**Title:**

**Understanding causal pathways within health systems policy evaluation through mediation analysis: an application to payment for performance (P4P) in Tanzania**

Authors:

Laura Anselmi, Centre for Health Economics, University of Manchester, Manchester

Peter Binyaruka, Ifakara Health Institute Dar es Salaam, Tanzania

Josephine Borghi, Department of Global Health and Development, London School of Hygiene and Tropical Medicine, London

Corresponding author: Laura Anselmi (laura.anselmi@manchester.ac.uk)

**Abstract**

Background: The evaluation of payment for performance (P4P) programmes has focused mainly on understanding contributions to health service coverage, without unpacking causal mechanisms. To test the causal pathways through which P4P schemes may (or may not) influence maternal care outcomes.

Methods: We used data from an evaluation of a P4P programme in Tanzania. Data were collected from a sample of 3,000 women who delivered in the 12 months prior to interview, 200 health workers at 150 health facilities from seven intervention and four comparison districts in Tanzania in January 2012 and in February 2013. We applied causal mediation analysis using a linear structural equation model to identify direct and indirect effects of P4P on institutional delivery rates and on the uptake of two doses of an anti-malarial drug during pregnancy. We first ran a series of linear difference-in-difference (DiD) regression models to test the effect of P4P on potential mediators, which we then included in a linear DiD model evaluating the impact of P4P on the outcome. We tested the robustness of our results to unmeasured confounding using semi-parametric methods.

Results: P4P reduced the probability of women paying for delivery care (-4.5 percentage points) which mediates the total effect of P4P on institutional deliveries (by 48%) and on deliveries in a public health facility (by 78%). P4P reduced the stock-out rate for some essential drugs, specifically Oxytocin (- 36 percentage points), which mediated the total effect of P4P on institutional deliveries (by 22%) and deliveries in a public health facility (by 30%). P4P increased kindness at delivery (5 percentage points), which mediated the effect of P4P on institutional deliveries (by 48%) and on deliveries in a public health facility (by 49%). P4P increased the likelihood of supervision visits taking place within the last 90 days (18

percentage points), which mediated 14% of the total P4P effect on the proportion of women receiving two antimalarial doses during antenatal care (IPT2).  Kindness during deliveries and the probability of paying out of pocket for delivery care were the mediators most robust to unmeasured confounding.

Conclusion: The effect of P4P on institutional deliveries is mediated by financing, and human resources factors and uptake of anti-malarials in pregnancy is mediated by governance factors. Further research is required to explore additional and more complex causal pathways.

**Background**

**Introduction**

Much of the focus of programme evaluation has been on outcome measurement and finding out whether or not a programme works, with randomised trials being considered to be the gold standard for causal inference [1]. However, when dealing with complex interventions, it is not enough to know whether they work, we also need to understand how they work [2]. Process evaluation enables us to get at the how and why questions, and unpack the 'black box' surrounding complex interventions and is increasingly promoted within evaluation research [3, 4].

One of the core functions of process evaluation is to shed light on causal mechanisms, or the process through which a programme influences an outcome [2, 5]. Examination of causal mechanisms is necessary in order to understand why a programme worked, or why it did not work, and whether the underlying theory was sound. It enables theory building and enhances intervention design [6], and can support the plausibility of outcome effects being associated with the intervention in a non-randomised study [7], increasing the internal validity of evaluation in social sciences [1, 5].

Practically, causal mechanisms can be identified by specifying intermediate outcomes or variables, referred to as mediators, that are on the causal pathway between the intervention and the outcome [6, 8]. The approach used to investigate causal mechanisms involves the estimation of causal mediation effects, or the breakdown of total causal effects into indirect effects (the effect of the intervention on the outcome that passes through the mediator) and

the direct effect (the effect of the intervention on the outcome through all other pathways) [9]. Causal mediation analysis has been employed to test change pathways within the evaluation of public health programmes, using individual level psychological [9-12] or physical characteristics [13], that may affect behaviour change outcomes. A recent study also considered the effect of community as well as individual level mediators [14]. To the best of our knowledge, to date, there has been only one study [15] considering the health system mediators which are relevant to the evaluation of interventions aimed at strengthening health systems.

Payment for performance (P4P) is an example of a programme which operates at the health system level with the aim of improving the quality and use of health services to enhance population health outcomes. P4P involves the payment of financial rewards to health workers (and sometimes to health facilities) based on their achievement of pre-specified performance targets. P4P has been widely used in the United Kingdom and the United States [16] and increasingly in low and middle income countries [17].

There is a growing body of evidence evaluating the impact of P4P [18]. Findings show that overall P4P has a positive effect on targeted service outcomes [19], although the evidence base in low-income settings is limited to a small number of studies [17, 20-25]. There has been less attention to the processes by which these outcomes are achieved, particularly in low and middle income settings [17, 26]. Three studies examined the implementation process challenges facing a P4P programme [27-29] and evaluations are increasingly looking at intermediate outcomes that may have affected service delivery [30][15]. However, existing studies do not conclusively shed light on the pathways through which P4P achieves

outcomes. Either they do not formally test the pathways, either they test them on a limited number of mediators [15].

The overall aim of the paper is to test the causal pathways through which payment for performance may (or may not) influence the utilisation of maternal health services. A previous study in Tanzania evaluated the impact of P4P on service use, quality, equity and health worker motivation over a 13 month period from January 2012 to February 2013 using linear difference-in-difference analysis [31]. The evaluation found a significant and positive effect on two of the targeted indicators:  an increase of 8.2 percentage points (CI: 3.6 to 12.8) in institutional deliveries, of 6.5 percentage points (CI: 1.3 to 11.7) in the rate of deliveries in public facilities and of 10.3 percentage points (CI: 4.3 to 16.2) in the proportion of women receiving two antimalarial doses during antenatal care [21].  In this paper we extend this analyses to examine the mediators of programme effect and to test the causal pathway to improved outcomes.


**Study setting**

In 2011 the Ministry of Health and Social Welfare of the Republic of Tanzania, introduced a P4P scheme in the Pwani region, with initial payments being made in mid-2012. The P4P scheme comprised four main components.

   1) P4P provided financial bonuses to health facilities and district and regional health managers based on achievement of maternal and child health (MCH) performance targets related to service coverage and quality of care. The targets were either for specific services (e.g. institutional delivery; postnatal care; family planning) or for care provided during a service (e.g. two doses of Intermittent Preventive Treatment for malaria (IPT2) during

antenatal care and HIV treatment for HIV positive pregnant women).  At the facility level, at least three-quarters of the bonus were distributed among health workers.  The health worker incentive represented about 10% of the average health worker monthly salary (about USD 30 per month). District and regional managers received bonus payments based on the performance of facilities in their district and region.

2) The remaining 25% of the bonus went to the health facility and could be invested in drugs, supplies or facility improvements. This represents roughly 4% of their average budget.

3) Supervision was more frequent as facility performance data were verified every six months by national, regional and district stakeholders, whereby achievements of targets, established by the Central Ministry of Health and Social Welfare, were measured and bonuses paid.

4) Primary care facilities had to open bank accounts in order to receive bonus payments, and could retain cost sharing revenue in these accounts, whereas before such funds were held at district level.  Health Facility Governing Committees, comprised of health workers and community members are responsible for managing facility resources, including P4P bonus payments, and representatives were to be present to withdraw bonus funds from the bank. However, the community members on the committee were not eligible for bonus payments.


**Conceptual framework**

Our analysis was guided by a theory of change for how P4P would affect the health system to improve outcomes, and a set of underlying assumptions about the change processes involved (Figure 1).

7

(Figure 1)

The increase in facility revenue from performance payments, together with financial autonomy resulting from facility-level bank accounts, may generate the need for increased accountability of resource allocation and use at the facility level, potentially stimulating health facility governing committees that are otherwise inactive, and improving relations between providers and communities [32]. Greater resources and more accountability over their use is expected to lead to improved availability of equipment, drugs and medical supplies at the facility, especially in relation to targeted services. P4P is also expected to directly affect supervision linked to the process of performance verification done by health care managers, as this results in more frequent contact between providers and managers, who examine registers and work conduct at the facility [33].

The direct financial incentives to health workers that are tied to service delivery, coupled with the changes in resource availability at the facility, and supervision practices are expected to impact on health workers' job satisfaction and increase motivation to adhere to clinical guidelines [34, 35] and treat patients respectfully. Health worker knowledge may also increase, through investment in training to improve skills linked to incentivised services, or through reallocation of staff to under-resourced or poor performing facilities. To stimulate service use and achieve targets, health workers may undertake more outreach activities and/or reduce user fees and/or be more likely to enforce exemptions for vulnerable groups [36, 37] or encourage enrolment in community health insurance, as this generates additional revenue for the facility.

We identified a set of indicators to measure each of the steps on the causal pathway (Table 1). The indicators were measured in household, facility and health worker surveys (Table 1).

A full discussion of the effects of P4P on the availability of medical supplies and drugs and on governance of facilities is presented elsewhere [33, 38].

(Table 1 here)

**Methods**

**Data sources**

Surveys were undertaken in all seven districts in Pwani region where P4P is being implemented and four neighbouring comparison districts with no P4P, with 75 facilities being sampled in each of the study arms, comprising six hospitals, 16 health centres, 11 non-public dispensaries and 42 dispensaries. A health facility survey was conducted at all facilities and 1-2 health workers per facility were interviewed. Interviews were conducted with women who had delivered in the past 12 months sampled within the catchment area of the facilities – a total of 3,000 women per round. Baseline data collection was conducted between January and March 2012 and endline data was collected 13 months later [31]. All data could be linked at the facility level [21].

**Data Analysis**

We used causal mediation analysis to identify steps on the causal pathway to the two significant outcomes in the main evaluation (delivery in a health facility and uptake of two doses of anti-malarial drugs during pregnancy). We also considered potential mediators of a third outcome, delivery in a public health facility, as we thought that mediators may differ

within public compared to non-public facilities.  We assessed mediation by applying the Linear Structural Equation Model (LSEM) Baron and Kenny [39] [6].  We estimated a single-mediator model to identify the effect of P4P on mediators and the effect of the latter on institutional deliveries and coverage of anti-malarials during pregnancy.  We followed a four step process to assessing mediation.

### Step1: Estimating the impact of P4P on outcomes

First, we replicated the analysis previously carried out by Binyaruka, Patouillard [21] to evaluate the effect of P4P on the selected outcomes using a linear difference-in-difference regression model:

$$Y_{ijt} = \beta_0^1 + \beta_1^1(P4P_j \times \delta_t) + \beta_2^1\delta_t + \beta_3^1 X_{ijt} + \gamma_j + \varepsilon_{ijt}^1 \qquad (1)$$

where $i$ is the sample of women who gave birth in the 12 months prior to the interview in the catchment area of facility $j$ at time $t$. $Y_{ijt}$ is a dummy taking value 1 if the service was received by a woman and 0 otherwise. $P4P_j$ is an indicator of whether P4P was implemented in the area where the woman was sampled from. We included facility fixed effects ($\gamma_j$) to control for facility-level unobserved time invariant characteristics and a dummy variable taking the value of 0 at baseline and 1 at endline ($\delta t$) to account for year fixed effects.  We also controlled for individual-level characteristics (education, religion, marital status, occupation, age, number of pregnancies) and household characteristics (insurance status, number of household members, household head education, and wealth based on ownership of household assets and housing particulars) that are known to affect outcomes ($X_{ijt}$). The effect of P4P on outcomes was estimated by $\beta_1^1$. Standard errors were clustered at the health facility level.

***Step 2: Identifying mediators***

Second, we tested for the effect of P4P on each of the potential mediators identified within the theory of change (Table 1).

As in (1) we used a linear difference-in-difference regression model:

$$M_{ijt} = \beta_0^2 + \beta_1^2(P4P_j \times \delta_t) + \beta_2^2\delta_t + \beta_3^2 X_{ijt} + \gamma_j + \varepsilon_{ijt}^2 \quad (2)$$

where $M_{ijt}$ is the potential mediator, and $\beta_1^2$ indicates the effect of P4P on the mediator. All mediators were measured at the health facility level. Items collected through the health worker survey were either averaged across health workers in the same facility, when they concerned individual judgement (satisfaction and motivation), or the highest value was retained when they concerned health facility characteristics (time and content of last supervision visit). Indicators of price, satisfaction with the service received and kindness during delivery, that were measured at the individual level, were averaged across women in the same facility catchment area, excluding the woman herself, to avoid direct reverse causality, simultaneity bias, and to test how the prevalent reported price and quality affected individual choice[40]. Although some mediators were measured at the individual level and some at the health facility level, Equation 2 was estimated at the individual level for all mediators, for comparability with step 1 and step 3. Sstandard errors were clustered at the health facility level.

**Step 3. Identifying direct and indirect causal effects**

Third, we evaluated the effect of P4P on the outcomes of interest, by re-estimating equation 1, including the potential mediators $M_{ijt}$ identified in step 2:

$$Y_{ijt} = \beta_0^3 + \beta_1^3(P4P_j \times \delta_t) + \beta_2^3\delta_t + \beta_3^3 X_{ijt} + \beta_4^3 M_{ijt} + \gamma_j + \varepsilon_{ijt}^3 \quad (3)$$

We ran the analysis separately for each maternal care outcome Yi and for each potential mediator $M_{ijt}$ identified in step 2.    If the estimated coefficient of $M_{ijt}$ ($\beta_4^3$) was significant and the effect of P4P was reduced compared to that estimated in (1) ($\beta_1^3$ was smaller than $\beta_1^1$), we can infer that the effect of P4P on $Y_{ijt}$ is mediated through $M_{ijt}$. For each set of outcome and mediators, $\beta_1^3$ measures the direct effect of P4P on $Y_{ijt}$, while the mediated (or indirect) effect was calculated as the product between $\beta_1^2$ and   $\beta_4^3$, and its significance verified by calculating their bootstrapped standard errors [6]. These analyses were run at the individual level. As for equations 1 and 2, equation 3 was estimated using a linear probability model and standard errors were clustered at the health facility level.

**Step 4: Sensitivity analysis**

The identified mediators can only be considered to be 'on the causal pathway' (enabling the measurement of causal mediation effects) under a set of two assumptions, referred to as 'sequential ignorability': first the intervention assignment is independent of outcomes and mediators; and second the observed mediator is independent of outcomes given the actual treatment status and pre-treatment confounders (there are no unmeasured confounders that affect both the mediator and the outcome) [41].

The first part of the assumption is satisfied if the treatment is assigned randomly, or assumed to be random given the pre-treatment covariates [8]. The use of difference in difference regression methods allows us to control for factors that may lead to the endogenous assignment of the intervention subject to the assumption of parallel trends. We verified that the pre-intervention trends in a selection of mediators and outcomes were parallel between intervention and comparison areas [21].

The second assumption is still required to identify the causal effect of the mediator on the outcome and cannot be formally tested [8, 42]. To address this, Imai et al. (2010) [43] propose a measure of the sensitivity to unmeasured confounding, Since the level of correlation between $\varepsilon_{ijt}^2$ and $\varepsilon_{ijt}^3$ reflects the presence of unobservables affecting both the mediator and the outcome, the level at which the mediation effect would be zero [44] provides an indication of how plausible the assumption is. The smaller the level of correlation, the less plausible the assumption. Imai et al. (2010) [43] develop their approach in a potential outcome framework using a semi-parametric approach for the identification of treatment direct and mediated effects. We set the prediction of potential outcomes to be based on equations 1, 2 and 3 used in the LSEM, so that we could compare results and the sensitivity analysis would apply to the results obtained. The sensitivity analysis provides the coefficient of correlation (rho) between $\varepsilon_{ijt}^2$ and $\varepsilon_{ijt}^3$ at which the average causal mediation effect (ACME) equals 0.

The LSEM approach to mediation analysis requires no interaction between the intervention and the average causal mediation effect, in other words that the average causal mediation effect is equivalent in intervention and comparison areas. We test this assumption by introducing an interaction term between treatment and mediator in equation (3) and testing its significance.

Since the outcomes are observed at the individual level, but the P4P scheme is implemented at the health facility level, we test the sensitivity of our results to the level at which the analysis is carried out by re-estimating equations 1 to 3 on the outcomes measured at the health facility level, based on averages of individuals within the facility catchment area.

We tested for clustering at the district level using a bootstrapping procedure which is recommended when the number of clusters is small [45, 46]. Since multiple hypothesis testing may lead to false rejection of the null hypothesis, we also applied a modified Bonferroni correction to the significance level that accounts for the correlation between the tested outcomes [30]. All statistical analyses were conducted using STATA 14.

## Results

### Descriptive statistics

The intervention and comparison groups are similar at baseline in relation to most of the outcomes and mediators considered (Table 2). However, in general, the comparison group performs slightly better than the intervention group in relation to the mediators.

(Table 2 here)

### Mediation analysis

As has been previously reported, there was a positive and significant effect of P4P on the rate of institutional deliveries (an 8.2 percentage point increase, CI: 3.6 to 12.8), on the rate of deliveries in public health facilities (a 6.5 percentage point increase, CI: 1.3; 11.7) and on the uptake of two doses of anti-malarial drugs (a 10.3 percentage point increase CI: 4.4; 16.1) [21] (Table 3). The effect of P4P was tested on all potential mediators in Table 2, but results are reported only for those significantly affected by P4P (Table 3).

P4P led to an increased availability of resources at the facility, notably a reduction in the disruption of services due to broken equipment (by -14.9 percentage points, CI: -29.3 to -

0.4); a reduction in the stock-out rate of essential medical supplies (by -14.8 percentage points, CI: -24.8 to -4.9) and drugs (by -17.2 percentage points, CI: -26.8 to -5.8), particularly those used during delivery including Oxytocin (by -36.2 percentage points, CI: -55.9 to -16.4) and Ergometrin (by -26.1 percentage points, CI: -48.2 to -4.0). As previously reported in [33], P4P resulted in more frequent supervision. There was a reduction in the probability of having received the last district or regional supervision more than 90 days prior to the survey (by -18 percentage points,CI: -32.0 to -4.0). P4P resulted in a significant increase in health worker knowledge (by 18.8 percentage points, CI: 10.4 to 27.2) and improved patient-provider interactions, measured by patient perceptions of provider kindness during deliveries (by 4.3 percentage points, CI: -0.4 to 9.0). P4P led to a reduction in user costs, measured as the reduced probability of paying out-of-pocket for institutional delivery by women in the catchment area of the facility (by 4.5 percentage points, CI: -9.5 to 0.6)) (Table 3). No effect was found on the remaining indicators on the causal pathway, notably, health worker motivation, outreach activities, and insurance enrolment.

However, among all the potential mediators identified, only a limited number of them significantly mediated the effect of P4P on the outcomes of interest (Table 4). The coefficient associated with P4P reported in Table 4 represents the direct programme effect when controlling for a given mediator; where this is less than that reported in the analysis without mediators, there is evidence of mediation. The indirect effect of P4P on the outcome, or the effect which passes through a given mediator, is calculated by interacting the coefficient associated with the mediator of interest in equation 3 with the effect of P4P on the same mediator in equation 2. The estimates of the direct and indirect (through the selected mediators) effects of P4P on outcomes are reported in Table 5 along with the results of sensitivity to sequential ignorability assumption (rho at which ACME equals 0).

(Table 4 here)

(Table 5 here)

The probability of paying for delivery and the perceived kindness of health workers during delivery mediate the effect of P4P on institutional deliveries, and the stock out rate of Oxytocin mediates the effect of P4P on deliveries in public facilities. When these are included as mediators, P4P has no significant direct effect anymore on the outcome (Table 4).

The reduction in the proportion of women who paid for delivery mediates 48% of the effect of P4P on institutional delivery and 78% of the effect of P4P on delivery in a public health facility (Table 5). The reduction in the stock-out rate of Oxytocin mediates 22% of the total effect of P4P on institutional delivery and 30% of the total programme effect on delivery in a public health facility (Table 5, columns 1 and 2). The kindness of providers during delivery mediates 48% of the total effect of P4P on institutional deliveries and 49% on deliveries in public facilities. The increase in the timeliness of supervision mediates 15% of the effect of P4P on the uptake of two doses of anti-malarial drugs during anti-natal care (Table 5, column 3), but did not mediate the effect of P4P on institutional deliveries. Uptake of two doses of antimalarial drugs did not appear to be a significant mediator of the effect of P4P on institutional deliveries (Table 4, columns 1 and 2), but it was borderline significant for deliveries in a public health facility.

**Sensitivity analysis**

16

The sensitivity analysis (Table 5) indicates that little correlation between the error terms of equations (2) and (3) (correlation coefficients ranging from 0.02 to 0.04) would be sufficient to reduce the mediated effect to zero for most mediators. However, a higher correlation coefficient would be required to reduce to zero the indirect effect of P4P through a reduction of payment at delivery and an increased health worker kindness, on institutional delivery (correlation coefficients 0.23 and 0.20 respectively) and on delivery in a public health facility (correlation coefficients 0.25 and 0.16 respectively).

When carrying out the analysis at the health facility level, the stock out rate of Oxytocin and the perceived kindness of health workers at delivery were still mediating the effect of P4P on institutional deliveries, while the proportion of women who paid for delivery was mediating the effect on deliveries in public facilities. However, the other mediators identified were no longer significant and no mediators for the uptake of two doses of anti-malarial drugs during antenatal care were identified. New mediators were also identified. For example, the health worker satisfaction with local leaders became mediator of delivery in a public health facility. None of the indirect effects were significant, however, as a consequence of the reduced statistical power due to the smaller number of observations.

A number of other sensitivity analyses were carried out. We tested for significance of the interaction between treatment and mediator in equation (3) and found no significant effect indicating that the average mediation effect is equivalent in treated and non-treated areas. We identified the same set of potential mediators when we tested for the effect of P4P correcting standard errors for clustering at the district level. When we adjusted the level of significance to account for multiple outcome testing, the reduction in the stock out rate of Oxytocin was the only mediator that remained significant.

**Discussion**

Causal mediation analysis has been put forward as an approach to understand causal mechanisms within process evaluation [2]. However, to date there is very little empirical evidence of its application within the evaluation of complex health interventions. Building on an existing impact evaluation, we set out to test the causal pathways through which P4P affected maternal care outcomes using causal mediation analysis. While our finding of P4P effects on core maternal outcomes is partly consistent with previous evaluation studies in Rwanda and Burundi [20, 22, 30, 47], ours is the first to formally test the pathways through which P4P affects outcomes.

Like in previous studies [15], we found that P4P affect the inputs to health care delivery available in the health facilities. However, we tested a wider range of mediators consistently to the hypothesis of an underlying theory of change on pre and post-intervention data, and found that they mediate a relevant proportion of the effect of P4P on maternal care service use.

Compared to other mediators, reductions in the probability of paying out of pocket and increased provider kindness during delivery mediated the largest share of P4P effect on institutional deliveries overall and in public facilities, and mediation effects were more robust to unmeasured confounding. The reduction in the rate of stock-out of Oxytocin (administered to induce or support labour and to manage the third stage of labour reducing the risk of postpartum haemorrhage [48]) also mediated 22% of the effect on institutional delivery (up to 30% in public health facilities), but the correlation coefficient was very low (0.04) suggesting the results are highly sensitive to unmeasured confounding. The effect of

18

P4P on the availability of Oxytocin is, however, consistent with our theory of change. The increased availability of Oxytocin may be due to additional resources made available through P4P to facilities, and/or greater communication with district authorities resulting from more frequent supervision. The increased availability of Oxytocin may be appreciated by women as a marker for quality of obstetric care, and management of bleeding, thereby influencing demand [49], though there is no literature highlighting women's preference for induction [50].

Although women are supposed to be exempt from payment for deliveries in public facilities, often such exemptions are incompletely enforced [51]. Also, when drugs are out of stock, women have to pay for them at private pharmacies. The mediation effect of the probability of paying for care is consistent with providers making a concerted effort to enforce exemptions to attract women to facilities for their delivery [36]. The probability of payment is also likely affected by the reduction in stock out of drugs related to delivery such as Oxytocin, that no longer have to be paid for privately by patients.

Previous results [21] also reported an effect of P4P on women's reports of health worker kindness during delivery. Health worker kindness, measured as the mean rank reported by other women in the same health facility catchment area, was found a significant mediator, suggesting that increased institutional deliveries could be due to expectations of higher quality of the service provided. This is consistent with our theory of change, whereby health workers modify their interactions and behaviour with patients to make services more attractive, to increase demand so as to meet the performance targets. Literature from a range of settings has highlighted the importance of provider attitude and kindness for women's demand for care at birth [52, 53]. Improved timeliness of supervision, which we

believe may be associated with the verification activities carried out as part of the P4P programme, significantly mediated 13% of the effect of P4P on the uptake of two doses of anti-malarials during pregnancy. This indicate that increased monitoring and coaching may lead health workers to improve health care process.

Referring back to our initial theory of change, the mediators which explained the largest share of total programme effect, and were most robust to unmeasured confounding, rely primarily on health worker response to the direct financial incentive. However, we did not find evidence of P4P increasing motivation, which was identified as a necessary precursor to behaviour change within the theory of change. This could be due to the limited sample size for the health worker survey, or invalid measurement of the underlying motivation construct, which was proxied as job satisfaction. It is also possible that health workers respond to incentives by changing their behaviour without experiencing greater job satisfaction. Our results also suggest that other components of the P4P programme were relevant to outcome achievements, notably the additional facility resources used to procure drugs and supplies, and more timely supervision, though these effects were less robust to unmeasured confounding. We found less evidence of the effect of financial autonomy. Ultimately such information is useful as it helps identify the programme's most effective components and 'levers' of demand.

In addition to illustrating likely mediators on the pathway to outcomes, our analysis also illustrates the application of causal mediation analysis to the evaluation of a health systems intervention, such as P4P, and specifically the consideration of health systems mediators, rather than individual level mediators, related to behaviour change. However, doing so does raise practical challenges.

First, when mediators operate at the level of the provider or health facility and outcomes are measured at the household or individual level it is unclear at which level the analysis should be carried out. We carried out the analysis at the individual level, as we were interested in the pathways to population outcomes, but we assessed the robustness of results to analysis at the facility level, and we found this did affect some of the mediators. The difference in results is in part due to the weighing proportional to the size of the sample from the health facility catchment areas, which varies from facility to facility, as well as to the reduced sample size and resulting lower statistical power.

Second, randomised trials of health systems interventions are often difficult to implement, and quasi-experimental methods may be the only way to assess causal effects, as in this study. However, to date causal mediation analysis has only been used alongside randomised controlled trials. We demonstrated its use within difference–in-difference analysis. This approach rests on the assumption of parallel trends between intervention and comparison groups in relation to outcomes as well as mediators. While we were able to assess pre-intervention trends in outcomes we could do it for only some mediators [21, 33, 38]. In the future, researchers should seek to gather pre-intervention time series data on outcomes as well as mediators. As in the main impact evaluation [21], we used a linear regression model to estimate P4P effects which allows us to use linear structural equation modelling to generate our estimate of mediation effect, although our outcomes and many of our mediators are binary. We had, however, previously demonstrated the robustness of our results to the use of non-linear models [21].

The selection of mediators for inclusion in the analysis was limited to those available within the surveys, so that potentially relevant mediators, such as level of funding available at the

facility level, could not not be included. Our approach relies on the accurate measurement of potential mediators and, where possible, we used tools that had been tested and applied in previous research to test for potential bias. Future studies should consider using qualitative methods alongside the impact evaluation, to validate and help explain mediators identified as being significant through mediation analysis.

The application of causal mediation analysis to the evaluation of P4P, generates an estimate of average causal pathways. The assumption is that all facilities experience the same pathway to impact, however, it is of course possible that facilities introduce different strategies to achieve outcomes and that there is some variation in pathways across facilities.

The assumption that interventions affect mediators, which in turn affect outcomes, presupposes a temporal ordering, of the change in mediators preceding that of outcomes. In our study we only had outcomes measured at two points in time: at baseline and endline. Hence, changes in mediators were measured at the same time as changes in outcomes. In the case of mediators measured at the individual level, this was problematic, as we would not expect a woman's report of kindness during her delivery to affect her delivery choice, rather we would expect her choice to be based on perceptions of kindness from the experience of other women. For this reason, we estimated the mediator excluding the woman herself. Further studies should seek to obtain measures of the mediator prior to that of outcomes, either through midline surveys, or by framing questions appropriately (for example, did you perceive that kindness during delivery had improved at your nearby facility prior to your birth?).

While we were able to identify significant mediators and explain how much of the overall effect of P4P each could explain, we were unable to determine the order of the causal chain.

Some mediators may cause other mediators, hence there is likely to be a hierarchy of outcomes (for example, increased availability of Oxytocin may affect health worker kindness, as increased drug availability improves their ability to do their job, which in turn affects service uptake). Epidemiology offers methods for quantifying the effects of multiple mediators, and their interactions, and decomposing them, but these methods are still very recent and with limited application [10, 11, 13, 42]. Most importantly, they rely on identifying assumptions, which are often unlikely to be satisfied or hard to prove within policy experiments. Further analysis should explore ways to examine more complex causal pathways, for example interactions between financing and human resources or governance factors, and to assess total mediated effect.

**Conclusions**

In this study we found that the effect of P4P on institutional deliveries was mediated by a reduction in the probability of women paying for delivery care and an increase in provider kindness during deliveries, and greater availability of drugs. The increase in coverage of IPT during antenatal care was mediated by more frequent supervision visits.

This study illustrates that there is great potential to apply the method of causal mediation analysis to help unpack the causal mechanisms of complex health systems interventions such as P4P, shedding light on how they impact the health system to achieve population health goals. We encourage further research of this kind to strength the evidence base about how health system intervention work and extend available methods.

**List of abbreviations**

DiD     Differerence in difference

LMICs   Low and Middle Income Countries

LSEM    Linear Structural Equations Model

P4P     Payment for Performance

**Authors' contributions:** JB conceptualised the study and conceptual framework and contributed to the first draft of the paper.  LA conceptualised and carried out the analysis and contributed to the first draft of the paper.  PB critically revised the paper. All authors edited the manuscript.

# References

1.    Ludwig, J., J.R. Kling, and S. Mullainathan, *Mechanism Experiments and Policy Evaluations.* Journal of Economic Perspectives, 2011. **25**(3): p. 17-38.

2.    Craig, P., et al., *Developing and evaluating complex interventions: New guidance*, 2008, Medical Research Council.

3.    Bonell, C., et al., *Realist randomised controlled trials: a new approach to evaluating complex public health interventions.* Social Science and Medicine, 2012. **75**(12): p. 2299-2306.

4.    Oakley, A., et al., *Process evaluation in randomised controlled trials of complex interventions.* British Medical Journal, 2006. **18**(332(7538)): p. 413-6.

5.    Moore, G.F., et al., *Process evaluation of complex interventions: Medical Research Council guidance.* British Medical Journal, 2015. **350**(h1258 ).

6.    MacKinnon, D.P., A.J. Fairchild, and M.S. Fritz, *Mediation Analysis.* Annual Review of Psychology, 2007. **58**(593): p. 59.

7.    Grant, A., et al., *Process evaluations for cluster-randomised trials of complex interventions: a proposed framework for design and reporting.* Trials, 2013. **12**(14): p. 15.

8.    Imai, K., et al., *Unpacking the Black Box of Causality: Learning about Causal Mechanisms from Experimental and Observational Studies.* American Political Science Review, 2011. **105**(4).

9.    Gunzler, D., et al., *Introduction to mediation analysis with structural equation modeling.* Shanghai Archives of Psychiatry, 2013. **25**(6): p. 390-4.

10.   Heckman, J., R. Pinto, and P.A. Savelyev, *Understanding the mechanisms through which an influential early childhood program boosted adult outcomes.* American Economic Review, 2013. **103**: p. 2052-2086.

11.   Heckman, J. and R. Pinto, *Econometric mediation analyses: Identifying the sources of treatment effects from experimentally estimated production technologies with unmeasured and mismeasured inputs.* Econometric Reviews, 2015. **34**(1-2): p. 6-31.

12.   Conti, G., J. Heckman, and R. Pinto, *The Effects of Two Influential Early Childhood Interventions on Health and Healthy Behaviors*, in *IZA Discussion Paper* IZA, Editor 2015: Bonn, Germany.

13.   DeStavola, B.L., et al., *Mediation analysis with intermediate confounding: structural equation modeling viewed through the causal inference lens.* American Journal of Epidemiology, 2015. **181**(1): p. 64-80.

14.   Abramsky, T., et al., *Ecological pathways to prevention: How does the SASA! community mobilisation model work to prevent physical intimate partner violence against women?* BMC Public Health, 2016. **16**(1).

15.   Ngo, D., S. Tisamarie, and S. Bauhoff, *Health system changes under pay-for-performance: the effects of Rwanda's national programme on facility inputs.* Health Policy and Planning, 2016. **first published online July 19, 2016 doi:10.1093/heapol/czw091**.

16.   Scott, A., et al., *The effect of financial incentives on the quality of health care provided by primary care physicians.* Cochrane Database Systematic Reviews, 2011. **9**.

17.   Witter, S., et al., *Paying for performance to improve the delivery of health interventions in low- and middle-income countries* Cochrane Database of Systematic Reviews 2012. **Feb 15**(2:CD007899).

18.   Ogundeji, Y., J. Bland, and T. Sheldon, *The effectiveness of payment for performance in health care: A meta-analysis and exploration of variation in outcomes.* Health Policy, 2016. **S0168-8510**(16): p. 30219-6.

19.   Hasnain, Z., N. Manning, and J.H. Pierskalla, *Performance-Related Pay in the Public Sector: A Review of Theory and Evidence.* World Bank Policy Research Working Paper, 2012. **6043**.

20. Basinga, P., et al., *Effect on maternal and child health services in Rwanda of payment to primary health-care providers for performance: an impact evaluation.* The Lancet, 2011. **377**(9775): p. 1421-8.

21. Binyaruka, P., et al., *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**(8).

22. Bonfrer, I., E. Van de Poel, and E. Van Doorslaer, *The effects of performance incentives on the utilization and quality of maternal and child care in Burundi.* Social Science & Medicine, 2014. **123**: p. 96-104.

23. Peabody, J., et al., *Financial incentives and measurement improved physicians' quality of care in the Philippines.* Health Affairs (Millwood), 2011. **30**(4): p. 773-81.

24. Yip, W., et al., *Capitation combined with pay-for-performance improves antibiotic prescribing practices in rural China.* Health Affairs (Millwood), 2014. **33**(3): p. 502-10.

25. Van de Poel, E., et al., *Impact of performance-based financing in a low-resource setting: a decade of experience in Cambodia Health Economics.* Health Economics, 2015. **online first**.

26. Witter, S., et al., *Performance-based financing as a health system reform: mapping the key dimensions for monitoring and evaluation.* BMC Health Service Research, 2013. **13**: p. 367.

27. Meessen, P., et al., *Reviewing institutions of rural health centres: the Performance Initiative in Butare, Rwanda.* Tropical Medicine & International Health, 2006. **11**(8): p. 1303-17.

28. Bertone, M.P. and B. Meessen, *Studying the link between institutions and health system performance: a framework and an illustration with the analysis of two performance-based financing schemes in Burundi.* Health Policy and Planning, 2013. **28**(8): p. 847-57.

29. Ssengooba, F., B. McPake, and N. Palmer, *Why performance-based contracting failed in Uganda - an "open-box" evaluation of a complex health system intervention.* Social Science and Medicine, 2012. **75**(2): p. 377-83.

30. Bonfrer, I., et al., *Introduction of performance-based financing in Burundi was associated with improvements in care and quality.* Health Affairs, 2014. **33**(12): p. 2179-2187.

31. Borghi, J., et al., *Protocol for the evaluation of a pay for performance programme in Pwani region in Tanzania: A controlled before and after study.* Implementation Science 2013, 8:80, 2013. **8**(80).

32. Falisse, J.-B., et al., *Community participation and voice mechanisms under performance-based financing schemes in Burundi.* 2012. **17**(5): p. 674-682.

33. Mayumana, I., et al., *Effects of Payment for Performance on accountability mechanisms: Evidence from the Pwani pilot, Tanzania* Social Science and medicine, 2016. **Forthcoming**.

34. Waddimba, A.C., et al., *Provider attitudes associated with adherence to evidence-based clinical guidelines in a managed care setting.* Medical Care Research Review, 2010. **67**(1): p. 93-116.

35. Gertler, P. and C. Vermeersch, *Using Performance Incentives to Improve Medical Care Productivity and Health Outcomes.* Working paper, 2013.

36. Huillery, E. and J. Seban, *Performance-based financing for health: experimental evidence from the Democratic Republic of Congo.* mimeo, available at: http://www.bsg.ox.ac.uk/sites/blavatnik/files/documents/paper_DRC_April2014.pdf, 2014.

37. Wang, H., et al., *An experiment in payment reform for doctors in rural China reduced some unnecessary care but did not lower total costs.* Health Affairs (Millwood). , 2011. **30**(12): p. 2427-36.

38. Binyaruka, P., M. Mamdani, and J. Borghi, *Improving quality of care through payment for performance: examining effects on the availability and stock out of essential commodities in Tanzania.* Tropical Medicine and International Health, 2016. **Forthcominng**.

39. Baron, R.M. and D.A. Kenny, *The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations.* Journal of Personality and Social Psychology 1986. **51**: p. 1173–82.

40. Manski, C.F., *Identification for Prediction and Decision*, ed. H.U. Press. 2009, Boston, M.A.

41. Imai, K., L. Keele, and T. Yamamoto, *Identication, inference, and sensitivity analysis for causal mediation effects.* Statistical Science, 2010. **25**: p. 51-71.

42. Keele, L., D. Tingley, and T. Yamamoto, *Identifying Mechanisms behind Policy Interventions via Causal Mediation Analysis.* Journal of Policy Analysis and Management, 2015. **34**(4): p. 937-963.

43. Imai, K., L. Keele, and D. Tingley, *A General Approach to Causal Mediation Analysis.* Psychological Methods, 2010. **15**: p. 309-334.

44. Hicks, R. and D. Tingley, *Causal Mediation Analysis.* The Stata Journal, 2011. **11**(4): p. 1-15.

45. Cameron, A., J. Gelbach, and D. Miller, *Bootstrap-Based Improvements for Inference with Clustered Errors.* The Review of Economics and Statistics, 2008. **90**(3): p. 414-427.

46. Cameron, A.C. and D.L. Miller, *A Practitioner's Guide to Cluster-Robust Inference.* Journal of Human Resources, 2015. **50**(2): p. 317-372.

47. Falisse, J.B., et al., *Performance-based financing in the context of selective free health-care: an evaluation of its effects on the use of primary health-care services in Burundi using routine data.* Health Policy and Planning, 2014. **30**(10): p. 1251-60.

48. Westhoff, G., A.M. Cotter, and J.E. Tolosa, *Prophylactic oxytocin for the third stage of labour to prevent postpartum haemorrhage.* Cochrane Database System Review, 2013. **Oct 30**(10:CD001808).

49. Wilunda, C., et al., *A qualitative study on barriers to utilisation of institutional delivery services in Moroto and Napak districts, Uganda: implications for programming.* BMC Pregnancy Childbirth, 2014. **14**(259).

50. Moore, J., et al., *Moving toward patient-centered care: Women's decisions, perceptions, and experiences of the induction of laborprocess.* Birth, 2014. **41**(2): p. 138-46.

51. Kruk, M.E., et al., *User fee exemptions are not enough: out-of-pocket payments for 'free' delivery services in rural Tanzania.* Tropical Medicine and International Health 2008. **13**(12): p. 1442-1451.

52. Kruk, M., et al., *Women's preferences for obstetric care in rural Ethiopia: a population-based discrete choice experiment in a region with low rates of facility delivery.* Journal of Epidemiology and Community Health, 2010. **64**(11): p. 984-8.

53. Larson, E., et al., *Moving Toward Patient-Centered Care in Africa: A Discrete Choice Experiment of Preferences for Delivery Care among 3,003 Tanzanian Women.* PLoS One, 2015. **10(8):e0135621**.

**Table 1. Health Financing, governance and human resources indicators tested as potential mediators linked to Theory of Change**

| Steps on the causal pathway /mediators | Indicators for measurement (data source) |
|---|---|
| Strengthened supervision | - Health workers received a supervision visit in the last 90 days (Health worker survey) |
| Active governing committees | - Health facility with a governing committee (Health facility survey) <br> - Health facility governing committees meetings held in last 90 days (Health facility survey and health worker survey) <br> - Minutes of health facility government committee meetings available (Health facility survey and health worker survey) <br> - Health workers satisfied with relationships between health facility and local leaders (Health worker survey) |
| Availability of drugs, medical supplies and equipment | - Index of stock out of 24 essential drugs in the past 90 days (Health facility survey) <br> - Index of stock out of 5 drugs used during delivery in the past 90 days (Health facility survey) <br> - Index of stock out of medical supplies in the past 90 days (Health facility survey) <br> - Index of stock out of SP [IPTp] (Health facility survey) <br> - Index of medical equipment functionality (Health facility survey) |
| Health worker more motivated and knowledgeable | - Intrinsic and extrinsic motivation score (Health worker survey) |
| Improved patient provider interactions | - Kindness during delivery ranked by women on a scale of 1-10 where 1 is minimum and 10 is maximum (Women survey) |
| Adherence to clinical protocol | - Uptake of IPT 2 during ANC (Women survey) |
| Increased outreach activities | - Frequency of outreach visits performed by health facility staff in past 90 days (Health facility survey and health worker survey) <br> - Outreach visits conducted by health facility staff in the last 90 days (Health facility survey and health worker survey) |
| Increase insurance enrolment | - Health insurance scheme available in the community (Health facility survey) <br> - Number of community health insurance scheme members per health facility (Health facility survey) |
| Reduction in user charges | - Probability of payment at delivery at the health facility (or public health facility) (Women survey) |

Note: Health workers kindness and knowledge have been re-ranked on a scale from 1 to 100 for ease of interpretation of the coefficients.

**Table 2. Summary statistics of maternal care outcomes and potential mediators at the baseline and endline and by intervention and comparison group**

| | Baseline | | | | | | | Endline | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Intervention | | Control | | Total | | Difference intervention control | Intervention | | Control | | Total | | Difference intervention control |
| | N | Mean (SE) | N | Mean (SE) | N | Mean (SE) | T-Test (P-value) | N | Mean (SE) | N | Mean (SE) | N | Mean (SE) | T-Test (P-Value) |
| **Outcome** | | | | | | | | | | | | | | |
| Facility based delivery | 1,389 | 84.7 (36.0) | 1,485 | 86.8 (33.9) | 2,874 | 85.8 (34.9) | -2.1 (0.113) | 1,449 | 89.2 (31.0) | 1,462 | 83.1 (37.5) | 2,911 | 86.2 (34.5) | 6.1 (0.000) |
| Public facility delivery | 1,389 | 76.7 (42.3) | 1,485 | 77.8 (41.6) | 2,874 | 77.3 (41.9) | -1.0 (0.510) | 1,449 | 81.6 (38.8) | 1,462 | 77.1 (42.0) | 2,911 | 79.3 (40.5) | 4.5 (0.003) |
| Taken at least 2 doses of malaria drug during pregnancy | 1,038 | 49.5 (50.0) | 1,212 | 56.7 (49.6) | 2,250 | 53.4 (49.9) | -7.2 (0.001) | 1,279 | 72.9 (44.4) | 1,252 | 69.2 (46.2) | 2,531 | 71.1 (45.3) | 3.8 (0.036) |
| **Potential mediators** | | | | | | | | | | | | | | |
| Health workers did not receive supervision visit in the last 90 days (%) | 1,315 | 11.3 (31.6) | 1,489 | 9.2 (28.9) | 2,804 | 10.2 (30.2) | 2.1 (0.073) | 1,165 | 1.5 (12.3) | 1,172 | 15.3 (36.0) | 2,337 | 8.4 (27.8) | -13.7 (0.000) |
| Health facility with a governing committee (%) | 1,202 | 74.3 (23.0) | 1,384 | 70.5 (45.6) | 2,586 | 72.2 (44.8) | 3.8 (0.029) | 1,119 | 80.6 (39.5) | 1,442 | 74.7 (43.5) | 2,891 | 77.7 (41.7) | 5.9 (0.000) |
| Governing committee met in the past 90 days (%) | 1,345 | 94.4 (23.0) | 1,444 | 93.0 (25.5) | 2,789 | 93.7 (24.3) | 1.4 (0.124) | 1,395 | 63.4 (48.2) | 1,442 | 44.5 (49.7) | 2,837 | 53.8 (49.9) | 19.0 (0.000) |
| Minutes of health facilities governing committee meetings available (%) | 1,249 | 92.7 (26.0) | 1,300 | 80.5 (39.7) | 2,549 | 86.5 (34.2) | 12.3 (0.000) | 1,193 | 90.1 (29.9) | 1,077 | 92.9 (25.6) | 2,270 | 91.5 (28.0) | -2.8 (0.016) |
| Health worker satisfied with relationship HF and local leaders (%) | 1,374 | 63.2 (48.3) | 1,489 | 58.2 (49.3) | 2,863 | 60.6 (48.9) | 4.9 (0.000) | 1,430 | 63.1 (48.3) | 1,462 | 57.1 (49.5) | 2,892 | 60.1 (49.0) | 6.0 (0.001) |

| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Drug stock-out index-general (0-1 index) (%) | 1,323 | 54.6 (23.3) | 1,468 | 45.7 (27.8) | 2,791 | 49.9 (26.2) | 8.9 (0.000) | 1,449 | 37.6 (25.3) | 1,462 | 45.7 (29.7) | 2,911 | 41.7 (27.8) | -8.0 (0.000) |
| Drugs at delivery -stock-out index (0-1 index) (%) | 1,323 | 54.3 (31.1) | 1,448 | 38.6 (34.4) | 2,771 | 46.1 (33.8) | 15.7 (0.000) | 1,449 | 36.7 (32.6) | 1,462 | 48.8 (37.9) | 2,911 | 42.8 (35.9) | -12.1 (0.000) |
| Oxytocin injection stock-out last 90days (%) | 1,302 | 42.0 (49.4) | 1,448 | 17.3 (37.9) | 2,750 | 29.0 (45.4) | 24.7 (0.000) | 1,449 | 22.3 (41.6) | 1,462 | 33.7 (47.3) | 2,911 | 28.0 (44.9) | -11.4 (0.000) |
| Ergometrin injection stock-out last 90days (%) | 1,286 | 65.4 (47.6) | 1,408 | 51.3 (50.0) | 2,694 | 58.1 (49.4) | 14 (0.000) | 1,449 | 42.9 (49.5) | 1,462 | 57.1 (49.5) | 2,911 | 50.1 (50.0) | -14.2 (0.000) |
| Misoprostol tablet stock-out last 90days (%) | 1,302 | 58.8 (49.2) | 1,428 | 49.1 (50.0) | 2,730 | 53.7 (49.9) | 9.7 (0.000) | 1,449 | 44.8 (49.7) | 1,462 | 55.7 (49.7) | 2,911 | 50.3 (50.0) | -10.9 (0.000) |
| Medical supplies stock-out index (0-1 index) (%) | 1,180 | 39.3 (25.3) | 1,227 | 25.6 (23.2) | 2,407 | 32.3 (25.2) | 13.8 (0.000) | 1,449 | 20.0 (19.1) | 1,462 | 21.3 (18.4) | 2,911 | 20.6 (18.8) | -1.3 (0.060) |
| Index of stock out of IPTp last 90 days (%) | 1,302 | 27.3 (44.6) | 1,486 | 19.5 (39.5) | 2,770 | 23.1 (42.1) | 7.9 (0.000) | 1,449 | 6.8 (25.1) | 1,462 | 10.9 (31.2) | 2,911 | 8.9 (28.4) | -4.2 (0.000) |
| Service delivery disrupted due to broken equipment last 90days (%) | 1,323 | 24.8 (43.2) | 1,468 | 12.5 (33.0) | 2,791 | 18.3 (38.7) | 12.3 (0.000) | 1,392 | 5.2 (22.2) | 1,462 | 6.4 (24.4) | 2,854 | 5.8 (23.3) | -1.2 (0.174) |
| Intrinsic motivation score | 1,374 | -7.7 (95.7) | 1,489 | 7.7 (74.1) | 2,863 | 0.3 (85.4) | -15.4 (0.000) | 1,429 | -7.2 (93.2) | 1,462 | 8.05 (73.7) | 2,891 | 0.5 (84.3) | -15.3 (0.000) |
| Extrinsic motivation score | 1,374 | -13.7 (68.2) | 1,489 | 1.1 (82.0) | 2,863 | -6.0 (76.1) | -14.8 (0.000) | 1,430 | -12.0 (68.5) | 1,462 | 0.5 (80.4) | 2,892 | -5.7 (75.0) | -12.6 (0.000) |
| Health worker knowledge (%) [a] | 1299 | 47.0 (27.4) | 1,467 | 67.7 (23.2) | 2,766 | 57.9 (27.0) | -20.0 (0.000) | 1,388 | 57.7 (9.8) | 1,347 | 59.6 (8.7) | 2,735 | 58.6 (9.3) | -19 (0.000) |
| Kindness at delivery (%) [a] | 1,389 | 74.8 (10.9) | 1,485 | 79.0 (10.1) | 2,874 | 77.0 (10.7) | -4.2 (0.000) | 1,449 | 88.9 (9.1) | 1,452 | 83.7 (11.7) | 2,911 | 83.8 (10.5) | 0.21 (0.590) |
| Outreach visits conducted by health facility staff in the last 90 days (%) | 1,358 | 60.6 (48.9) | 1,424 | 58.5 (49.3) | 2,782 | 59.5 (49.1) | 2.1 (0.258) | 1,449 | 73.5 (44.1) | 1,462 | 60.0 (49.0) | 2,911 | 66.7 (47.1) | 13.4 (0.000) |
| Frequency of outreach visits performed | 1,358 | 2.0 | 1,424 | 2.0 | 2,782 | 2.1 | -0.0 | 1,449 | 2.5 | 1,462 | 2.03 | 2,911 | 2.3 | 0.48 |

| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| by staff in last 90 days | | | | | | | | | | | | | | |
| | | (2.8) | | (2.2) | | (2.5) | (0.843) | | (3.27) | | (3.0) | | (3.1) | (0.000) |
| Health insurance scheme available in the community (%) | 1,358 | 79.8 | 1,429 | 55.9 | 2,787 | 67.5 | 24.0 | 1,449 | 84.0 | 1,462 | 62.4 | 2,911 | 73.2 | 21.5 |
| | | (40.1) | | (49.7) | | (46.8) | (0.000) | | (36.7) | | (48.4) | | (44.3) | (0.000) |
| Number of community health insurance scheme members per HF | 991 | 23.3 | 1,144 | 14.8 | 2,135 | 18.8 | 8.5 | 1,369 | 36.9 | 1,424 | 32.5 | 2,793 | 34.7 | 4.3 |
| | | (40.9) | | (31.9) | | (36.6) | (0.000) | | (52.9) | | (56.2) | | (54.6) | (0.038) |
| Percentage of women who paid for delivery in a HF (%) | 1,389 | 19.9 | 1,485 | 14.3 | 2,874 | 17.0 | 5.5 | 1,449 | 0.133 | 1.462 | 0.119 | 2,911 | 0.126 | 0.014 |
| | | (20.0) | | (23.8) | | (22.2) | (0.000) | | (0.205) | | (0.222) | | (0.214) | (0.082) |

(a) Kindness and knowledge originally ranked in from 1 to 10, but rescaled here in 1 to 100 for ease of interpretation of the coefficient. (b) Calculated for each woman as average of the responses of all women in the same health facility catchment area, excluding the woman herself.

## Table 3.  Effect of P4P on institutional delivery and on potential mediators

| | Effect of P4P (% change) | Observations |
|---|---|---|
| **Maternal care outcomes** | | |
| Facility based delivery | 8.2*** (3.6; 12.8) | 5,747 |
| Public facility delivery | 6.5** (1.3; 11.7) | 5,747 |
| Taken at least 2 doses of malaria drug | 10.3*** (4.4; 16.1) | 4,759 |
| **Potential mediators** | | |
| Health workers did not receive supervision visit in the last 90 days | -18.0** (-32.0; -4.0) | 5,100 |
| Drug stock-out index-general (0-1 index) | -17.2***BS (-28.6; -5.8) | 5,659 |
| Drugs at delivery stock-out index (0-1 index) | -27.0***BS (-43.4; -10.6) | 5,639 |
| Oxytocin injection stock-out last 90days | -36.2***BS (-55.9; -16.4) | 5,618 |
| Ergomentrin injection stock-out last 90days | -26.1** (-48.2; -4.0) | 5,562 |
| Medical supplies stock-out index (0-1 index) | -14.8***BS (-24.8; -4.9) | 5,278 |
| Service delivery disrupted due to broken equipment last 90days | -14.9** (-29.3; -0.4) | 5,602 |
| Health worker knowledge | 18.8*** BS (10.4; 27.2) | 5,461 |
| Kindness ranks for HW at delivery | 4.3* (-0.4; 9.0) | 5,747 |
| Percentage of women who paid for delivery in a HF | -4.5* (-9.5; 0.6) | 5,750 |

95% Confidence intervals in parentheses; * $p<0.10$, ** $p<0.05$, *** $p<0.01$; % indicates percentage points change. BS: Significant at 5% level with Bonferroni adjusted p-value for multiple outcomes: Bonferroni adjusted p-value for potential mediators within three major groups: Financing 0.0047, Governance 0.0017, Human resources 0.0414. Indicators within each category in Table 1 have been grouped into: Governance (Strengthened supervision, Active governing committees, Increased outreach activities), Financing (Availability of drugs, medical supplies and equipment, Increase insurance enrolment, Reduction in user charges) and Human resources (Health worker more motivated and knowledgeable, Improved patient provider interactions, Adherence to clinical protocol).

## Table 4. Effect of P4P and potential mediators on maternal care outcomes (results from equation 3)

| | Facility based delivery (% change) | Public facility delivery (% change) | Taken at least 2 doses of malaria drug (% change) |
|---|---|---|---|
| **Effect of P4P without mediators** | 8.2*** | 6.5** | 10.3*** |
| | (3.6; 12.8) | (1.3; 11.7) | (4.4; 16.1) |
| **Effect of P4P mediated by** | | | |
| **Health worker did not receive supervision in the last 90 days** | | | |
| - **Coefficient on P4P (B$_1$)** | 5.6** | 4.3 | 10.7*** |
| | (4.0; 10.7) | (-1.5; 10.2) | (4.1; 17.3) |
| - **Coefficient on mediator (B$_4$)** | -3.1 | -0.8 | -8.5*** |
| | (-8.5; 2.3) | (-8.2; 6.6) | (-14.6; -2.4) |
| Drug stock-out index-general (0-1 index) | | | |
| - Coefficient on P4P (B$_1$) | 7.8*** | 5.5** | 11.5*** |
| | (3.0; 12.6) | (0.2; 10.9 | (5.5; 17.4) |
| - Coefficient on mediator (B$_4$) | -3.6 | -2.9 | 7.0 |
| | (-9.2; 1.9) | (-9.0; 3.2) | (-1.8; 15.8) |
| Drugs at delivery -stock-out index (0-1 index) | | | |
| - Coefficient on P4P (B$_1$) | 7.4*** | 4.6* | 10.6*** |
| | (2.5; 12.3) | (-0.6; 9.9) | (4.6; 16.5) |
| - Coefficient on mediator (B$_4$) | -3.2 | -4.3* | 0.8 |
| | (-7.6; 1.2) | (-9.0; 0.3) | (-5.6; 7.2) |
| **Oxytocin injection stock-out last 90days** | | | |
| - **Coefficient on P4P (B$_1$)** | 6.3*** | 3.8 | 10.7*** |
| | (1.5; 11.1) | (-1.6; 9.1) | (4.6; 16.9) |
| - **Coefficient on mediator (B$_4$)** | -4.9*** | -5.3*** | 0.9 |
| | (-8.4; -1.4) | (-8.9; -1.7) | (-4.2; 6.1) |
| Ergometrin injection stock-out last 90days | | | |
| - Coefficient on P4P (B$_1$) | 8.1*** | 5.7** | 10.8*** |
| | (2.8; 13.4) | (0.1; 11.4) | (4.7; 16.8) |
| - Coefficient on mediator (B$_4$) | 1.0 | 0.5 | 0.7 |
| | (-3.7; 5.6) | (-4.4; 5.5) | (-4.0; 5.4) |
| Medical supplies stock-out index (0-1 index) | | | |
| - Coefficient on P4P (B$_1$) | 8.0*** | 4.9* | 10.9*** |
| | (3.1; 12.9) | (-0.8; 10.6) | (4.3; 17.5) |
| - Coefficient on mediator (B$_4$) | -6.8 | -8.0 | 1.3 |
| | (-16.5; 2.8) | (-18.2; 2.2) | (-9.8; 12.4) |
| Service delivery disrupted due to broken equipment last 90days | | | |
| - Coefficient on P4P (B$_1$) | 7.7*** | 5.4** | 9.9*** |
| | (3.0; 12.4) | (0.1; 10.6) | (3.8; 16.0) |
| - Coefficient on mediator (B$_4$) | -3.7 | -3.7 | -0.5 |
| | (-8.8; 1.5) | (-10.3; 2.9) | (-8.8; 7.7) |
| Health worker knowledge | | | |
| - Coefficient on P4P (B$_1$) | 8.6*** | 6.6** | 11.2*** |
| | (4.0; 13.2) | (1.5; 11.7; ) | (4.7; 17.6) |
| - Coefficient on mediator (B$_4$) | -5.6 | -1.7 | -3.4 |
| | (-14.5; 3.2) | (-12.0; 8.7) | (-17.1; 10.3) |
| **Health worker kindness at delivery** | | | |
| - **Coefficient on P4P (B$_1$)** | 4.3 | 3.3 | 10.0*** |
| | (-2.2; 10.8) | (-3.2; 9.9) | (4.0; 15.9) |
| - **Coefficient on mediator (B$_4$)** | 9.1*** | 8.0*** | 0.3 |
| | (6.4; 11.8) | (4.7; 10.3) | (-1.9; 26) |
| **Percentage of women who paid for delivery in a HF** | | | |
| - **Coefficient on P4P (B$_1$)** | 4.0 | 1.5 | 9.4*** |
| | (-2.5; 10.5) | (-5.0; 8.0) | (3.5; 15.3) |
| | -93.5*** | -112.9*** | -17.9* |
| - **Coefficient on mediator (B$_4$)** | (-122.7; -64.4) | (-141.0; -84.8) | (-35.8; 0.1) |

Taken at least 2 doses of malaria drug during pregnancy

| | | |
|---|---|---|
| - Coefficient on P4P ($B_1$) | 9.3*** | 8.2*** |
| | (2.5) | (2.9) |
| - Coefficient on mediator ($B_4$) | 1.4 | 2.1* |
| | (1.2) | (1.2) |

95% Confidence intervals in parentheses; * p<0.10, ** p<0.05, *** p<0.01; % indicates percentage points change; Bold font indicates identified mediators.

## Table 5. Indirect effect of potential mediators on maternal care outcomes

| | Facility based delivery (% change) | | Public facility delivery (% change) | | Taken at least 2 doses of malaria drug (% change) | |
|---|---|---|---|---|---|---|
| **P4P** | | | | | | |
| Total effect | 8.2 | *** | 6.5 | *** | 10.3 | *** |
| | (3.6; 12.8) | | (1.3; 11.7) | | (4.4; 16.1) | |
| Standard Error | | | | | | |
| | | | | | | |
| **Indirect effect of P4P on outcome through mediators** | | | | | | |
| | | | | | | |
| **Health worker did not receive supervision last 90 days** | | | | | | |
| Indirect effect (coefficient interaction) | | | | | 1.5 | *** |
| Confidence Interval | | | | | (0.1, 3.0) | |
| Percentage of total effect explained | | | | | 15 | |
| Correlation at which ACME=0 | | | | | -0.0343 | |
| | | | | | | |
| **Oxytocin injection stock-out** | | | | | | |
| Indirect effect (coefficient interaction) | 1.8 | *** | 1.9 | *** | | |
| | (0.7, | | | | | |
| Confidence Interval | 2.9) | | (0.8, 3.1) | | | |
| Percentage of total effect explained | 22 | | 30 | | | |
| Correlation at which ACME=0 | -0.0439 | | -0.0437 | | | |
| | | | | | | |
| **Average kindness  (excluding woman herself)** | | | | | | |
| **Indirect effect (coefficient interaction)** | 3.9 | *** | 3.2 | *** | | |
| **Confidence Interval** | (3.0, 4.7) | | (2.4, 3.9) | | | |
| **Percentage of total effect explained** | 48 | | 49 | | | |
| **Correlation at which ACME=0** | 0.2085 | | -0.0202 | | | |
| | | | | | | |
| **Mean of women who paid for delivery (excluding woman herself)** | | | | | | |
| Indirect effect (coefficient interaction) | 3.9 | *** | 5.0 | *** | | |
| Confidence Interval | (3.0, 4.8) | | (3.9, 6.3) | | | |
| Percentage of total effect explained | 48 | | 78 | | | |
| Correlation at which ACME=0 | -0.2326 | | -0.2590 | | | |

Note: Results obtained from LSEM model. Correlation at which ACME=0 is derived using Imai et al sensitivity analysis.

**APPENDIX: Sensitivity analysis with data averaged at the health facility level**

**Table 1A. Effect of P4P on institutional delivery and on potential mediators**

|  | Effect of P4P (% change) | Observations |
|---|---|---|
| **Maternal care outcomes** |  |  |
| Facility based delivery | 7.87*** | 300 |
|  | (2.4, 13.3) |  |
| Public facility delivery | 5.1 | 300 |
|  | (-1.0, 11.2) |  |
| Taken at least 2 doses of malaria drug | 11.2*** | 300 |
|  | (3.9, 18.5) |  |
| **Potential mediators** |  |  |
| Health worker did not receive supervision in last 90 days | -19.2** | 266 |
|  | (-38.1; -0.2) |  |
| Governing committee met in the past 90 days | 30.1*** | 291 |
|  | (8.1; 52.1) |  |
| Drugs at delivery -stock-out index (0-1 index) | -31.3*** | 294 |
|  | (-51.8; -10.8) |  |
| Oxytocin injection stock-out last 90days | -38.2*** | 293 |
|  | (-62.8; -13.5) |  |
| Ergometrine injection stock-out last 90days | -34.9** | 290 |
|  | (-62.7; -7.0) |  |
| Antimalarial drugs stock-out last 90days | -22.8** | 294 |
|  | (-42.6; -3.0) |  |
| Service delivery disrupted due to broken equipment in last 90days | -18.6** | 292 |
|  | (-36.5; -0.7) |  |
| Health worker knowledge | 18.9*** | 300 |
|  | (8.4; 29.4) |  |
| Health worker kindness at delivery | 3.27 | 300 |
|  | (-2.42; 9.0) |  |
| Percentage of women who paid for delivery in a HF | -7.6 | 300 |
|  | (3.0) |  |

95% Confidence intervals in parentheses; * p<0.10, ** p<0.05, *** p<0.01; % indicates percentage points change.

**Table 2A. Effect of P4P and potential mediators on maternal care outcomes (results from equation 3)**

| | Facility based delivery (% change) | Public facility delivery - Public Health facilities only (% change) | Taken at least 2 doses of malaria drug (% change) |
|---|---|---|---|
| **Effect of P4P without mediators** | 7.9*** | 7.4** | 11.2*** |
| | (2.8) | (3.3) | (3.7) |
| **Effect of P4P mediated by** | | | |
| Health worker did not receive supervision in last 90 days | | | |
| - Coefficient on P4P ($B_1$) | 3.7 | 0.3 | 12.8*** |
| | (-3.1; 10.4) | (-7.4; 8.1) | (4.5; 21.2) |
| - Coefficient on mediator ($B_4$) | -0.3 | 2.7 | -8.7* |
| | (-7.6; 6.9) | (-5.6; 11.0) | (-17.6; 0.2) |
| Governing committee met in the past 90 days | | | |
| - Coefficient on P4P ($B_1$) | 8.0*** | 5.7* | 11.8*** |
| | (2.2; 13.8) | (-0.9; 12.3) | (3.9; 19.8) |
| - Coefficient on mediator ($B_4$) | 1.8 | -1.4 | 1.0 |
| | (-2.9; 6.5) | (-6.7; 4.0) | (-5.5; 7.4) |
| Drugs at delivery -stock-out index (0-1 index) | | | |
| - Coefficient on P4P ($B_1$) | 7.1** | 3.0 | 11.5*** |
| | (1.4; 12.9) | (-3.5; 9.5) | (3.6; 19.4) |
| - Coefficient on mediator ($B_4$) | -4.2* | -5.5* | -1.2 |
| | (-9.2; 0.7) | (-11.0; 0.1) | (-8.0; 5.5) |
| **Oxytocin injection stock-out last 90days** | | | |
| **- Coefficient on P4P ($B_1$)** | **5.8**\*\* | **1.9** | **12.3**\*\*\* |
| | (0.2; 11.5) | (-4.5; 8.3) | (4.3; 20.3) |
| **- Coefficient on mediator ($B_4$)** | **-6.3**\*\*\* | **-6.9**\*\*\* | **0.4** |
| | (-10.3; 2.3) | (-11.5; -2.3) | (-5.2; 6.1) |
| Ergometrin injection stock-out last 90days | | | |
| - Coefficient on P4P ($B_1$) | 8.3*** | 4.5 | 11.9*** |
| | (2.3; 14.3) | (-2.2; 11.2) | (3.8; 20.0) |
| - Coefficient on mediator ($B_4$) | 0.6 | 0.2 | -0.9 |
| | (-3.2; 4.5) | (-4.2; 4.5) | (-6.1; 4.3) |
| Antimalarials stock-out last 90days | | | |
| - Coefficient on P4P ($B_1$) | 7.9*** | 5.7* | 13.0*** |
| | (2.9) | (3.4) | (3.9) |
| - Coefficient on mediator ($B_4$) | -2.4 | -5.2* | 2.5 |
| | (2.6) | (3.0) | (3.5) |
| Service delivery disrupted due to broken equipment last 90days | | | |
| - Coefficient on P4P ($B_1$) | 7.7*** | 3.9 | 11.4*** |
| | (2.0; 13.3) | (-2.5; 10.3) | (3.7; 19.2) |
| - Coefficient on mediator ($B_4$) | -5.0* | -5.3 | -1.5 |
| | (-10.6; 0.7) | (-11.7; 1.1) | (-9.2; 6.3) |
| Health worker knowledge | | | |
| - Coefficient on P4P ($B_1$) | 7.6** | 4.3 | 10.1** |
| | (1.3; 13.9) | (-2.7; 11.3) | (1.7; 18.5) |
| - Coefficient on mediator ($B_4$) | -7.2 | -3.2 | 0.5 |
| | (-17.9; 3.5) | (-15.2; 8.8) | (-13.8; 14.8) |
| Health worker kindness | | | |
| - Coefficient on P4P ($B_1$) | 7.8*** | 5.2 | 11.6*** |
| | (2.2; 13.3) | (-1.1; 11.4) | (4.1; 19.0) |
| - Coefficient on mediator ($B_4$) | 0.3 | -0.3 | -0.11 |
| | (-1.4; 2.0) | (-0.23; 0.16) | (-0.34; 0.12) |
| Proportion of women who paid for delivery in a HF | | | |
| - Coefficient on P4P ($B_1$) | 8.2*** | 3.0 | 9.9** |
| | (2.5; 13.8) | (-3.2; 9.1) | (2.4; 17.5) |
| - Coefficient on mediator ($B_4$) | 3.8 | -27.7*** | -16.6 |
| | (-12.6; 20.2) | (-2.4; 1.7) | (-38.5; 5.2) |

95% Confidence intervals in parentheses; * p<0.10, ** p<0.05, *** p<0.01; % indicates percentage points change; Bold

font indicates identified mediators.

## Table 3A. Indirect effect of potential mediators on maternal care outcomes

| | Facility based delivery (% change) | | Public facility delivery (% change) | |
|---|---|---|---|---|
| **Effect of P4P without mediators** | | | | |
| Total effect | 7.9 | *** | 7.4 | ** |
| Standard Error | (2.4; 13.3) | | (0.9; 13.9) | |
| | | | | |
| **Indirect effect of P4P on outcome through mediators** | | | | |
| | | | | |
| **Oxytocin injection stock-out** | 3.9 | | 5.0 | |
| Indirect effect (coefficient interaction) | 0.024 | | 0.030 | |
| Confidence Interval | (-0.00; 0.138) | | (-0.083; 0.143) | |
| Percentage of total effect | | | | |
| | | | | |
| **Health worker kindness** | | | | |
| Indirect effect (coefficient interaction) | 0.1 | | | |
| Confidence Interval | (-0.019; 0.020) | | | |
| Percentage of total effect | | | | |
| | | | | |
| **Percentage of women who paid for delivery in a HF** | | | | |
| Indirect effect (coefficient interaction) | | | 2.3 | |
| Confidence Interval | | | (0.4; 4.9) | |
| Percentage of total effect | | | 31.1 | |

## Table 4A. P4P Average causal mediation effect (ACME) and direct effects, sensitivity analysis using Imai et al. (2010) approach

| | Mediation analysis | | | Sensitivity results | |
| --- | --- | --- | --- | --- | --- |
| | Mean | [95% Conf. | Interval] | | |
| **Mediators for delivery in health facility** | | | | | |
| **Oxytocin stock-out** | | | | | |
| ACME | 0.017 | 0.006 | 0.028 | | |
| Direct Effect | 0.059 | 0.021 | 0.095 | Correlation at which ACME=0 | -0.0439 |
| Total Effect | 0.077 | 0.041 | 0.109 | | |
| % of Tot Eff mediated | 0.226 | 0.159 | 0.424 | | |
| **Health worker kindness** | | | | | |
| ACME | 0.039 | 0.031 | 0.048 | | |
| Direct Effect | 0.042 | 0.007 | 0.075 | Correlation at which ACME=0 | 0.2085 |
| Total Effect | 0.081 | 0.045 | 0.116 | | |
| % of Tot Eff mediated | 0.482 | 0.337 | 0.864 | | |
| **Percentage of women who paid for delivery in HF** | | | | | |
| ACME | 0.041 | 0.032 | 0.051 | | |
| Direct Effect | 0.040 | 0.004 | 0.072 | Correlation at which ACME=0 | -0.2326 |
| Total Effect | 0.081 | 0.045 | 0.113 | | |
| % of Tot Eff mediated | 0.511 | 0.364 | 0.911 | | |
| **Mediators for delivery in public HF** | | | | | |
| **Oxytocin stock-out** | | | | | |
| ACME | 0.019 | 0.008 | 0.031 | | |
| Direct Effect | 0.037 | 0.000 | 0.077 | Correlation at which ACME=0 | -0.0437 |
| Total Effect | 0.057 | 0.021 | 0.096 | | |
| % of Tot Eff mediated | 0.342 | 0.202 | 0.933 | | |
| **Health worker kindness** | | | | | |
| ACME | 0.033 | 0.026 | 0.042 | | |
| Direct Effect | 0.031 | -0.008 | 0.068 | Correlation at which ACME=0 | 0.161 |
| Total Effect | 0.065 | 0.024 | 0.102 | | |
| % of Tot Eff mediated | 0.516 | 0.325 | 1.374 | | |
| **Percentage of women who paid for delivery in HF** | | | | | |
| ACME | 0.051 | 0.039 | 0.062 | | |
| Direct Effect | 0.014 | -0.025 | 0.050 | Correlation at which ACME=0 | -0.259 |
| Total Effect | 0.064 | 0.025 | 0.100 | | |
| % of Tot Eff mediated | 0.791 | 0.506 | 2.051 | | |
| **Mediators for IPT2 uptake** | | | | | |
| **Health workers did not receive supervision visit in last 90 days** | | | | | |
| ACME | 0.016 | 0.002 | 0.030 | | |
| Direct Effect | 0.102 | 0.038 | 0.161 | Correlation at which ACME=0 | -0.0343 |
| Total Effect | 0.117 | 0.056 | 0.174 | | |
| % of Tot Eff mediated | 0.133 | 0.089 | 0.280 | | |

Note: ACME: Average Causal Mediation Effect

**Figure 1: Theory of change of P4P pathways to impact via health system strengthening**



1. Health workers become more motivated as a result of incentives; knowledge levels increase through investment in training to meet targets; or through re-allocation of staff to poor performing facilities; 2. Verification activities result in strengthened supervision; 3.Strengthened supervision motivates health workers; 4. Local bank accounts and facility revenue give a role to governing committees; 5. Additional funds are used to procure drugs and supplies through the committee; 6. Increased availability of drugs and supplies increases motivation levels; 7. Increased availability of drugs and supplies reduces the user costs of care; 8. More motivated health workers improve interactions with patients; 9. Increased motivation, knowledge and availability of drugs increases adhere to clinical care guidelines; 10. To meet utilisation targets, health workers undertake more outreach activities in the community; 11. Health workers encourage enrolment in insurance to increase facility revenue; patients are more willing to join due to better perceived quality of care; 12. Insurance coverage reduces user costs of care; 13. Service utilisation increases as a result of quality improvements and increased affordability of care.