.05 (0.17)	0.02 (0.21)	0.09 (0.15)	0.09 (0.16)	0.09 (0.15)
$P4P_{it}^* - 10\text{ months}$	-0.06 (0.23)	0.07 (0.16)	0.04 (0.21)	0.13 (0.14)	0.11 (0.15)	0.13 (0.14)
$P4P_{it}^* - 11\text{ months}$	-0.06 (0.23)	0.06 (0.17)	0.01 (0.21)	0.11 (0.16)	0.05 (0.16)	0.10 (0.16)
$P4P_{it}^* - 12\text{ months}$	0.03 (0.23)	0.02 (0.17)	0.10 (0.21)	0.06 (0.15)	0.16 (0.14)	0.05 (0.15)
$P4P_{it}^* - 13\text{ months}$	-0.07 (0.28)	-0.09 (0.23)	0.05 (0.27)	-0.05 (0.22)	0.03 (0.23)	-0.05 (0.22)
$P4P_{it}^* - 14\text{ months}$	-0.02 (0.28)	-0.04 (0.24)	0.01 (0.27)	0.02 (0.22)	-0.06 (0.23)	0.01 (0.22)
Observations	2019	2019	2004	2004	2004	2004
Controls 1	No	No	Yes	Yes	Yes	Yes
Controls 2	No	No	No	No	Yes	Yes
Facility fixed effects	No	Yes	No	Yes	No	Yes

Note: Standard errors clustered at the facility level in parenthesis, p-value <0.1 *, <0.05 **, <0.01 ***. Controls 2 include all the variables listed in Table 1, the type of facility (dispensary, health center, hospital) and the facility ownership (public or private). Controls 1 include all the variables listed in Table 1, except for the distance to the nearest hospital and health center and whether the household owns a motorcycle or a car.

References

Akin, J.S., Hutchinson, P., 1999. Health-care facility choice and the phenomenon of bypassing. *Health Pol. Plann.* 14 (2), 135–151.

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.

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., 2015a. Effect of paying for performance on utilisation, quality, and user costs of health services in Tanzania: a controlled before and after study. *PLoS One* 10 (8). <https://doi.org/10.1371/journal.pone.0135013>.

Binyaruka, P., Patouillard, E., Powell-Jackson, T., Greco, G., Maestad, O., Borghi, J., 2015b. Effect of paying for performance on utilisation, quality, and user costs of health services in Tanzania: a controlled before and after study. *PLoS One* 10 (8), e0135013.

Binyaruka, P., Robberstad, B., Torsvik, G., Borghi, J., 2018a. Who benefits from increased service utilisation? Examining the distributional effects of payment for performance in Tanzania. *Int. J. Equity Health* 17 (1), 14. <https://doi.org/10.1186/s12939-018-0728-x>.

Binyaruka, P., Robberstad, B., Torsvik, G., Borghi, J., 2018b. Does payment for performance increase performance inequalities across health providers? A case study of Tanzania. *Health Pol. Plann.* 33 (9), 1026–1036. <https://doi.org/10.1093/heapol/czy084>.

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 (1), 80. <https://doi.org/10.1186/1748-5908-8-80>.

Fleming, L.C., Ansumana, R., Bockarie, A., Alejandre, J., Bangura, U., Jimmy, D.H., Jacobsen, K.H., 2016. Inpatient healthcare provider bypassing by women and their children in urban Bo, Sierra Leone. *The Pan African medical journal* 23. <https://doi.org/10.11604/pamj.2016.23.146.8706>.

Fritsche, G.B., Soeters, R., Meessen, B., 2014. Performance-based Financing Toolkit. The World Bank.

Gauthier, B., Wane, W., 2011. Bypassing health providers: the quest for better price and quality of health care in Chad. *Soc. Sci. Med.* 73 (4), 540–549. <https://doi.org/10.1016/j.socscimed.2011.06.008>.

Gergen, J., Josephson, E., Coe, M., Ski, S., Madhavan, S., Bauhoff, S., 2017. Quality of care in performance-based financing: how it is incorporated in 32 programs across

- 28 countries. *Glob. Health: Science and Practice* 5 (1), 90–107. <https://doi.org/10.9745/GHSP-D-16-00239>.
- Kanté, A., Exavery, A., Phillips, J., Jackson, E., 2016. Why women bypass front-line health facility services in pursuit of obstetric care provided elsewhere: a case study in three rural districts of Tanzania. *Trop. Med. Int. Health* 21 (4), 504–514.
- Karkee, R., Lee, A.H., Binns, C.W., 2015. Bypassing birth centres for childbirth: an analysis of data from a community-based prospective cohort study in Nepal. *Health Pol. Plann.* 30 (1), 1–7. <https://doi.org/10.1093/heapol/czt090>.
- Kovacs, R.J., Powell-Jackson, T., Kristensen, S.R., Singh, N., Borghi, J., 2020. How are pay-for-performance schemes in healthcare designed in low- and middle-income countries? Typology and systematic literature review. *BMC Health Serv. Res.* 20 (1), 291. <https://doi.org/10.1186/s12913-020-05075-y>.
- Kruk, M.E., Mbaruku, G., McCord, C.W., Moran, M., Rockers, P.C., Galea, S., 2009a. Bypassing primary care facilities for childbirth: a population-based study in rural Tanzania. *Health Pol. Plann.* 24 (4), 279–288. <https://doi.org/10.1093/heapol/czp011>.
- Kruk, M.E., Paczkowski, M., Mbaruku, G., de Pinho, H., Galea, S., 2009b. Women's preferences for place of delivery in rural Tanzania: a population-based discrete choice experiment. *Am. J. Publ. Health* 99 (9), 1666–1672. <https://doi.org/10.2105/AJPH.2008.146209>.
- Larson, E., Gage, A.D., Mbaruku, G., Mbatia, R., Haneuse, S., Kruk, M.E., 2019. Effect of a maternal and newborn health system quality improvement project on the use of facilities for childbirth: a cluster-randomised study in rural Tanzania. *Trop. Med. Int. Health* 25 (5), 636–646.
- Leonard, K.L., Mliga, G.R., Haile Mariam, D., 2002. Bypassing health centres in Tanzania: revealed preferences for quality. *J. Afr. Econ.* 11 (4), 441–471. <https://doi.org/10.1093/jae/11.4.441>.
- MacKinnon, J.G., Webb, M.D., 2017. Wild bootstrap inference for wildly different cluster sizes. *J. Appl. Econ.* 32, 233–254.
- MacKinnon, J.G., Webb, M.D., 2018. The wild bootstrap for few (treated) clusters. *Econom. J.* 21, 114–135.
- 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.
- Rao, K.D., Sheffel, A., 2018. Quality of clinical care and bypassing of primary health centers in India. *Soc. Sci. Med.* 207, 80–88.
- 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 Pol. Plann.* 31 (9), 1297–1309.
- Roodman, David, Morten Ørregaard Nielsen, James G. MacKinnon, Webb, Matthew D., 2019. Fast and wild: bootstrap inference in stata using boottest. *STATA J.* 19 (1), 4–60.
- Sabde, Y., Chaturvedi, S., Randive, B., Sidney, K., Salazar, M., De Costa, A., Diwan, V., 2018. Bypassing health facilities for childbirth in the context of the JSY cash transfer program to promote institutional birth: a cross-sectional study from Madhya Pradesh, India. *PLoS One* 13 (1), e0189364.
- Salazar, M., Vora, K., Costa, A.D., 2016. Bypassing health facilities for childbirth: a multilevel study in three districts of Gujarat, India. *Glob. Health Action* 9, 32178. <https://doi.org/10.3402/gha.v9.32178>.
- Shah, R., 2016. Bypassing birthing centres for child birth: a community-based study in rural Chitwan Nepal. *BMC Health Serv. Res.* 16 (1), 597. <https://doi.org/10.1186/s12913-016-1848-x>.
- WB, 2018. RBF Health: Projects. <https://www.rbfhealth.org/projects>. (Accessed 6 March 2019).
- Witter, S., Fretheim, A., Kessy, F.L., Lindahl, A.K., 2012. Paying for performance to improve the delivery of health interventions in low-and middle-income countries. *Cochrane Database Syst. Rev.* (2).