

This is a repository copy of *Paying for performance for health care in low- and middle-income countries::an economic perspective*.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/id/eprint/135417/>

Version: Published Version

Monograph:

Chalkley, Martin John orcid.org/0000-0002-1091-8259, Mirelman, Andrew orcid.org/0000-0002-7622-0937, Siciliani, Luigi orcid.org/0000-0003-1739-7289 et al. (1 more author) (2016) Paying for performance for health care in low- and middle-income countries::an economic perspective. Discussion Paper. CHE Research Paper . Centre for Health Economics, University of York , York, UK.

Reuse

Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.



**Paying for Performance for
Health Care in Low- and Middle-
Income Countries:
An Economic Perspective**

Martin Chalkley, Andrew Mirelman,
Luigi Siciliani, Marc Suhrcke

CHE Research Paper 140

Paying for performance for health care in low- and middle-income countries: an economic perspective

^a Martin Chalkley

^a Andrew Mirelman

^b Luigi Siciliani

^a Marc Suhrcke

^a Centre for Health Economics, University of York, York, UK

^b Department of Economics and Related Studies, University of York, York, UK

December 2016

Background to series

CHE Discussion Papers (DPs) began publication in 1983 as a means of making current research material more widely available to health economists and other potential users. So as to speed up the dissemination process, papers were originally published by CHE and distributed by post to a worldwide readership.

The CHE Research Paper series takes over that function and provides access to current research output via web-based publication, although hard copy will continue to be available (but subject to charge).

Acknowledgements

This report is financially supported by the International Decision Support Initiative (iDSI) - funded by the Bill & Melinda Gates Foundation, the UK Department for International Development, and the Rockefeller Foundation.

Further copies

Copies of this paper are freely available to download from the CHE website www.york.ac.uk/che/publications/. Access to downloaded material is provided on the understanding that it is intended for personal use. Copies of downloaded papers may be distributed to third-parties subject to the proviso that the CHE publication source is properly acknowledged and that such distribution is not subject to any payment.

Printed copies are available on request at a charge of £5.00 per copy. Please contact the CHE Publications Office, email che-pub@york.ac.uk, telephone 01904 321405 for further details.

Centre for Health Economics
Alcuin College
University of York
York, UK
www.york.ac.uk/che

Contents

Abstract.....	ii
1 Introduction	1
2 P4P from an economic perspective	2
2.1 Delegation and conditionality.....	2
2.2 The design of P4P schemes.....	2
2.2.1 Linear versus non-linear incentive schemes.....	2
2.2.2 Budget neutrality	4
2.2.3 Sticks or carrots?	4
2.3 Unintended consequences	5
2.3.1 Multitasking	5
2.3.2 Gaming	6
2.3.3 Selection or cherry picking.....	6
2.3.4 Equity concerns.....	6
2.3.5 Crowding out.....	7
3 P4P programme evidence in LMIC.....	8
3.1 Programme characteristics	8
3.2 Have they worked? Evidence from evaluated P4P programmes	9
4 Conclusion.....	11
5 Tables	12
6 References	18

Abstract

Pay for Performance (P4P) arrangements, which are fixtures of health systems in high-income countries (HIC), have been deployed across many low- and middle-income country (LMIC) settings as well. P4P programs in HICs have typically addressed the challenge of 'over delivery', controlling costs while maintaining adequate services and getting the best clinical practice, or quality of care. In LMICs, health systems are similarly concerned with issues of quality, but they may also grapple with problems of low demand, lack of resources and poor governance. By revisiting the overall framework for understanding P4P arrangements, their benefits and their risks in the context of healthcare delivery, this paper draws on experiences with P4P in HIC to assess how the insights from economic theory apply in practice in LMICs. Issues of programme design and unintended consequences are summarized and LMIC case examples of where these concepts apply and are missing from the evidence of P4P programs in LMIC settings are also reviewed. The evidence on P4P in LMICs is still in its infancy, both in terms of evidence of impact (especially as far as health outcomes are concerned), and in in terms of the attention to potential unintended consequences. However, it is critical to return to basic economic understanding of how the contractual arrangements and incentives of P4P inform program design and ultimately impact health outcomes and service delivery.

1 Introduction

The idea that a transfer of a payment should be linked to observing and verifying some tangible outcome delivered has an obvious intuitive appeal. The funder can influence what gets done and receives evidence to ensure that what was intended was done. In the jargon of economics, the verification arrangements of the funder to delivery organisation generate *incentives* to *perform* appropriately. For convenience we refer to these arrangements as Pay for Performance (P4P).

P4P is a ubiquitous part of the landscape of health care in high-income countries (HIC) and increasingly discussed and implemented in low- and middle-income country (LMIC) health systems (Miller and Babiarz 2014). There are potentially important insights to be gained from the economics of incentives in terms of when their use may be appropriate, what the unintended consequences might be, and how those unintended consequences might be mitigated or avoided altogether.

HICs have typically relied on P4P programmes as part of their efforts to address the challenge of ‘over delivery’, i.e. to control costs while maintaining adequate services and getting the best clinical practice, or quality of care. In LMICs, health systems are similarly concerned with issues of quality, but they may also grapple with problems of low demand, lack of resources and poor governance. Despite obvious differences in terms of focus between HIC and LMIC settings, the insights gained from experience with P4P in HICs are potentially valuable for LMIC health systems.

The purpose of this paper is to draw on the development of P4P in HICs and to assess how the insights from economic theory apply in practice in LMICs. We begin by revisiting the overall framework for understanding P4P arrangements, their benefits and their risks in the context of healthcare delivery. We next discuss the issues of programme design and unintended consequences and summarize the evidence regarding the impact of P4P programmes in LMIC settings.

2 P4P from an economic perspective

2.1 Delegation and conditionality

At a very general level there is a problem that those wanting to ensure the delivery of things have to delegate the actual delivery. In health care as in the delivery of many services, the actual delivery process needs to be undertaken by ‘experts’, ‘professionals’ or ‘healthcare workers’. In the language of economics, the delegator is termed the *principal* and the delegate is termed the *agent*. Whilst there are a diversity of principals and agents in many health care settings, for convenience in our exposition we use the terms ‘*funder*’ and ‘*delivery organisation*’, respectively. Between the funder and the delivery organisation, there is an agreement often referred to as the ‘*contract*’.

One approach to the problem of delegation is for the funder to simply articulate what it would like to be done and to pay the delivery organisation a reasonable sum to achieve that. We would call this an *unconditional payment* in the sense that the delivery organisation will be paid irrespective of any evidence of its performance. In a health care setting this corresponds to the funder simply making a payment to the delivery organisation and essentially letting it “get on with the job”. Such completely unconditional payments for health care delivery are increasingly rare.

If the funder could simply instruct the delivery organisation and verify that what it has required has actually been done, then there would not be any problem. The whole matter of incentives would be handled simply by an agreement of the form “*do what I want and I will pay for your effort in doing it*”. We could call this a *fully specified payment* -- the delivery organisation will receive a financial transfer only if they satisfy all of the precise requirements that the funder stipulates.

But in reality there are limitations to this approach. First, the funder may not be able to establish exactly what was done, or even if they can establish it, they may need to convince a separate actor that is charged with enforcing the agreement. Second, the funder may not be sure what they actually want done – i.e. it may depend on what the delivery organisation is going to learn or observe. Third, there may be costs to the funder and the delivery organisation for monitoring, documenting and validating that the services were delivered.

P4P can be viewed as specifying a *conditional* payment. The funder may want, for instance, to have a particular target population vaccinated, with the target defined in terms of the most vulnerable or those individuals who would realise the largest health benefit in a locality. A conditional payment might take some items which are simple to observe, verify characteristics of individuals and pay a price per service use for those individuals who could be verified as having received the service. A key observation is that, in LMICs, as in HICs, there is a growing reliance on these conditional payments – a movement towards more P4P in the provision of health care (Eijkenaar et al. 2013). The next section describes issues to consider when designing contracts between the funder and the delivery organisation.

2.2 The design of P4P schemes

2.2.1 Linear versus non-linear incentive schemes

P4P schemes can be designed in a number of different ways in terms of the strength and nature of the provider incentive. A common distinction is between a *linear* payment with a fixed amount paid for each additional unit of performance observed (for example, each person vaccinated) and a *non-linear* payment where payment is conditioned on thresholds and where per-unit payment can vary when volume is below, above or between different thresholds.

Linear payment systems, with a fixed per-unit price have the advantage of being simple to implement but do not perform well in the presence of large variations in the characteristics of those who have to be treated. Delivery organisations are likely to differ in aspects such as: costs, altruism,

and population served. This heterogeneity calls for the payment to be non-linear and gradually adjusted for different degrees of performance (Baron and Myerson 1982). The payment received increases with performance but not necessarily at a linear rate. **Figure 1** provides an example and compares a non-linear scheme where the unit price increases with the degree of performance with a linear one with a constant price. For example when some providers have a much lower marginal cost than others, a linear incentive scheme would either leave large profits to the efficient providers or drive the less efficient providers out of the market, which might then impact on delivery for individuals who can only be served by inefficient providers. A non-linear incentive scheme can mitigate (but not eliminate) this tension by paying higher unit prices for some levels of production, but this entails additional contractual complexity (Laffont and Martimort 2002).

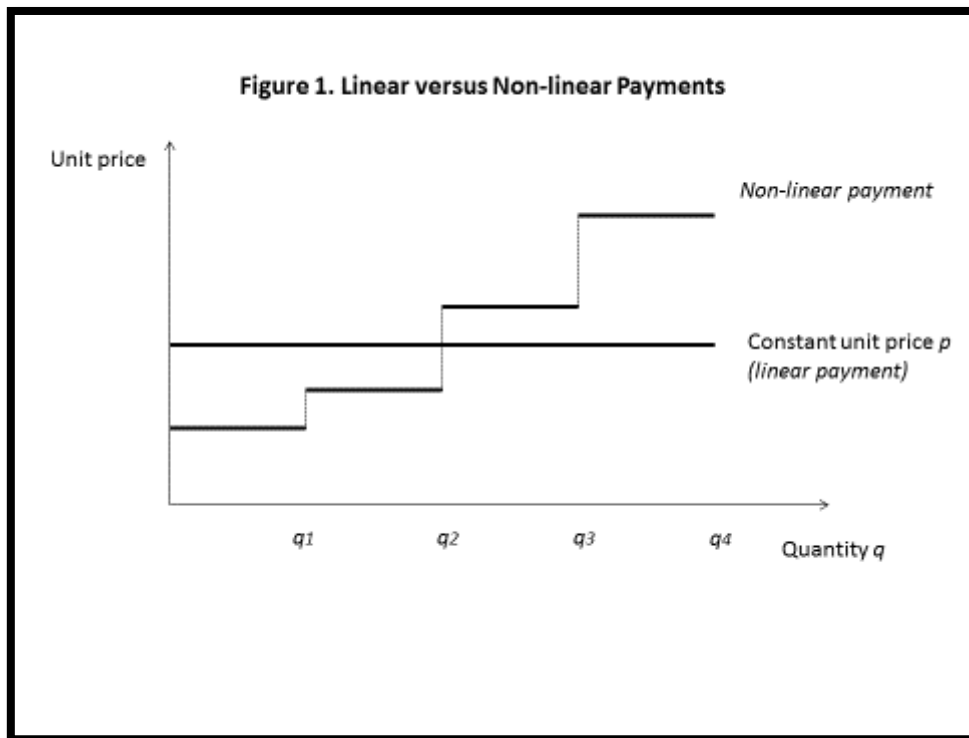


Figure 1. Linear versus non-linear payments

The central message here is that non-linear schemes can generate more sophisticated incentives but they are more difficult (or costly) to design and implement. As might be expected, emerging P4P schemes in LMICs may start with linear arrangements. For example, in the health P4P scheme set up by the Rwandan government, a linear payment was made based on a scaled score from 0 to 1, with a quality score of 1 meaning a facility received 100% of the payment. Scores of 0.5 and 0.8 would receive 50% and 80% of the payment respectively (Basinga et al. 2011). In another P4P programme in the Philippines, the Philippines Child Health Experiment, physician bonuses were paid linearly based on the number of patients seen after the physician met a quality threshold based on clinical performance vignettes (Peabody et al. 2014).

There are very few complex, non-linear contract designs in LMIC P4P schemes. However, a withholding design, where payment is not made unless a given performance threshold is met, is an example of a simplified non-linear contract. Withholding type schemes are seen in several LMIC programmes. In a performance contracting scheme set up by the Cambodian government to have NGO's provide basic care, payments were withheld to contractors if progress on a set of performance metrics is unsatisfactory (Bloom et al. 2006). In another example in Nicaragua's *Red de Protección Social (RPS)* programme where providers were paid based on reaching a specific coverage

target to specific groups of beneficiaries (0-2 year old, pregnant women, adolescents, etc.). The providers were paid on a per groups basis if they were able to provide coverage to a threshold percentage of beneficiaries (Regalia and Castro 2007).

2.2.2 Budget neutrality

A critical aspect in the design of any P4P scheme is whether it replaces a current payment scheme or if it is introduced on top of a current payment scheme. In the former case, this can be referred to as 'budget neutral'. An example of a budget-neutral scheme would be when the delivery organisation replaces a fixed budget with a linear payment (fixed price) system with no fixed budget element. An example of the second is where a top-up price is added to a fixed budget and there is an additional cost to the funder.

In theory, there is no reason why a funder cannot introduce P4P and finance it by withholding resources from other revenues of the delivery organisation. For example, a P4P scheme combined with a smaller fixed budget component can replace a system fully based on a fixed budget. In practice, this may be difficult in certain institutional contexts, and introducing a P4P scheme while also providing additional resources, at least in the first years, may help convince delivery organisations to adopt the scheme.

In practice the issue of budget neutrality is important for evaluating the impact of a scheme, as well as its cost-effectiveness. An increase in performance may come from the conditioning on performance, or it may come from the increased generosity in overall resources. Additionally, some P4P schemes may be introduced as payment in a specific project budget or as a modest percentage of the overall budget.

An example of this issue is highlighted in a recent evaluation of a P4P programme in two counties in Shandong, China to curb irrational drug use. In one treatment group, a move was made from fee for service (FFS) to capitation with 20% of the capitated budget being withheld based on performance. Since none of the health centres received the full 20% back, the authors rightly note that P4P scheme is actually providing a lower budget overall than a group that receives 100% of the capitation (Sun et al. 2016). An opposite example can be seen in a national performance-based contracting programme in Cambodia, where a bonus (or, carrot)-based incentive led to providers with performance contracts receiving more resources overall than those providers without performance contracts (Van de Poel et al. 2016).

Examples of programmes where the P4P component comprise a fraction of the overall budget come from programmes by the NGO, Cordaid, in Tanzania and Zambia. In Tanzania, the P4P programme comprised 8-10% of the total health budget for the region where it was implemented, while in the Zambia programme, the contribution was 17%. The remainder of the budgets were not performance-based and came from sources such as the ministry of health, cost recovery and insurance. In each of these P4P programmes, 50% of this portion of the overall budget was provided up front and the other 50% was paid based on performance (Canavan and Swai 2008, Vergeer and Chansa 2008).

2.2.3 Sticks or carrots?

A P4P scheme can be designed in such a way that it gives extra resources for the additional, or improved, care provided (a 'carrot' or bonus) or can give an amount of payment and then 'withhold' resources in case of under-performance (a 'stick' or punishment). Withholding may be a more natural option if the initial resources are provided to set up the services, i.e. to build capacity and cover the key costs to make the provision of services viable. In theory, the two schemes should be equivalent. In practice, they may lead to different outcomes. Delivery organisations may receive a grant to build capacity so that most of the money is provided even if the performance target is below the necessary threshold ('withhold').

In the previous section, we pointed out that two programmes, one in China and one in Cambodia, have different contract designs with respect to withholding payment or paying with bonuses. A literature review of LMIC P4P programmes also found that there is a split, and in some cases programmes can have a blend of withholding and bonuses (Grittner 2013). The lack of consensus on which design is preferred means that this needs to be carefully considered during the programme design stage with respect to the context, political acceptability and desired goals.

2.3 Unintended consequences

Some fundamental issues in conditioning payment arise when the delivery organisation knows, or is able to observe things, that are hidden from the funder. This is generically referred to as *imperfect information* in the economics literature. In this section, we discuss the underlying concepts of five important types of unintended consequences that may arise in P4P programmes. Four of these are related to the conditionality (multi-tasking, gaming, selection and equity) while the remaining one – crowding-out – is a cross-cutting theme relevant to incentivizing health care more generally. These potential unintended consequences have emerged as topics of interest due to experiences with P4P in HICs and LMICs, and each one is discussed here in more detail (Witter et al. 2012, Roland and Dudley 2015).

2.3.1 Multitasking

One of the most common potential problems with P4P is what is known as *multitasking*. Although some dimensions of performance are quantifiable and contractible, others are not. In health care it is often assumed that *quality* broadly falls into the latter category, i.e. it is non-quantifiable and non-contractible. The concern is that P4P schemes could generate improvements in the quantifiable dimensions of care at the expense of reductions elsewhere (Eggleston 2005). So by paying for *volume* we might sacrifice *quality* or *appropriateness* of treatment.

Multitasking happens when different aspects of performance are substitutes – that is, increasing one reduces the other. In such cases the P4P scheme has to be designed with care. There has been a lot of theoretical interest in how P4P schemes might be adapted to account for multitasking. Responses to multi-tasking that have been explored include introducing other regulatory or monitoring mechanisms or reducing the power (the level of reward relative to the marginal benefit) of the incentive scheme to balance the benefits from improved performance with reductions from other care (Kaarboe and Siciliani 2011). The general message from multi-tasking is that you get what you pay for, but you might not want what you get.

The evidence for multitasking in LMIC P4P schemes is very weak, in the sense that it has rarely been a focus of the research to date. A 2012 Cochrane review of P4P schemes in LMICs found that only two out of nine identified studies looked at multi-tasking and whether incentivized activities were traded off with non-incentivized ones (Witter et al. 2012). In an evaluation of the P4P programme in Tanzania run by the NGO, Cordaid, the evaluators found that the incentivized focus on curative interventions may distract from preventive services (Canavan and Swai 2008). In a separate Cordaid P4P programme in Zambia, a focus on inpatient turnover rates in health centres was thought to distract from the focus of these centres to deliver primary health care services (Vergeer and Chansa 2008). However, in both studies, no significant multi-tasking was found.

Further evidence of the distortions that a P4P programme can cause is mentioned in a qualitative review of the P4P scheme in the health sector in Rwanda where providers felt they had to focus on remunerated activities over non-remunerated ones (Kalk et al. 2010). In addition to substitution among health services, providers also felt they were neglecting health-producing activities in favour of fulfilling bureaucratic conditions for the reward, such as doing required paperwork (Kalk et al. 2010).

2.3.2 Gaming

A second concern with P4P is the potential for gaming: the data used to measure performance might be manipulated by the delivery organisation to inflate reported performance. For a P4P scheme to work as intended information on contractible dimensions of care needs to be reliable. This is not always the case however because health information systems tend to be complex, and often the information upon which payment is conditioned is self-reported by the delivery organisation. For example, for child immunization it could be difficult for the funder to verify that the immunization took place. The ideal response to gaming is for the funder to base the P4P scheme on information which is outside of the control of the delivery organisation and which is easily measurable, so that there is little scope for misclassifying patients in more remunerative groups.

If some gaming is unavoidable, the funder will either have to introduce some effective monitoring systems (for example a system of random audit) although these tend to be costly, or there may again be a case for reducing the power of the incentive scheme (Kuhn and Siciliani 2009).

An example of gaming is seen again in the qualitative evaluation of the Rwanda health sector scheme. Providers mentioned that they had retrospectively filled in reporting forms inaccurately (Kalk et al. 2010). This behaviour was justified by the providers in interviews as being due to that fact that the P4P programme was externally imposed and the performance indicators were counterproductive to producing health.

2.3.3 Selection or cherry picking

A related but distinct concern to gaming is what is known as *cherry picking*. If the cost of providing an incentivized service differs across patients, and differences in costs are not reflected in the P4P payment, the delivery organisation may have a financial incentive to select patients with low cost and avoid patients whose costs are above the tariff (Ellis 1998). The problem is exacerbated if patients with lower costs have the lowest capacity to benefit from the treatment. It may be a smaller concern if it is precisely those patients with lower costs that stand to gain most from the scheme.

One potential response for the funder is to design P4P schemes which differentiate payment according to the expected costs. But this may be problematic in at least two ways. One, the funder may have little reliable information to condition the payment on; and two, differentiating payments for different types of patients opens up the door to gaming. The delivery organisation has a financial incentive to assign patients to the most remunerative category for payment, a practice which is sometimes referred to as 'upcoding' (Dafny 2005).

There have not been any studies in LMIC P4P schemes that have found evidence of a selection effect, or cherry-picking. Examples, however, do exist from the HIC contexts. For instance, one study of a substance abuse treatment programme in the United States found that a performance-based system led to the selection of less severe cases for treatment (Shen 2003).

2.3.4 Equity concerns

Somewhat related to the above selection problem, another key concern with the introduction of P4P is that improvement in performance may come at the cost of reduced equity, understood as a widening of the gap in health or health care utilisation between lower and higher socio-economic groups. For example, if there were cherry picking or selection occurring, then it is likely that the most socio-economically advantaged patients (with better health) would benefit most from the scheme, further exacerbating the health gap between the rich and the poor.

Similarly, if multitasking is an issue and the incentivized dimensions of care benefit mostly the patients in better health, then the introduction of P4P will further increase disparities between groups based on health, which may have implications for socioeconomic disparities as well.

The funder can in some instances address this issue by identifying the groups of patients who are more vulnerable and introduce an additional incentive (e.g. a top-up payment) for those groups. Such an incentive, while possibly beneficial, is likely to be imperfect. Although the vulnerable group has on average worse health than the non-vulnerable group, there may be some patients in the non-vulnerable group with worse health than the healthiest in the vulnerable group.

An example is in relation to income. Poor individuals are likely to be in worse health and more costly to treat. A general incentive scheme is likely to benefit richer individuals (who are cheaper to treat) more than poorer individuals who are more costly to treat. An incentive scheme targeted at the poor (as opposed to the general population) may be helpful to reduce socioeconomic inequalities in health and address the highest needs but needs to account for the differences in patient cost as well as the cost of targeting.

In a re-analysis after the primary impact evaluation of the health sector P4P scheme in Rwanda, it was found that there are important inequities in the effect of the scheme (Lannes et al. 2016). Improvements in utilisation were found to be highest for more affluent groups, and in some cases, to even decrease service use for the poor. The Rwanda P4P did not include differential payment, or top-up payments to incentivize providing services to worse off groups. On the other hand, P4P schemes such as the World Bank supported performance-based financing (PBF) programmes in Burundi did provide higher capitation rates to services delivered in remote areas (Witter 2013, Bonfrer et al. 2014a).

2.3.5 Crowding out

Acknowledging all issues identified so far, some analysts have argued that even if care is perfectly conditioned on payment, P4P may have adverse consequences on the intrinsic motivation of delivery organisations. For example, Le Grand argues that the introduction of prices may turn healthcare delivery organisations from *knights* to *knaves* (Le Grand 2003). It is easily argued that health care delivery organisations are intrinsically motivated. Doctors have to endure long years of training and nursing, midwifery and community health work are vocational jobs often associated with low pay. The key concern is that although the delivery organisation may respond to the financial incentive introduced by P4P, this could be offset by a reduction of intrinsic motivation. In turn, this makes the introduction of P4P less effective. In practice, it is hard to reliably attribute low response of P4P schemes to intrinsic motivation as opposed to other contextual factors given that intrinsic motivation is difficult to define or observe.

Finding any evidence of crowding out in LMIC P4P schemes is also difficult. In health activities of the sectoral Public Sector Reform Programme in Tanzania, it was identified that if the financial incentives from the P4P scheme did not adequately account for context, there would be potential for crowding out (Leonard and Masatu 2010, Songstad et al. 2012). The effect of a P4P programme may also have the opposite effect. For example, in the health sector scheme in Rwanda, there is qualitative evidence that a P4P scheme increased motivation through providing better working conditions for providers (Kalk et al. 2010).

3 P4P programme evidence in LMIC

In this section, we summarise the characteristics and the effectiveness of several P4P schemes across LMIC settings. To do so, we conducted a purposive review of the P4P literature in relevant databases (e.g. PubMed, EconLit, and Google Scholar) and from previous systematic reviews.¹ To narrow the list of selected programmes, we only included those that had contracts between funders and health-care delivery organisations and where a quantitative evaluation was conducted. Ultimately, we selected 14 P4P programmes using a single case example per country and covering as disparate set of countries as possible, to examine in depth. We have listed each of these programmes and their characteristics in **Table A1.1-1.3**.² Of the 14 programmes, six were performance based financing (PBF) schemes, six were contracting-out models, one had characteristics of both over the scheme's life course and one final scheme was a purely public sector scheme. Geographically, the schemes took place across Latin America, Asia and Africa. Most of the schemes were implemented either in the late 1990's or early 2000's and several had pilot programmes before proceeding to scaling up. A programme in Costa Rica was introduced as early as in 1988 and the most recent year of introduction in our sample is 2007 for the programme in Zambia, though in terms of programme scale-up, Burundi's programme was most-recently scaled up from a pilot in 2011.

3.1 Programme characteristics

P4P programmes for health in LMICs may have funders that include country governments, bilateral donor agencies, multinational development banks and international not-for-profit organisations (NGOs). Delivery organisations typically include public-sector providers, private-sector providers, NGO providers or faith-based organisations. There are different models for contractual arrangements, which have been summarised in previous reviews (Eldridge and Palmer 2009). Examples of common contracting models include the 'contracting out' model and the PBF model (Liu et al. 2008, Lagarde and Palmer 2009). In each contracting model, the government may serve as either a funder or delivery organisation, and external groups such as NGOs and multinationals may serve as either service providers (in the case of contracting out) or as sources of financing (in the case of PBF), or as both.

What is conditioned on – and used for – payment in LMIC P4P programmes also varies by programme. P4P programmes are a move away from paying for inputs, or just providing resources for service delivery and reward based on outputs or outcomes. There are three primary items that may be conditioned on: health service activities, process measures of quality and health outcomes. The first two may be considered as outputs while the third is a measure of outcome. Output-based performance is typically based on service utilisation, which makes sense if it is controllable by the service provider. Frequently, these outputs can include elements of quality as well such as completing a full-course of vaccination or receiving the total number of recommended antenatal care visits.

Complex quality-based metrics have been used in several programmes. One example is the Burundi PBF's 220 item checklist to assess quality (Bonfrer et al. 2014a). Other programmes have employed sophisticated methods such as clinical performance vignettes (CPV), which was done in the Philippines' Quality Improvement Demonstration Study (QIDS), a P4P programme rolled out to 30 district hospitals in the country (Peabody et al. 2014). Other quality metrics may consist of patient-reported outcomes or even waiting times. Performance that was based on health outcomes has also been utilized in programmes – examples can be seen in the performance based contracting

¹ A list of these reviews and associated program countries are in Appendix **Table A2**.

² The Senegal and Madagascar programmes are presented as one since they are based on the same model and have been evaluated jointly.

programme in Cambodia and in donor funded child nutrition programmes in Senegal and Madagascar (Marek et al. 1999, Van de Poel et al. 2016).

3.2 Have they worked? Evidence from evaluated P4P programmes

In section 2.3, we have highlighted examples of unintended consequences in LMIC P4P programmes, although this evidence is limited. Here we describe some findings from quantitative impact evaluations of the selected 14 programmes. As reported in previous systematic reviews, this is intended to give a sense of the relative success or failure of these programmes. For brevity, each of the programmes from **Table A1** is named by the country where it was implemented.

Impact evaluation methods. Most evaluation designs adopted a before and after design and looked at differences between a treatment and a control group. This allows for difference-in-difference analyses, though typically there is little discussion of the validity of the underlying assumption of the ‘treatment’ and ‘control’ groups moving in parallel prior to the treatment (i.e. the parallel trends assumption). Four of the programmes (Afghanistan, Cambodia, the Philippines and Rwanda) employed randomisation in their roll-out which allowed for an arguably more rigorous evaluation.

Improvements in process measures. In many of the selected P4P evaluations, positive findings for process improvement were found; however, the schemes in Tanzania and Zambia did not find any process improvement at all. Common process measures included immunisation rate, antenatal care (ANC) visits completed, tetanus vaccine delivered during ANC visit and facility delivery. While the effect sizes differed across programmes, they may potentially be quite large. This may partly be explained by the differences in baseline levels of various process measures. For example, in the Rwanda programme, there was a 23% increase in institutional deliveries, which had a baseline of 35%, but there is a negligible effect on ANC visits, which have a baseline of 95% (Basinga et al. 2011). In some cases this could be due to the difference in a pilot vs. a scale-up programme. This was seen in the case of the Burundi PBF program where impacts on institutional delivery and antenatal care (ANC) visit were found in the pilot, but not in the scaled version of the program (Bonfrer et al. 2014a, 2014b).

Improvements in quality measures. Several of the selected programmes found improvements in quality measures. One example was in Afghanistan where the programme included a ‘balanced scorecard’ with five domains of in total 20 quality indicators. Of these, significant impacts were only seen for three indicators in the domain of service provision. These three were: time spent with patients, completeness of medical histories and the amount of counselling provided (Engineer et al. 2016). The percentage point differences in intervention and comparison groups for these three indicators were 5.9%, 6.2% and 4.1% respectively.

Improvements in health outcomes. Two examples of programmes that reported impacts on health outcomes were in the Philippines and in Senegal/Madagascar. The former found an improvement in self-reported health and the latter found a reduction in severe and moderate malnutrition. Neonatal mortality rate (NMR) has been another health outcome of interest, and is potentially sensitive to the incentivized maternal and child health efforts that are a focus of several P4P schemes. However, results from the schemes in Cambodia and Rwanda failed to find any impact on NMR (Chari and Okeke 2014, Van de Poel et al. 2016). This also highlights the point that health outcomes are produced by a complex array of factors, including quality of services, and that incentive programmes do not always lead to the desired effects – simply incentivizing service delivery is not enough to improve outcomes.

Complementarities with demand-side programmes. Several of the selected P4P schemes were also introduced with demand-side incentive programs such as vouchers and cash transfers. An evaluation of a combined P4P scheme and cash transfer scheme in Nicaragua found that there were

positive impacts from such a combined approach. Additionally, in the short-term the removal of the demand-side cash transfer did not diminish the positive impact (Regalia and Castro 2007). The programme in Cambodia combined a voucher for delivery with a P4P scheme and found that there was a significant impact in terms of improved institutional delivery (Van de Poel et al. 2016). However, this complementary impact was not seen for the poorest women, leaving the question of how to ensure more equitable outcomes from these combined approaches.

Across a diverse set of programmes, country contexts and evaluation approaches, we show predictably that there are mixed results. In part, this may be due to the variable methods applied for evaluating impact, but this may also be due to differences in the scheme design and context (i.e. demand and supply factors) where they are conducted. Overall, there is more evidence of effects on process, or output measures than there are for health outcomes, and there have been several examples where quality measures have been examined.

4 Conclusion

P4P can be understood as a response to being imperfectly able to condition payment to match all aspects of the delivery of health care that a funder could potentially be concerned with. The imperfection is important because in practice all schemes encounter trade-offs – by paying for each person treated, one might get more treatment but increase the risk of missing the most vulnerable, or costly individuals. Schemes are limited by the information that can be observed to condition payment on and the ability of the funder to ensure that the delivery organisation does not manipulate data. There is no guarantee that P4P will actually improve outcomes. If unintended consequences, such as multi-tasking, gaming and selection are important enough then it may be better to make an unconditional payment. In any event the detail is crucial – what is being contracted over, who is doing the contracting, what they can observe, how they choose to structure the payment (linear or non-linear, with sticks or carrots) will all play a role in determining the success of a P4P scheme. Schemes that appropriately condition payment will be highly complex, and understanding the practical constraints that can operate may lead to relatively simple schemes, with a smaller number of choices that need to be made in terms of design.

These issues have been extensively discussed and analysed in relation to health care in high income countries, most especially with regard to hospital services where there has been a substantial take-up of fixed price (linear) prospective payment systems for hospital services (Roland and Dudley 2015). This has led to the establishment of incentives for improving quality of health care and controlling costs (Roland and Dudley 2015, Markovitz and Ryan 2016). However, since these issues are specific to the concerns and priorities of high income country health care systems, it remains to be seen how they will play out in LMICs that tend to face not only shortages in the quality but also in the quantity of health care provided. As this paper has shown, the evidence on P4P in LMICs is still in its infancy, both in terms of evidence of impact (especially as far as health outcomes are concerned), and in particular in terms of the attention to potential unintended consequences, and how they may be contained.

5 Tables

Table A1.1 Extracted P4P programmes – characteristics

Table. Characteristics of selected P4P programmes in LMIC*

		Afghanistan**	Bangladesh	Bolivia	Burundi	Cambodia	Costa Rica	DRC	Haiti	Nicaragua	Philippines	Rwanda	Senegal & Madagascar	Tanzania	Zambia
Characteristics	Programme name	un-named (contract out)	Bangladesh urban primary health care (contract out)	Bolivia urban PHC (El Alto health service - contract out)	Burundi nationwide PHC P4P	Cambodia rural PHC and district hospital svcs (contract out)	PHC worker cooperatives	PBF in Southern Kivu	Bonus for NGOs delivering PHC in rural areas	Social protection network (RPS)	Philippine Child Health Experiment/ Quality Improvement Demonstration Study (QIDS)	Rwanda PBF	Senegal: Community Nutrition Project (CNP); Madagascar: Secaline project. (contracting out)	Tanzania Mission PBF	Zambia Mission PBF
	Time frame	2003-2005	1998 pilot, 2005 scale up	1999	2006 pilot with national scale-up in 2011	1998 pilot	1988, scale up in 2000	2005	FFS 1995, switch to P4P in 1999	2000-2002 pilot, extended for 5 years	2003-2007	3 districts, Butare in 2001 and Cyangugu in 2002, Kigali in 2005. P4P component introduced in 2005.	Senegal: 1996 pilot, 1998 scale up. Madagascar: 1994 start and scale up	2006-2008	2007-2008
	Principal	World Bank /USAID/EC	Gov. and ADB	Gov. and MOH	NGO & donors	MOH and ADB	Gov.	Gov. and NGOs	Donor and iNGO	Gov. and IADB	Gov. (Philippine Health Insurance Corp.)	Donor Gov and iNGO	WBank/WFP/ German Dev. Bank/Japan/ UNICEF and Gov.	iNGO	iNGO
	Agent	NGOs and MOPH	4 Local NGOs	Local NGO	health facilities	NGO's	primary care cooperatives	health facilities	NGOs	NGOs and private for profit providers	Physicians in district community hospitals	public and private not for profit facilities	Local NGOs	faith-based facilities	faith-based facilities
	Services	Essential PHC services	immunization, prenatal and obstetric care, FP, disease management	PHC services	PHC services	rural PHC, district hospital services	PHC services	preventive care, TB, HIV/AIDS testing	basic healthcare services	child health, reproductive health, maternal health	hospitals services	maternal and child health, curative services, HIV/AIDS (immunization, prenatal care, deliveries)	Nutrition services	essential health services package	HIV, maternal health, hospital svcs (PHC in one facility)
	Targeting***	General population	Poor households	General population	Not identified	Poor households	General population	General population	General population	Poor households. Geographic combined with hh-level assessment of assets.	Hospitals were primarily from poor districts	General population	Poor households. 1st stage geographic, 2nd stage based on nutrition/health status.	General population	General population

Table A1.2 Extracted P4P Programmes – Programme Design

		Afghanistan**	Bangladesh	Bolivia	Burundi	Cambodia	Costa Rica	DRC	Haiti	Nicaragua	Philippines	Rwanda	Senegal & Madagascar	Tanzania	Zambia
Payment Scheme Design	Scheme design	Payment of fixed amount plus performance bonus	specific coverage targets	Contract based on achieving process and outcome indicators	Fixed amount plus quality bonus	mix of contracting-in and contracting-out	payment based on service production and coverage	Fixed amount plus quality bonus	Fixed amount plus quality bonus	payment based on achieving pre-defined targets	payment if certain CPV quality score is met	pay for incremental svcs (pilot). Pay according to quality-adjusted quantity of svcs (scale-up).	Payment of fixed amount plus performance WFP/German Dev. performance/ Japan/UNICEF coverage country payment based on service production and coverage	fixed payment plus performance bonus	fixed payment plus performance bonus
	Indicators	Outputs, nationally defined management indicators, based on MICS 2003 (World Bank). Inputs (EC/USAID).	Output only. (e.g. # of centres providing immunisation, FP, or lab tests)	Ouput only. (eg. # of instit. Deliveries and output visits)	menu of utilisation indicators	incidence of sickness, incidence of diarrhea in children, infant mortality, service utilisation	service utilisation, general mortality, child mortality	coverage indicators of basic services (e.g. FIC, women protected from tetanus, assisted deliveries, HIV+ on ARV)	Output: % of clinics with 4 methods of FP. Outcome: use of ORT for diarrhea, immunization coverage, coverage of 3 antenatal visits. High-level: coordination with MOH	group-specific performance targets	Health status of children under 6. Indicators of weight, height and blood test related to pneumonia and diarrhea	assisted deliveries, FIC, tetanus immunization of Preg Women, acceptance of FP and HIV testing.	(outcome and impact level) % malnourished children, child anthropometry, % children weighed monthly, % women attending education sessions	(multiple levels). Availability of essential drugs, % of supervised deliveries, # new VCT, utilisation of output and input	(multiple levels). Availability of essential drugs, % of supervised deliveries, # new VCT, utilisation of output and input
	Quality indicators****	(Unsure)	% clients reporting waiting times are acceptable, % prescriptions with specific diagnosis.		Composite quality index (153 indicators, later reduced)	Perceived quality of care used as indicator					Clinical performance vignettes (CPV)	Quantity adjusted quality (post scale up). Quality indicators include inputs (staffing, drugs, etc.) and processes			
	Payment type	Bonus (World Bank). Withholding (USAID)	Bonus	Withholding* ****	Payment based on service utilisation and reward withheld if quality target not met.	Penalties to NGOs for not achieving targets. Bonuses to health workers	Withholding	Fixed amount per targeted action per month plus bonus (as withheld bonus?)	Bonus	Withholding	Bonus	Bonus paid for incremental service provision during pilot. Quality-adjusted quantity during scale up.	Withholding	Up front payment and retrospective payment (withholding)	Up front payment and retrospective payment (withholding)

					Quantitative-based payments paid monthly (FFS), quality-related payments given as quarterly bonuses (15% of quantitative payment, increased to 25% later).	HCW salaries: 55% basic, 15% bonus for punctuality, 30% bonus for performance if monthly financial targets were met.	NA	Reward is 15% of fixed amount if quality score was 100% and proportionally less for lower scores.	95% of budget as fixed and up to 10% bonus for achieving targets	3% of annual budget paid up front. Payments were made quarterly or biannually for reaching coverage of groups enrolled in RPS programme as "all or nothing". FFS for services for up to 10% of households not in RPS.	amount of bonus = no. of patients times 100 Philippine pesos	Monthly payment per indicator multiplied by quality index (range: 0-1). Varying % of performance bonus (40%-95%) forwarded to staff, by district	NA	Withholding. Guaranteed 50% and performance based 50% reward paid every 6 mos. P4P programme is only 8-10% of total budget overall so reward was around 4%.	Withholding. Guaranteed 50% and performance based 50% reward paid every 6 mos. P4P programme is 17% of total budget overall.
	Payment Size and Schedule	10% of contract value, paid in stages. Final 5% at end of contract(World Bank)	NA	NA											
	Who Received \$\$\$	NGO	NGO	NGO	facilities, then have autonomy over how it is used. No more than 50% can go to staff incentives, rest must go to improve service quality	NGO	Cooperatives received and had autonomy over funds	health facilities	NGO	contracted health care provider	Physician	health facilities, have discretion over usage	NGO	health facilities (though donor determined max % of allocation on staff, infrastructure, running costs, etc.)	health facilities (though donor determined max % of allocation on staff, infrastructure, running costs, etc.)

Table A1.3 Extracted P4P Programmes – Evaluation Evidence

		Afghanistan**	Bangladesh	Bolivia	Burundi	Cambodia	Costa Rica	DRC	Haiti	Nicaragua	Philippines	Rwanda	Senegal & Madagascar	Tanzania	Zambia
Evaluation Evidence	Evidence of improved non-monetary incentives	YES			external evaluation found provider motivation improved when they had autonomy over where incentives allocated				YES			YES		YES	YES
	Evidence of perverse incentives					YES (Bloom 2006)						Suggestive (Kalk 2010), also some evidence of gaming		Suggestive (Canavan 2008)	Yes (Vergeer 2008)
	Quantitative evaluation design	broad comparisons	Controlled before and after study (Mahmud 2002)	Before and after comparison with a control group/district (double differences)	Controlled cohort study (Rudasingwa 2015)	Randomised assignment of districts + double differences	interrupted time series (single difference) over 10 years compared to publicly managed facilities (Gauri 2004, Loevinsohn 2008)	Before and after comparison with a control group/district (double differences)	Before and after (without control group)	Before and after comparison with a control group/district (double differences)	Randomized assignment to district hospitals + DiD	Randomised assignment of districts + double differences	Before and after in Senegal	Controlled before and after	Controlled before and after
	Evaluation results	Rapid increase in services that appear better in programmes with P4P. Better performance on quality scorecards for P4P programmes	11% improvement in quality score, 3.4% improvement on household survey indicators	21% DiD for deliveries and 1% DiD for bed occupancy (MC vs. control)	38-66% increase in quality indicators. Bonfrer 2014a,b find increase in quality and utilisation of MCH (but not equitably and not for all services).	SDC and MC better than control on 7 indicators. (21.3% SDC vs. control; 9.3% MC vs. control). Van de Poel 2015 find more deliveries in public facilities.	22% more general visits and 42% more dental in cooperatives and decrease in medical expenditures	Patients in P4P pay less for similar or better quality	single diff: -3% prenatal care, +32% vaccination coverage	5% increase in children <3 accessing preventive health checks (18.1% intervention, 13.1% control)	Improvements in self-reported health (improved 7% in treated) and wasting (incr. 9% in control).	23% increase in instit. Deliveries. Increase in child preventive visits. Improvements in prenatal quality scores.	severe and moderate malnutrition declined 6% and 4%	Nothing significant	Nothing significant

Table A2. Programmes Covered in Literature Reviews of P4P Programmes

Source	Grittner 2013	Morgan 2011	Eichler 2013	Witter 2012 (Cochrane)	Witter 2013 (agent)	Loevinsohn 2005 (contracting)	LaGarde 2009 (Cochrane)	Liu 2008
Number	PBF programs	PBI programs	PBI programs	P4P of health svcs	P4P with some MNCH component	Contracting programs	Contracting out	Contracting Out
1	Afghanistan	Afghanistan	Afghanistan	Burundi	Afghanistan (ngo, gov)	Cambodia	Bolivia	Bangladesh
2	Bangladesh (urban)	Belize	Bangladesh	China	Argentina (province)	Bangladesh (rural)	Cambodia	Bangladesh
3	Bolivia	Benin	Cambodia	DRC	Brazil (local gov)	Bangladesh (urban)	Pakistan	Bolivia
4	Burundi	Burundi	DRC	Egypt (pending results)	Burundi (facility)	Bolivia		Cambodia
5	Cambodia	DRC (Cordaid)	Egypt (social insur)	Philippines	Cambodia (district)	Guatemala		Costa Rica
6	Costa Rica	DRC (EC)	Haiti	Rwanda	Cambodia (midwife)	Haiti		Croatia
7	DRC	DRC (Wbank)	Nepal	Rwanda (Basinga)	DRC (facility)	India (urban TB)		Guatemala
8	Haiti	Egypt	Philippines	Rwanda (Soeters)	DRC (ngo)	India (private practitioners)		Haiti
9	Nicaragua	Honduras	Rwanda	Tanzania	Egypt (facility)	Madagascar & Senegal		India
10	Rwanda	Rwanda		Uganda (pending results)	Haiti (ngo)	Pakistan		Madagascar
11	Senegal & Madagascar	Senegal		Vietnam	India (gov agency)			Senegal
12	Tanzania	Tanzania		Zambia	India (maternal hcw)			South Africa
13					Liberia (ngo)			
14					Nepal (hcw)			
15					Nicaragua (facility)			
16					Pakistan (hcw)			
17					Philippines (facility)			
18					Philippines (maternal hcw)			
19					Rwanda (facility)			
20					South Sudan (intl ngo)			
21					Tanzania (facility)			
22					Tanzania (hcw)			
23					Uganda (facility)			
24					Zambia (facility)			

Others

Miller 2014 - lists 26 P4P programs that haven't been formally evaluated

References

- Baron DP, Myerson RB. (1982). Regulating a monopolist with unknown costs. *Econometrica* 50(4): 911-930.
- Basinga P, Gertler PJ, Binagwaho A, Soucat AL, Sturdy JR, Vermeersch C. (2010). *Paying primary health care centers for performance in Rwanda*. Policy Research Working Paper No.5190. World Bank, Washington DC.
- Basinga PP, Gertler J, Binagwaho A, Soucat AL, Sturdy J, Vermeersch CM (2011). Effect on maternal and child health services in Rwanda of payment to primary health-care providers for performance: an impact evaluation. *Lancet* 377(9775): 1421-1428.
- Bertone MP, Mangala A, Kwété D, Derriennic Y. (2011). *Review of the results-based financing experiences in the Democratic Republic of the Congo*. Bethesda, Md.: Health Systems, 20, 20.
- Bhushan I, Keller S, Schwartz B. (2002). *Achieving the twin objectives of efficiency and equity: contracting health services in Cambodia*. ERD Policy Brief Series No.6. Asian Development Bank.
- Bloom E, Bhushan I, Clingingsmith D, Hong R, King E, Kremer M, Loevinsohn B, Schwartz JB. (2006). *Contracting for health: evidence from Cambodia*.
- Bonfrer I, Soeters R, Van De Poel E, Basenya O, Longin G, Van De Looij F, Van Doorslaer E. (2014a). Introduction of performance-based financing in Burundi was associated with improvements in care and quality. *Health Affairs*, 33, 2179-2187.
- Bonfrer I, Van De Poel E, Van Doorslaer E. (2014b). The effects of performance incentives on the utilization and quality of maternal and child care in Burundi. *Social Science & Medicine*, 123, 96-104.
- Busogoro J-F, Beith A. (2010). *Pay-for-performance for improved health in Burundi*. Bethesda, MD: Health Systems, 202, 20.
- Canavan A, Swai G. (2008). *Payment for performance (P4P) evaluation: 2008 Tanzania Country Report for Cordaid*. KIT Development Policy & Practice. Amsterdam, Royal Tropical Institute, KIT.
- Cercerone J, Briceno R, Gauri V. (2005). Contracting PHC services: the case of Costa Rica. In: Laforgia, G. (ed.) *Health systems innovations in Central America: lessons and impact of new approaches*. Washington, DC: World Bank.
- Chari AV, Okeke EN. (2014). *Can institutional deliveries reduce newborn mortality?* RAND Labor and Population Working Paper, RAND: 47.
- Dafny L. (2005). How do hospitals respond to price changes? *American Economic Review* 95(5): 1525-1547.
- Eggleston K. (2005). Multitasking and mixed systems for provider payment. *J Health Econ* 24(1): 211-223.
- Eichler R, Auxila P, Antoine U, Desmangles B. (2007). Performance-based incentives for health: six years of results from supply-side programs in Haiti. *Center for Global Development working paper*.
- Eijkenaar F, Emmert M, Scheppach M, Schoffski O. (2013). Effects of pay for performance in health care: a systematic review of systematic reviews. *Health Policy* 110(2-3): 115-130.

- Eldridge C, Palmer N. (2009). Performance-based payment: some reflections on the discourse, evidence and unanswered questions. *Health Policy Plan* 24(3): 160-166.
- Ellis RP. (1998). Creaming, skimping and dumping: provider competition on the intensive and extensive margins. *J Health Econ* 17(5): 537-555.
- Engineer CY, Dale E, Agarwal A, Agarwal A, Alonge O, Edward A, Gupta S, Schuh HB, Burnham G, Peters DH. (2016). Effectiveness of a pay-for-performance intervention to improve maternal and child health services in Afghanistan: a cluster-randomized trial. *Int J Epidemiol* 45(2): 451-459.
- Gauri V, Cercone J, Briceno R. (2004). Separating financing from provision: evidence from 10 years of partnership with health cooperatives in Costa Rica. *Health Policy and Planning*, 19, 292-301.
- Grittner AM. (2013). *Results-based financing: evidence from performance-based financing in the health sector*. Bonn, Germany, German Development Institute.
- Kaarboe O, Siciliani L. (2011). Multi-tasking, quality and pay for performance. *Health Econ* 20(2): 225-238.
- Kalk A, Paul FA, Grabosch E. (2010). 'Paying for performance' in Rwanda: does it pay off? *Trop Med Int Health* 15(2): 182-190.
- Kuhn M, Siciliani L. (2009). Performance indicators for quality with costly falsification. *Journal of Economics & Management Strategy* 18(4): 1137-1154.
- Laffont J-J, Martimort D. (2002). *The theory of incentives: the principal-agent model*. Princeton, NJ, Princeton University Press
- Lagarde M, Palmer N. (2009). The impact of contracting out on health outcomes and use of health services in low- and middle-income countries. *Cochrane Database Syst Rev*(4): CD008133.
- Lannes L, Meessen B, Soucat A, Basinga P. (2016). Can performance-based financing help reaching the poor with maternal and child health services? The experience of rural Rwanda. *Int J Health Plann Manage* 31(3): 309-348.
- Lannes L, Meessen B, Soucat A, Basinga P. (2015). Can performance-based financing help reaching the poor with maternal and child health services? The experience of rural Rwanda. *The International Journal of Health Planning and Management*.
- Lavadenz F, Schwab N, Straatman H. (2001). Redes públicas, descentralizadas y comunitarias de salud en Bolivia. *Rev Panam Salud Publica*. Vol.9 (3) Washington.
- Le Grand J. (2003). *Motivation, agency, and public policy: of knights and knaves, pawns and queens*. Oxford Scholarship Online.
- Leonard KL, Masatu MC. (2010). Professionalism and the know-do gap: exploring intrinsic motivation among health workers in Tanzania. *Health Econ* 19(12): 1461-1477.
- Liu X, Hotchkiss DR, Bose S. (2008). The effectiveness of contracting-out primary health care services in developing countries: a review of the evidence. *Health Policy Plan* 23(1): 1-13.
- Loevinsohn B. (2008). *Performance-based contracting for health services in developing countries: a toolkit*, World Bank Publications.

Loevinsohn B, Harding A. (2005). Buying results? Contracting for health service delivery in developing countries. *The Lancet*, 366, 676-681.

Mahmud H, Ullah Khan A, Ahmed S. (2002). Mid-term health facility survey—urban primary health care project. *Dhaka: Mitra and Associates*.

Marek T, Diallo I, Ndiaye B, Rakotosalama J. (1999). Successful contracting of prevention services: fighting malnutrition in Senegal and Madagascar. *Health Policy Plan* 14(4): 382-389.

Markovitz AA, Ryan AM. (2016). Pay-for-performance: disappointing results or masked heterogeneity? *Med Care Res Rev*.

Meessen B, Kashala J-P I, Musango L.(2007). Output-based payment to boost staff productivity in public health centres: contracting in Kabutare district, Rwanda. *Bulletin of the World Health Organization*, 85, 108-115.

Miller G, Babiarz KS. (2014). Pay-for-performance incentives in low- and middle-income country health programs. in Culyer AJ. (Ed) *Encyclopedia of Health Economics*. Elsevier: 457-466.

Peabody JW, Shimkhada R, Quimbo S, Solon O, Javier X, McCulloch C. (2014). The impact of performance incentives on child health outcomes: results from a cluster randomized controlled trial in the Philippines. *Health Policy Plan* 29(5): 615-621.

Regalia F, Castro L. (2007). *Performance-based incentives for health: demand- and supply-side incentives in the Nicaraguan Red de Proteccion Social*. CGD Working Paper. Washington, DC, Center for Global Development: 52.

Roland M, Dudley RA. (2015). How financial and reputational incentives can be used to improve medical care. *Health Serv Res* 50 Suppl 2: 2090-2115.

Shen Y. (2003). Selection incentives in a performance-based contracting system. *Health Serv Res* 38(2): 535-552.

Rudasingwa M, Soeters R, Bossuyt M. (2015). The effect of performance-based financial incentives on improving health care provision in burundi: a controlled cohort study. *Global journal of health science*, 7, 15.

Rusa L, Schneidman M, Fritsche G, Musango L. (2009). Rwanda: Performance-based financing in the public sector. in *Performance incentives for global health: potential and pitfalls*. Center for Global Development.

Skiles MP, Curtis SL, Basinga P, Angeles G, Thirumurthy H. (2015). The effect of performance-based financing on illness, care-seeking and treatment among children: an impact evaluation in Rwanda. *BMC Health Services Research*, 15, 1.

Soeters R, Griffiths F. (2003). Improving government health services through contract management: a case from Cambodia. *Health Policy and Planning*, 18, 74-83.

Soeters R, Musango L, Meessen B. (2005). Comparison of two output based schemes in Butare and Cyangugu provinces with two control provinces in Rwanda. *Global Partnership on Output Based Aid (GPOBA)*.

Soeters R, Peerenboom PB, Mushagalusa P, Kimanuka C. (2011). Performance-based financing experiment improved health care in the Democratic Republic of Congo. *Health Affairs*, 30, 1518-1527.

Sondorp E, Palmer N, Strong L, Wali A. (2009). Afghanistan: Paying NGOs for performance in a postconflict setting. *Eichler R, Levine R; the Performance-Based Incentives Working Group. Performance incentives for global health: potential and pitfalls. Washington, DC: Center for Global Development*, 139-64.

Songstad NG, Lindkvist I, Moland KM, Chimhutu V, Blystad A. (2012). Assessing performance enhancing tools: experiences with the open performance review and appraisal system (OPRAS) and expectations towards payment for performance (P4P) in the public health sector in Tanzania. *Global Health* 8: 33.

Sun X, Liu X, Sun Q, Yip W, Wagstaff A, Meng Q. (2016). The impact of a pay-for-performance scheme on prescription quality in rural China. *Health Econ* 25(6): 706-722.

Van de Poel E, Flores G, Ir P, O'Donnell O. (2016). Impact of performance-based financing in a low-resource setting: a decade of experience in Cambodia. *Health Econ* 25(6): 688-705.

Vergeer P, Chansa C. (2008). *Payment for performance (P4P) evaluation: 2008 Zambia Country Report for Cordaid*. KIT Development Policy & Practice. Amsterdam, KIT: 94.

Witter S. (2013). *Pay for performance for strengthening delivery of sexual and reproductive health services in low- and middle-income countries: Evidence Synthesis Paper*. Washington, DC, The World Bank: 72.

Witter S, Fretheim A, Kessy FL, Lindahl AK. (2012). Paying for performance to improve the delivery of health interventions in low- and middle-income countries. *Cochrane Database Syst Rev*(2): CD007899.