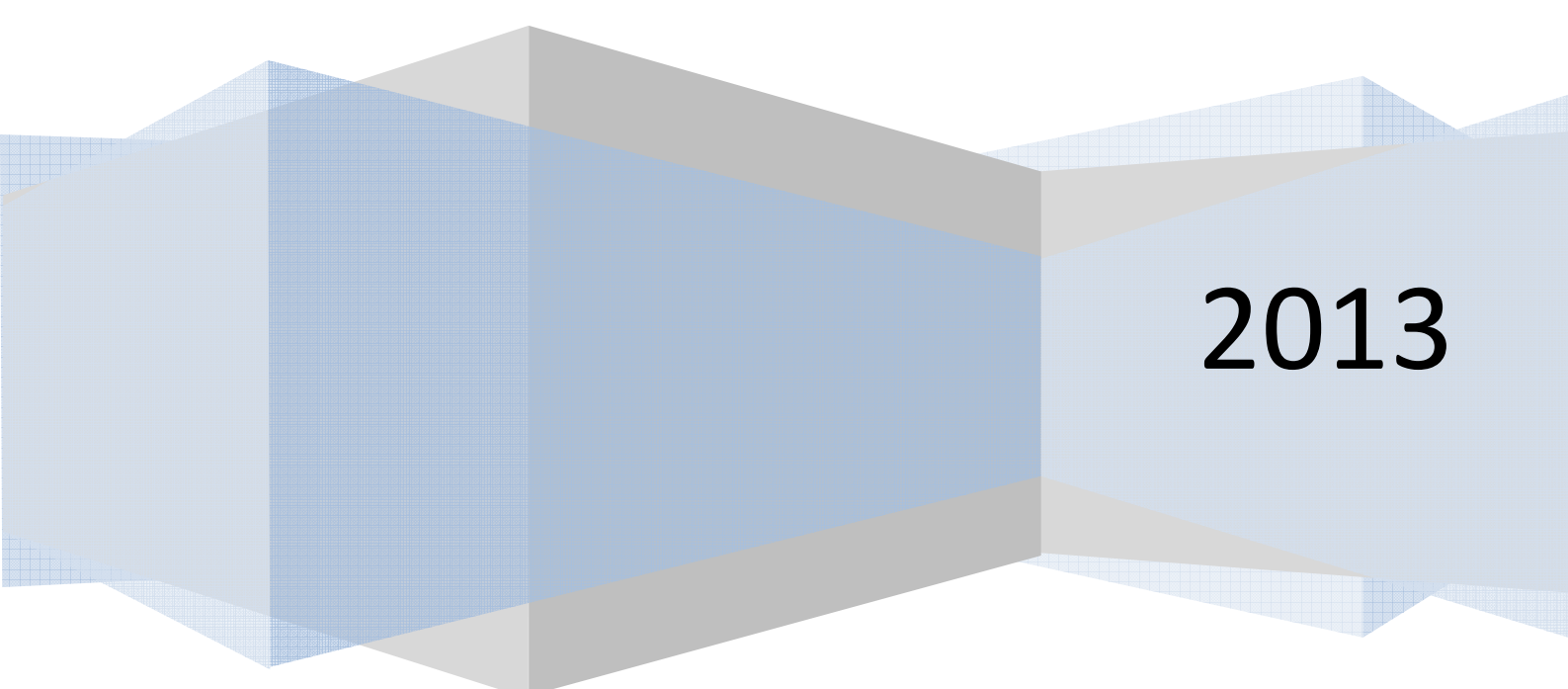


The World Bank

Pay for performance for strengthening delivery of sexual and reproductive health services in low- and middle-income countries Evidence synthesis paper

Sophie Witter



2013

Acknowledgements

This paper was commissioned by the World Bank as a document to support government decision-making in the East Asia Region.

Many thanks to Seemeen Saadat for her assistance with the literature search and summary tables. The paper has also benefited from guidance from Aparnaa Somanathan. Any errors however are the responsibility of the author's.

Executive summary

Background

This paper aims to bring together the global evidence on paying providers for performance (P4P), its impact on the delivery of sexual and reproductive health services, and the conditions under which it may have been effective. It is based on a literature review carried out in November-December 2011, with some updating in 2013. It synthesises evidence from policies and projects which have been documented and published to date. The sources include the few available published impact evaluations as well as the more extensive internal reports focussing on early implementation experiences. It focuses on supply-side measures, and complements a recent report on demand-side financing for SHR services in low and middle-income countries.

Definitions, modalities, goals and risks

P4P refers to the transfer of money or material goods conditional on taking a measurable action or achieving a predetermined performance target. It goes by a number of different names and has a range of modalities, including different levels of payment (to governments, local governments, NGOs, facilities and individual health workers) and combinations of these levels, different targeted outcomes, payment systems and magnitude of transfers.

While in higher income countries, P4P has been to a large extent aimed at improving quality of care, in LMICs, the objectives have been wider, including to increase the allocative efficiency of health services (by encouraging the provision of high priority and cost effective services); to increase their technical efficiency (by making better use of existing resources such as health staff); and to improve equity of outcomes (for example, by encouraging expansion of services to hard-to-reach groups). Some have also argued that P4P has the potential to transform health systems.

However, P4P relies on a set of assumptions, which may be more or less accurate for specific contexts, and also poses potentially serious risks. These need careful monitoring and management. They include:

- distortion – encouraging health staff to ignore important services which are not rewarded with incentives, including untargeted services and more complex dimensions of performance
- cherry-picking – focussing health workers on services which provide greatest gain with least effort over others which may be as or more important; this may also increase inequity in some cases (where poorer populations are harder to reach, for example)
- gaming - improving or cheating on reporting rather than improving performance
- financial dependency – problematic if P4P is not sustained and has caused raised expectations from staff
- inefficiency – high operating costs and low returns may make this a poor investment
- fiduciary risks – if funds are poorly controlled, then leakage and corruption may be increased
- coercion – if clients are pressured to accept a service or undergo a procedure to increase provider rewards
- undermining intrinsic motivation of staff, and cooperative behaviour between staff

Analysis of policies and projects

P4P has spread quickly in the health sector in low- and middle-income countries over the past decade. It has been deployed in relation to two groups of services in particular – control of infectious diseases and sexual, reproductive and child health services. More than twenty countries are now applying P4P in some form. Details are presented on 25 documented case studies, covering Latin America, Asia and Africa, including two global health initiatives; two cases of performance-based

contracting of local authorities and seven of NGOs; nine examples of PBF directed mainly at health facilities; and five examples focussed on individual health worker incentive payments.

However, the literature on P4P remains limited. While there are many reports written by designers and implementers of P4P programmes, there are few independent evaluations using robust study design. A recent systematic review of the topic has highlighted a number of areas of concern in relation to the quality of evidence. Further work is now underway to generate more impact evaluation evidence in 16 countries.

A conceptual framework is developed and applied here to analyse evidence on the impact of P4P, including on utilisation of care, quality of care, organisation of services, provider motivation and behaviour, equity and patient payments, unintended consequences, health outcomes, costs and cost effectiveness, and sustainability. The most common domain which is reported is utilisation, but even here only seven of the 25 studies had findings which could be linked to the P4P intervention and the findings were mixed. There were some positive impacts for some indicators relating to facility deliveries and family planning, but no consistent pattern across schemes, in terms of which indicators responded positively, or the magnitude of response. This is likely to be linked to the different designs of schemes, the different payment amounts and systems, and the different starting points and degree of effort involved in changing indicators.

For almost all impact areas there is a striking lack of evidence (especially for effects on equity, provider behaviour, organisation of care, and health outcomes). Cost effectiveness has not been assessed for any of the programmes. Deeper and broader studies are required to understand P4P and its impact on targeted indicators as well as wider systemic effects.

Lessons on design and implementation

The experiences documented are reviewed to derive lessons relating to design of P4P programmes, including on how performance is defined; the level of targeting of payments; how payments are calculated and how large they should be; how many targets should be used; types of services which are suitable; verification requirements; complementary measures; and purchasing arrangements. Many of these issues are necessarily very contextual.

Some of the more developed schemes are based in post-conflict areas, and there is some evidence that these can present suitable conditions for P4P, paradoxically (commentators emphasise the necessary organisational preconditions, which are ambitious – nevertheless, there may be more need and scope for reform in post-conflict areas).

Despite the rhetoric relating to outcomes, most schemes fund outputs and some specify quite detailed levels of process. This raises the questions of the degree to which P4P increases provider autonomy, which is one of the channels through which it might bring about gains.

Conclusions and outstanding questions

It is clear that P4P can have a positive influence on health outputs, at least in some contexts, and in the short term. However, there are a number of important areas which remain to be better understood. These include:

- the long-term impact of P4P, assuming it is intended as a long-term financing mechanism as opposed to a short-term behaviour modifier
- its health systems impact, particularly in terms of:
 - integration/fragmentation of financing channels
 - effectiveness of purchasing arrangements

- information systems
- pay policy
- non-targeted services (and how adverse effects can be controlled)
- capacity building
- governance arrangements
- its impact on health outcomes, which study designs have not yet been able to reveal
- its equity effects and impact on patient payments
- analysis of the cost-effectiveness of the approach, particularly related to alternative methods of achieving similar goals (wider performance management tools; increasing provider autonomy; direct facility funding; pay reforms)
- its sustainability, given the high degree of donor dependence to date (financial and technical)
- its strengths and weaknesses as a mechanism for international aid, especially in terms of consistency with the Paris Declaration and its impact on aid harmonisation and transactions costs
- how to increase its efficiency – particularly reducing the high overhead costs recorded in most schemes to date
- its effects on health workers' motivation and behaviour (in the LMIC setting), and how to maintain motivation over longer periods of time
- a better understanding of the contextual factors favouring the use of P4P, including in relation to the process of its introduction, and to the necessary systems capacity prior to introduction
- learning on design issues, such as how to design payment systems to minimise inflationary tendencies and promote quality of care, as well as on the magnitude of payments required to improve performance in different settings

The paper concludes that overall evidence base remains weak. P4P may be beneficial in some settings, but should not be introduced without a careful diagnosis of the blockages in the health system, and consideration of the advantages and disadvantages of all strategies to address them.

Contents

Acknowledgements.....	2
Executive summary.....	3
Objectives of paper.....	8
Research methods.....	8
Background.....	9
Definitions and modalities.....	9
Rationale for P4P.....	10
Assumptions and risks.....	11
Scale and scope of PBF for sexual & reproductive health.....	12
How policies were introduced.....	16
Evidence to date of impact.....	17
Framework for assessment.....	17
Quality of current evidence.....	18
Utilisation of services/coverage of programmes.....	19
Quality of care.....	20
Organisation of care.....	21
Provider motivation and behaviour.....	21
Equity, access and patient payments for care.....	22
Unintended consequences.....	23
Health outcomes.....	24
Costs and cost effectiveness.....	24
Magnitude of payments.....	24
Relative size of payments.....	24
Transaction costs.....	25
Cost-effectiveness.....	25
Sustainability.....	26
Observations and lessons learned on design.....	27
Defining performance.....	27
Level of targeting of payments.....	27
Size of payments.....	27
How payments are calculated.....	28
Number of targets.....	28
Payment systems.....	29
Verification.....	29
Service or target type.....	29
Stand-alone or linked to demand-side measures.....	30

Purchasing arrangements	30
Implementation lessons.....	30
Application in different contexts	31
Organisational requirements	31
Discussion and some outstanding questions.....	32
Conclusions	35
References and bibliography	37

List of tables and figures

Table 1A Description of PBF schemes.....	43
Table 1B Description of PBF schemes.....	49
Table 2A: Impact of PBF schemes	60
Table 2B: Impact of PBF schemes	64
Table 3 Cost and sustainability of PBF schemes	69
Figure 1 Conceptual framework for assessing impact of PBF programmes	18

Objectives of paper

Improving the performance of health care delivery systems is an important objective, both in high-income settings and, even more critically, in low- and middle-income (LMIC) settings, where resources for health are much more constrained. Pay for performance (P4P) is currently receiving increased attention as a strategy for improving the performance of healthcare providers, organisations and governments. It is also promoted as an important tool for achieving the health Millennium Development Goals, improving the effectiveness of development aid, and motivating patients to improve their attendance at health facilities and compliance with recommended health interventions. While P4P is used across a variety of services, targets relating to sexual and reproductive health (SRH) are at the heart of many of the recent schemes. However, there is currently a lack of rigorous evidence on the effectiveness of these strategies in improving health care and health, particularly in lower income countries (Oxman and Fretheim 2008; Eldridge and Palmer 2009; Witter et al. 2012).

The objective of this paper is to summarise the evidence to date on the impact of RBF mechanisms on the delivery of sexual and reproductive health (SHR) services in low and middle-income countries.

Research methods

Search strategy

A literature search was conducted in October-November 2011 of multiple databases - Jstor, The Cochrane Library, Pubmed, IngentaConnect; Springer Link; Science Direct. In addition, we searched Google scholar and Google as well as the World Bank's internal library and the Results Based Financing Website (www.rbfhealth.org). We also reviewed reference lists.

The following key words were used: pay for performance; performance-based funding/finance; output-based funding/aid; results-based funding/finance; target payment; performance-based contracting; supply side financing/funding.

They were combined with SHR terms: reproductive health OR sexual health OR maternal health OR deliveries OR obstetric care OR family planning OR neonatal care OR antenatal care OR postnatal care OR sexually transmitted infections OR HIV OR abortion care OR fertility.

After review the paper was updated in 2013, although a full search of all databases was not re-run at this point.

Study selection

A number of recent papers on P4P have been discursive/viewpoints. These are still relevant for the discussion, but we have focussed on extracting findings from papers containing primary data on the impact of a RBF scheme covering SRH services in a low or middle-income setting.

Within SHR, the following are the main focal areas (following the WHO definition of SHR¹):

- improving antenatal, perinatal, postpartum and newborn care;
- providing high-quality services for family planning, including infertility services;
- eliminating unsafe abortion;

¹ http://www.who.int/reproductivehealth/about_us/en/index.html

- combating sexually transmitted infections, including HIV, reproductive tract infections, cervical cancer and other sexual and reproductive health morbidities;
- promoting sexual health

Exclusions from the review

While it is conceivable that pay increases designed to increase motivation and retention of staff might fall within this definition, in this review we focus on reforms which are explicitly linked to changing patterns of activity, output or outcome indicators (thus excluding routine changes to pay or public funding flows or user fee regimes).

Another summary paper has recently addressed the use of demand-side financing for SHR services in LMICs (Witter 2013). This review therefore focuses on evidence of the impacts of supply-side measures focussed on service providers.

Studies relating purely to high income countries were also excluded (using the World Bank classification of countries).

Background

Definitions and modalities

Pay for performance is commonly understood to refer to the transfer of money or material goods conditional on taking a measurable action or achieving a predetermined performance target (Eichler 2006). The plethora of terms commonly used in this field – results-based financing, performance-based incentives, pay for performance, performance-based contracting, conditional cash transfers, cash on delivery, and others – can cause confusion. Some have taken on a specific meaning (Musgrove 2011), although there is not always full consensus. Conditional cash transfers are commonly used to denote payments or near-cash transfers such as vouchers to beneficiaries. Some terms are specific to aid, such as cash on delivery or output-based aid. Performance-based contracting is used when contracts are drawn up with non-state actors, such as non-governmental organizations. Performance based funding is often used to specify a particular model in which providers are paid retrospectively according to verified outputs, modified by quality measures. Other labels are more general, including results-based financing, and pay for performance. We use the term P4P here to capture all types of supply-side conditional financing.

While paying for performance is a relatively simple and ancient concept, it includes a wide range of interventions that vary with respect to the level at which the incentives are targeted (recipients of healthcare, individual providers of healthcare, health care facilities, private sector organizations, public sector organizations and national or sub-national levels). Paying for performance interventions can also be used to reward a wide range of measurable actions, including health outcomes, delivery of effective interventions (for instance immunization), utilization of services (such as prenatal visits or births at an accredited facility), and quality of care (Witter et al. 2012). Paying for performance interventions typically also include ancillary components, such as increasing the availability of resources for health care, education, supplies, technical support or training, monitoring and feedback, increasing salaries, construction of new facilities, improvements in planning and management or information systems etc (Oxman and Fretheim 2008).

Paying for performance typically takes three main forms:

- Conditional cash payment (payment per output or outcome)
- Conditional provision of material goods

- Target payments (payments for reaching a certain level of coverage, which can be defined in absolute terms or relative to a starting point)

The arrangements in relation to payments and expected outputs (in quantity or quality terms) are often expressed in a contract between purchaser and provider.

It seems likely that schemes targeted at different levels will vary in their characteristics. It may therefore be useful to analyse schemes according to the following categories:

1. Those targeted at national level (e.g. aid programmes, global health initiatives etc.)
2. Those targeted at intermediate levels – regions or districts, for example, or non-governmental organisations
3. Those targeted at facility level
4. Those targeted at individual health worker level

Other factors, such as the magnitude of the payments, the nature of the targeted indicators, starting levels of coverage for those services, and the systems for assessing them would also be expected to play an important role determining the outcome of P4P schemes.

Rationale for P4P

On one level, paying for performance, in terms of outputs, is not new – it has taken the form of user fees, which in many low and middle income countries it remains one of the main forms of health financing. It is also commonly used in insurance and contracting payment systems. However, public funding for health (including aid funding, where this is channelled through governments) has traditionally not been linked to specific activities, but has taken the form of budget flows, which are linked to indicators such as staffing levels or bed numbers (for facilities), inputs (such as estimated drug needs), population numbers (for regions and districts, in some cases) and also historical trends in expenditure (all modified by overall budget constraints) (Witter et al. 2012).

These bureaucratic mechanisms offer the advantage of stability, and predictability, and rely on local clinical judgement as to how and what services to offer. The disadvantage, however, is that health systems based on budget funding and salaried staff can lack incentives to improve quality, to increase outputs and to improve outcomes. Paying for performance aims to reintroduce those incentives by linking pay (at individual or facility level) to desired activities and/or outcome indicators. It may in addition increase resources (by providing supplementary funding) or may be an alternative mechanism for channelling existing funding resources (substituting for existing funds).

In Organisation for Economic Co-operation and Development (OECD) countries, paying for performance is generally described as a tool for improving quality (Christianson, Leatherman, & Sutherland 2007; Petersen et al. 2006). In LMICs, however, it generally has wider objectives. These include (Witter et al. 2012):

- to increase the allocative efficiency of health services (by encouraging the provision of high priority and cost effective services)
- to increase their technical efficiency (by making better use of existing resources such as health staff)
- to improve equity of outcomes (for example, by encouraging expansion of services to hard-to-reach groups)

Others emphasise the potential of some forms, such as PBF, to transform health sectors, introducing client-oriented public finance models inspired by the new public management mode (Meessen et al. 2011). PBF aims at improving provider performance through allowing providers greater control over

resources and encouraging system efficiencies. It aims to shift the performance risk to the provider of service, making them directly responsible for the agreed outputs. When designed properly, it is argued, PBF can be an efficient tool for improving organization, production, management, and quality of services through influencing the behaviour of healthcare providers (Eichler 2006, Brenzel 2009, Meessen et al. 2011). There is also a widespread perception that existing methods of funding providers are unreliable in getting resources to the front-line and that P4P may offer a more effective channel.

Assumptions and risks

Paying providers for performance is clearly premised on the assumption that for these three dimensions to shift, a change in behaviour on the provider side is required. If, however, the barriers to service uptake and use are more connected with demand side factors (such as low affordability of services), then paying for performance for providers alone will not be effective. However, by incentivising providers to increase outputs, it is hoped that they will in turn take measures to boost demand, for example by reducing fees or making services more attractive.

One theoretical advantage of performance pay is that explicit financial incentives are provided even when patient demand for health care is unresponsive to (unable to accurately assess) technical quality of care. This theoretical advantage relies on a host of assumptions, including the ability to assess quality accurately, the linkage of P4P performance systems with appropriate quality measures, robust information systems and the absence of adverse consequences (Witter et al. 2012).

Clearly, incentives would be expected to operate differently at these different levels: incentives to individuals are likely to be more directly motivating (incentives to organisations only affect behaviour indirectly, if passed on in some way to individuals), but may undermine cooperation (unlike organisational incentives, which might be expected to reinforce cooperation).

P4P relies on the power of extrinsic motivation. However, there is a substantial literature which emphasises other factors which motivate health professionals, including professional and social status and altruism (see, for example, Deci et al. 1999). Moreover there may be other barriers to changing professional behaviour, even when professionals are motivated, including patient factors, lack of time, lack of technical skills, lack of resources, and organisational constraints (Witter et al. 2012).

There is also the risk that financial incentives may dilute professionals' intrinsic motivation and this is the subject of widespread debate around public sector motivation in higher income countries (Marquand 2004, Myers 2008). On the other hand, where health workers' pay is low in absolute terms, incentives may be an important channel to improve motivation through increasing their income levels. The balance of effects is likely to depend on design and implementation, which could reinforce or undermine motivation.

The timescale of evaluation is another important consideration. Financial incentives might be effective in the short run for simple and distinct, well-defined behavioural goals, but these are not necessarily sustained in the longer run.

The potential failure to sustain P4P schemes also offers the significant risk of demotivating the workforce. Loss aversion suggests that the demotivating effects of reduced or discontinued payments might be greater than the original motivation of increases.

Other risks which have been identified (Oxman & Fretheim 2008; Witter et al. 2012) include:

- distortion – ignoring important services which are not rewarded with incentives, including untargeted services and more complex dimensions of performance
- cherry-picking – focussing on services which provide greatest gain with least effort over others which may be as or more important; this may also increase inequity in some cases (where poorer populations are harder to reach, for example)
- gaming - improving or cheating on reporting rather than improving performance
- financial dependency
- inefficiency – high operating costs and low returns may make this a poor investment
- fiduciary risks – if funds are poorly controlled, then leakage and corruption may be increased
- coercion – if clients are pressured to accept a service or undergo a procedure to increase provider rewards (e.g. for family planning uptake)

Scale and scope of PBF for sexual & reproductive health

There is now nearly 20 years of accumulated experience of P4P in the health sector of LMICs. A recent report states that over 20 countries have introduced PBF (Meessen et al. 2011), but this is almost certainly out of date, as a number of new schemes have begun since then. Many of these countries, such as India, Afghanistan and Rwanda, have employed PBF specifically with the aim of improving maternal and child health outcomes, the two most lagging areas within the MDGs globally.

The schemes commonly cover a range of services, including primary care provision, family planning services, and maternal and neonatal care. In terms of levels of payment, they are most commonly provided to facilities or individual providers. A brief description is given here of some of the better documented programmes which support SRH services, gathered according to the main level at which payments are made (see also the summary in Table 1). Many programmes provide a mix (e.g. payments to district teams and to facilities). There are many more as yet undocumented schemes.

Payments to governments

Global health initiatives, such as the Global Alliance for Vaccines and Immunizations (GAVI) and the Global Fund to fight against AIDS, Tuberculosis and Malaria (GFATM), use P4P mechanisms linked to contracts between the donor agency and national governments. They are best known for their contribution to immunisation and control of infectious diseases. However, more recent health system strengthening (HSS) grants may be linked to SHR services. For example, in Cambodia, GAVI HSS grants were given to 10 districts for improving maternal, reproductive and child health. Part of the grant was paid to health centres as a fee-for-service incentive to the team, paid for a limited subset of activities, including \$0.50 per consultation with children under 5, \$1 per ANC, PNC, Tetanus Typhoid, DPT-HB and measles immunization, iron/folate supplementation and birth-spacing visit (Hawkins 2011). There is no evaluation as yet of the impact of these grants, which also include unconditional elements related to, for example, remoteness.

Payments to mid-level structures

Payments to local governments

P4P arrangements between central and local (for example, provincial) governments are more common in Latin America. Argentina's Plan Nacer scheme focuses on uninsured pregnant women and children under the age of 6 through an incentive mechanism between the National Ministry of Health and the Provincial Government, and between the Provincial Government and health-care providers. The World Bank provided a loan that pays \$5 per person per month in eligible provinces. Provinces receive a capitation fee of \$10 per person per month. A first payment of 60% is made on the basis of the number of people enrolled, with the remaining 40% paid out in relation to meeting such targets as the number of women with a first antenatal visit before week 20, share of children who are vaccinated for measles, and the number of children born with healthy weight. The provinces write contracts with individual health providers, for the purchase of 72 services in all. Provinces determine their own fee schedules and administrative arrangements.

Brazil's Family Health Project, supported by the World Bank, also provides P4P-based transfers to local government. This makes per capita transfers to local municipalities on the basis of planned increases in certain services, such as safe delivery of babies for low-income women, monitoring of infants' nutritional status and growth, and treatment of poor children for various illnesses. If the municipalities reach these targets and several others, they will continue to be eligible for future financial transfers; otherwise, the level of central government support will be reduced (Hecht et al. 2004).

Payments to NGOs & contractors

Performance-based contracting through NGOs for service delivery is often favoured in post-conflict setting. One of the earliest examples was in Haiti, where, since 1999, USAID has funded local NGOs with a P4P-component to its payment mechanism. Since 2005, NGOs received 94 percent of the estimated budget needed to deliver a defined package of services to a catchment population in quarterly payments. In addition, NGOs could earn the 6 percent "withheld" plus another 6 percent if they achieved predetermined performance targets. Indicators focussed on child health (e.g. fully immunized children under one) and maternal health (proportion of pregnant women receiving at least four prenatal care visits and proportion of women with institutional deliveries). NGO-reported results are validated through random administrative audits at the facility level and random household visits to verify that services were received.

In Afghanistan, NGOs were contracted to deliver a basic package of services, funded by a variety of donors (mainly the World Bank, USAID and the EC). The modalities differed between donors. For the USAID project, MSH specified targets to implementing NGOs, with the sanction of non-payment if targets were not met. For the World Bank, bonuses of 10% were offered if targets were met (Sondorp et al. 2009).

Cambodia was another early example of contracting of services to NGOs, with a performance-based element. Two models - contracting in and contracting out of services - were tested. Targets for service coverage were set for the NGOs. At health worker level, staff in contracted in districts received pay which was partly performance related, while staff in contracted out districts received fixed higher pay (Bloom et al. 2007).

The World Bank project in DRC, now covering 89 health zones with total population coverage of 10 million, has adopted a performance-based contract on two levels (NGO and health worker). A total

of 10% of the project budget is earmarked for incentives, equivalent to \$0.40 per capita in the recent phase, 2007-10 (Johannes et al. 2008).

Many P4P schemes are in the early stages of implementation. In South Sudan, for example, a portion of the funding from donors to lead agencies who are implementing the Umbrella Programme for Health Sector Development since 2009 is based on hitting approximate targets for 13 indicators. These include inputs, processes and outputs (such as increased coverage of FP, ANC and health facility deliveries) (Morgan 2011). In Liberia, USAID began supporting contracts with NGOs that pay partly based on results in 2009.

An example from a more stable context is the Innovations in Family Planning project, funded by USAID since 1994 in Uttar Pradesh state, India (Rowan 2009). This channels funding to a local NGO to develop plans to increase FP access, use, and quality. A P4P mechanism is used to fund the NGO, SIFPSA, which in turn provides cost-based reimbursements of activities by implementers (NGOs and public departments). The performance indicators are largely process-related, with funds being paid out once activities are completed.

Output-based aid projects, like the Kenya and Uganda voucher programmes for reproductive health care, may also channel international funds direct to non-governmental organisations. Although these programmes offer demand-side finance (DSF), the funding modality fits with P4P in that payments from the donors are made according to the number of services delivered by the organisation managing the voucher schemes (see Witter 2013 for further discussion of DSF experiences).

Payments to facilities

In Rwanda, a number of donor-supported PBF pilot projects in different provinces were used to inform a national scheme. In 2005, the government decided to introduce incentives as a supplement to input-based budgets at primary health care centres. Bonuses were established for 14 maternal and child healthcare output indicators (e.g. children who completed vaccinations on time, women who received appropriate tetanus vaccines during prenatal care) and 10 clinical services and care indicators related to HIV. The bonuses were adjusted in proportion to each facility's progress on structural and process indicators of health care quality. Facilities reported their monthly indicators to steering committees that were responsible for authorizing payment. The reports were verified by auditors who would control the monthly invoices at the health centre level. In addition, on a quarterly basis, a different team would visit each health facility to evaluate their health care quality indicators. Payments went directly to facilities, which had full discretion in their use. Of 80 facilities surveyed in 2006-2008, the payments represented an average 22 percent increase in funds above the regular input-based budget, 77 percent of which was used to increase take-home pay for staff (Basinga et al. 2010).

The central Africa region hosts a number of PBF schemes which use a similar design – not only Rwanda, which is the best documented and known, but also Burundi and the Democratic Republic of Congo (DRC). In these cases, these policies began with pilot schemes managed by international NGOs, which were later scaled up, at least in Rwanda and Burundi, with continued technical and financial support. In Burundi, health facilities receive payments for delivering a list of priority services (maternal and child health, family planning, TB, and HIV). Scores on quality assessments provide potential increases of up to 15 percent of total fees received (Busogoro & Beith 2010).

In other countries, smaller scale NGO projects have been piloted (e.g. CORDAID-supported programmes in Zambia and Tanzania, working with Catholic dioceses and facilities) (Toonen et al.

2009). A pilot programme with a quasi-experimental design was implemented in Uganda in 2003-6 but not rolled out. 118 health facilities – largely mission private not-for-profit (PNFP) facilities – were randomised to a control group, a group without P4P but with increased autonomy in its use of existing resources, and a P4P-group, which was able to gain an additional 11% of its block grant for meeting three of the six performance targets. The targets were increased numbers of ANC visits, supervised deliveries, uptake of family planning, children immunised, malaria treatment for children and outpatient visits (Morgan 2010).

Many performance-based incentive programmes have just recently started – for example, in 2011 primary level facilities and staff in Sierra Leone started to receive performance-related incentives. As in Burundi, these supply-side measures were designed to complement a policy of providing free care for mothers and under-five year olds. Primary Health Units were to receive P4P funds every quarter based on delivery of six key reproductive and child health interventions, including ANC, FP, supervised deliveries and PNC (Amara 2011). Fixed amounts were to be paid per targeted output. The P4P funds were to be divided between incentives for staff and investment or operational costs for the facility. District health teams and Local Councils would supervise and verify service delivery and use of the P4P funds.

Payments to health workers

In a number of countries, P4P payments targeted at health workers and community health workers have been wrapped up in programmes offering demand-side incentives to increase facility-based deliveries. The largest example is India, where the Janani Suraksha Yojana (JSY) aims at reducing maternal and child mortality through both demand and supply side incentives. The JSY is a national programme that is administered by each state independently. The programme is centrally funded from the federal budget and various donors. Supply side incentives in the form of cash payments are provided to female health workers, the Accredited Social Health Activists (ASHA). The ASHAs act as bridge between the public health system and communities, and are responsible for registering pregnant women for facility-based births, arranging for in-facility deliveries, accompanying pregnant women to facilities and staying with them during their deliveries. These health workers are also responsible for follow-up visits and ensuring that newborns are duly immunized. Payments are received on verification of service provision, with the main trigger being an in-facility delivery (Dagur et al. 2010; Lim et al. 2010; UNFPA 2009).

A similar policy has been implemented in Nepal, with the Safe Delivery Incentive Programme (SDIP), which combines fee exemption with cash transfer to women and a flat-rate payment to health staff carrying out deliveries (both in facilities and at home). This was introduced in 2005 and continues to be operated as part of the 2009 Aama programme (Witter et al. 2011b).

Some health worker incentive schemes are operated at small scale as part of NGO-run projects. This was the case in Bhattagram, Pakistan, where Save the Children offered performance-based incentives to government health staff as part of its district health services reconstruction project (Witter et al. 2011a). In this case, performance was measured monthly at facility level (based on 27 supervision targets and 8 performance targets), but paid directly into health worker bank accounts.

In other cases, policies cover several provinces. An example is the Philippines, where women health teams, which include a midwife, “barangay” (village) health worker, and a traditional birth attendant, receive payment for every poor mother referred and for women who deliver in a health institution in several provinces (Gonzales et al. 2009). These performance-related grants were introduced in 2008 as part of an integrated Maternal Neonatal and Child Health strategy. The grants

to local government units covered provision of family planning for poor women, in the first case, and then support for facility deliveries (including cash payments to women, to traditional birth attendants for referring women in to facilities for their delivery, to the health staff and health teams). \$22 is paid to the team per facility delivery. In some provinces this includes the \$11 transfer to the mother, while in others this is paid separately. The programme is funded by the World Bank in two provinces.

In other cases, health worker incentive schemes are operated nationwide. In Cambodia, for example, the government has, since 2007, paid midwives a \$15 allowance per live birth in a health centre or health post, or a \$10 allowance per delivery in a regional or national hospital. Funds are allocated for this incentive in the MoH budget. A study of MCH incentives in one province found that many health centres were paying a share of the midwifery incentive to the Village Health Support Group (\$2.40) or Traditional Birth Attendant (\$1.20) when they referred pregnant women to the HC for delivery. (An MoH circular encourages this.) Some HCs instead used part of the incentive to pay for a gift given to the mothers to reward them for coming to the HC for delivery. The study also reports that in this province at least, the incentive (along with other forms of performance based payment) was usually shared equally among staff, although the incentive is paid to individual midwives (Murakami 2009, cited in Hawkins 2011).

In Tanzania, the Government of Norway agreed to contribute US\$32 million over five years to reduce maternal and child mortality, with pay for performance as one of the strategies to be used. This led to the introduction of a pay for performance scheme in 2008. The scheme consisted of bonuses to public and mission health facilities that attained performance targets related to maternal and newborn health and timely and accurate data collection (Morgan & Eichler 2009). The scheme was rolled out nationwide without piloting. However, donors declined to fund bonuses, due to differences of views on the process and design. The bonus payments were made to facilities, and district and regional health teams, to be shared equally amongst employees. As rules on awarding bonuses were not clear, donors felt that it would be used as a universal salary top-up (Morgan & Eichler 2009).

How policies were introduced

International donors have played an important role in promoting and funding P4P and demand-side pay for performance programmes (USAID 2010). The World Bank supports P4P programmes in a number of countries, including pilots in eight countries that will include impact evaluations funded by a trust fund. Norway has bilateral arrangements in a number of countries (India, Nigeria, Malawi, Tanzania, Pakistan), in addition to the funds provided through the World Bank Health Results Trust Fund. The Department for International Development (DFID) also contributes to the Trust Fund and is considering a broader results-based financing strategy. KfW (German Development Bank) is supporting output-based aid (vouchers) and is beginning to support broader P4P programs with both supply- and demand-side incentives. Belgian Technical Cooperation has supported P4P in Burundi, Rwanda, and the DRC. AusAID has supported development of P4P designs in a number of Asian countries. Currently, the GAVI Alliance (GAVI), the Global Fund, and the World Bank are working on operationalizing a joint platform for health systems strengthening that intends to incorporate P4P. The Inter-American Development Bank (IDB) and the World Bank have supported conditional cash transfers in a number of countries, primarily in the Latin America and the Caribbean region.

The role of international NGOs has been significant in piloting and managing schemes in many countries, at least in the initial phase. In Haiti, performance-based contracting was introduced by USAID in the late 1990s. The programme was initially administered in country with support from

Management Sciences for Health (MSH), a US-based international NGO. MSH administered the programme through 3 NGOs on a limited scale at first, and later expanded to 25 NGOs, covering 2.7 million people, with USAID support.

In Africa, P4P schemes in Tanzania, Zambia, Burundi, DRC, and Rwanda have been pioneered by international NGOs such as Cordaid and Healthnet TPO (Vergeer & Collins 2008; Canavan & Swai 2008; Vinard 2011). In some cases these were later scaled up, with government and donor support (e.g. in Burundi and Rwanda).

The process of developing the P4P schemes is not detailed for many of these schemes, but it is likely that a wider degree of participation in the process of development is one of the factors explaining the more successful schemes. For example, the Zambia and Tanzania pilots were agreed at diocesan level, with minimal involvement of government representatives or indeed facility managers, which is thought to be one of the factors behind their disappointing results (Toonen et al. 2009). Similarly, in Tanzania, the level of donor-government consensus on the design was limited (Morgan 2010). In a number of cases, particularly in relation to performance-based contracting, the main funders and contract managers have been external, with government playing a minimal role.

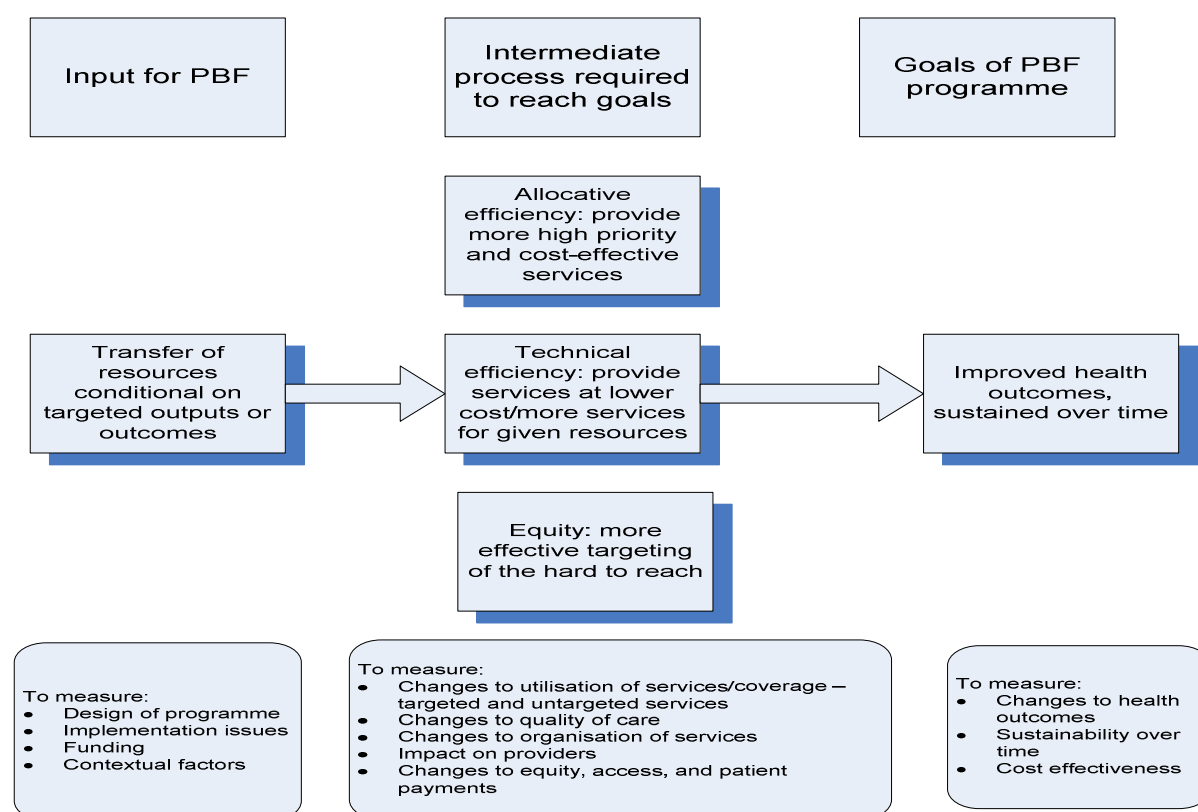
A clear problem analysis is recommended before introducing P4P (Eichler and De 2008) but it is not clear whether all schemes started from a clear understanding of the main blockages and of why P4P might be the best solution to them. A more recent manual emphasises that P4P is not a model but an approach, and one which should start from a bottom-up action-research approach involving all key actors (Toonen and van der Wal, 2012).

Evidence to date of impact

Framework for assessment

Based on the objectives laid out above for P4P programmes, a conceptual framework for assessing them has been developed (Figure 1). It follows the logic of inputs, intermediate goals and ultimate goals and defines some of the areas of enquiry in assessing the effectiveness and impact of a P4P programme. The paper looks for evidence on these different nodes, starting with impact and then moving to observations on preconditions for success (which link to the input elements in the figure).

Figure 1 Conceptual framework for assessing impact of P4P programmes



Quality of current evidence

Before examining the current findings on impact of P4P schemes, it is important to assess the overall strength of the evidence base. As seen above, P4P schemes are wide-ranging in approach, and many are recent. Much of their documentation is internal. Of the 100 or so documents reviewed for this paper, only a few were peer-reviewed articles. Many are descriptive briefs or reports by project funders and implementers.

A recent systematic review of P4P in low and middle-income settings concluded that ‘overall, the quality of evidence is graded as low or very low, with limited numbers of studies reporting on specific indicators, high risk of bias in most studies, and inconsistency of findings. We conclude that there are few robust studies of PBF available from a low- or middle-income context and it is premature to draw any firm conclusions on its effectiveness or factors that determine its effectiveness’ (Witter et al. 2012). Only one study was assessed as having low risk of bias (Peabody et al. 2011), which is not included in this review as it focussed on curative child health services.

Problems which were common amongst the studies identified (Witter et al. 2012) included:

- Non-random allocation of the intervention
- Additional resources and ancillary components that may be responsible for impacts rather than conditional payments
- Other confounders (e.g. contextual differences between intervention and non-intervention groups)
- Lack of rigorous before and after measures of effect
- Lack of consideration of wider systemic issues (e.g. adverse impacts on other services)

- A plethora of targets (outcome measures) and consequently a high risk of selective outcome reporting (i.e. reporting statistically significant results and not reporting results that are not statistically significant)
- Conflicting interests (due to P4P being evaluated by individuals and organisations that are advocating and implementing it)
- Evidence of publication bias (with negative findings being less likely to be published than positive ones)
- Not measuring health outcomes, which means that the relationship between process measures and health benefits is uncertain

These concerns are shared by other reviewers (Eldridge & Palmer 2009; Oxman & Fretheim 2008) and by the World Bank, which acknowledges that there is a lack of strong evidence base, despite nearly 20 years of involvement in P4P programmes (Brenzel et al. 2008). This has led to new focus on generating robust evidence, supported by a toolkit for conducting impact evaluations (Vermeersch et al. 2012). There are currently 16 impact evaluations underway, funded by the Health Results Innovation Trust Fund, with others in preparation, testing a variety of designs².

In addition to concerns about the internal validity of studies, there are concerns about the narrow frame for evaluating P4P programmes to date. There have been calls in recent years for a broader health system framework for monitoring and evaluating PBF, rather than focussing on targeted output indicators, as is commonly the case in current studies (Witter et al. 2013), for a greater focus on understanding PBF processes and mechanisms (Witter et al. 2013, Sengooba et al. 2011, Macq & Chiem 2011). Others also point to the need for broader assessment of P4P programmes against the DAC criteria (Perrin 2013).

Utilisation of services/coverage of programmes

Most schemes identified (see Table 1) target a common set of services, including antenatal care, immunisation, family planning, assisted deliveries, postnatal care and, in some cases, child nutrition and monitoring. Some also fund general outpatient care (e.g. DRC and Burundi) and indicators linked to inpatient care (e.g. inpatient turnover in Zambia). As the aim of the P4P programme is to increase coverage and utilisation of these services, it is not surprising that this is the output/outcome indicator most commonly reported in studies (see Table 2).

Nevertheless, of the 25 schemes and studies identified, only seven included information on utilisation which could be linked in any way to the P4P intervention (some because of demand-side interventions, others because of lack of controls of any description).

The most robust is the Rwanda study, which provided equivalent additional resources to control facilities. It found (Basinga et al. 2010):

- No impact of PBF on the probability of any prenatal care or on the probability of completing four or more visits
- A statistically significant impact on the probability of institutional delivery (7% absolute increase, rising from 35% before to 42% after)
- For children, a significant increase in the likelihood of a preventive visit in the four weeks prior to the survey, but no impact on the likelihood of full vaccination

² See <http://www.rbfhealth.org/project/our-projects> for latest details.

- Related analysis of payments for HIV testing found an increase of 6.1% in the probability of individuals having ever been tested

Results from other studies have to be interpreted with particular caution, given the risk of bias issues highlighted above. Broadly, the Tanzania and Zambia studies found that performance of the intervention (mission) facilities was similar to or worse than the 'control' government ones for the indicators tracked (outpatients visits, antenatal care, voluntary counselling and testing, in-patients and institutional deliveries).

Outputs for which relative risks could be calculated, based on the original data, were assessed in the systematic review (Witter et al. 2012). In Burundi, a statistically significant difference was found for institutional deliveries, favouring the intervention sites (RR: 1.79), but in the DRC, the reverse was found (statistically significant difference, but RR of 0.75). Coverage of bed nets was statistically significantly higher in Burundi (RR: 1.9). In Tanzania, inpatient admissions were significantly lower in intervention sites (RR: 0.82). In Burundi, pregnant women were statistically significantly more likely to be fully vaccinated (RR: 1.13). For all other indicators, no statistically significant difference was found.

The performance-based contracting programme in Haiti showed success in some areas, with immunization coverage increasing between 13 and 24 percentage points, and increased births by skilled attendants from 17 to 27 percentage points (Eichler et al. 2007). However, prenatal and postnatal care did not respond significantly, which is attributed to a 'strong patient behavioural element'. Later analysis of this scheme found that a 39% increase in primary health care services could be attributed to the incentives element, and that there were no unintended effects on unrewarded services (Zeng et al. 2012). It should be noted however that the scheme had an unusual design, in which incentives (averaging 6% of funding to the NGOs) was awarded annually for services which were identified after the year-end (to avoid gaming and neglect of untargeted services).

The Uganda pilot found that the PBF intervention group (despite winning bonuses in relation to targets) performed less well than the group which simply received more autonomy, and about as well as the control group (Morgan 2010).

Overall, then, the findings on utilisation of care are mixed. There is no pattern across schemes, in terms of which indicators responded positively, or the magnitude of response. This is likely to be linked to the different designs of schemes, the different payment amounts and systems, and the different starting points and degree of effort involved in changing indicators.

Quality of care

Very few studies provided details on independently assessed quality of care, although six out of 25 provided some measure of quality of care, including process indicators and patient and staff perceptions. Again, overall findings are mixed.

In Rwanda, quality of prenatal care was assessed by comparing activities undertaken during prenatal visits with the local clinical practice guideline and by investigating whether a tetanus typhoid vaccination was given during prenatal check-ups. Significant improvements in both measures are reported for the intervention group (Basinga et al. 2010). In addition to these measures, the payment of incentives was linked to a composite quality measure, based on quarterly direct observation by district supervisors and medical records review. The scores for this broader quality assessment are not reported. However, a later analysis found that incentives reduced the gap between provider knowledge and practice of appropriate clinical procedures by 20 percent, and that

there was a strong complementarity between performance incentives and provider skill (Gertler and Vermeersch, 2012).

Other studies provide anecdotal information of unintended negative consequences for quality of care. In one study (with a very limited sample), staff reported neglecting essential but untargeted duties, such as potentially life-preserving activities in the intensive care unit of hospitals, and counter-productive behaviour such as not distributing the last drug box of the pharmacy to avoid a stock-out (Kalk et al. 2010). There is also anecdotal evidence of similar distortions and possible perverse effects from a number of other studies (Macro International 2009; Eichler 2006).

In Cambodia, improvements in process indicators such as staff availability are reported in contracted areas, but also negative perceptions by patients (Bloom et al. 2007).

In Uganda, exit polls showed that the perceived availability of medicines, attitude of staff, and the prices charged by the facility worsened in the view of the respondents in intervention areas (Morgan 2010).

Patient-assessed quality was found to be higher in PBF districts for most indicators (before and after the intervention) in the DRC project (Soeters and Kiwanuka 2008). Significant improvement was found in patient assessments of quality. However, managers were found to be dissatisfied in most facilities, particularly with the level of external support.

In Burundi, there was an improvement in the PBF provinces for quality of care, as assessed by health professionals in hospitals and health centres. Quality of care as perceived by households declined for both groups (PBF and controls), with no significant difference between the two (Soeters and Kiwanuka 2009).

Organisation of care

One of the claims for P4P is that it can allow services to be organized in a more efficient way. It is therefore interesting to examine the findings on this from the studies which were identified, but here the evidence is even more limited. Most studies do not report impact in terms of the range of services offered or how they are organized. This was also reported by the systematic review, which found that 'none reported on changes to organisation or delivery of services, on impacts for management and information systems or on wider impacts for financing or resource allocation, which is surprising, given the nature of the intervention' (Witter et al. 2012).

In some cases, such as the Global Fund, additional resources have allowed for the expansion of services, though these cannot be attributed to the P4P mechanism as such. In other cases, capacity building of NGOs was undertaken as an ancillary measure alongside the performance-based contracting (e.g. in Burundi). In Cambodia, better availability of services was reported in contracted-in districts, but again the study does not support attribution of this to P4P.

Although the impact of introducing P4P on organizational features may be hard to isolate, it appears from some studies that greater autonomy at facility level may in itself produce benefits (e.g. in Uganda, where facilities with autonomy out-performed the bonus and control groups).

Provider motivation and behaviour

Given that one of the main purposes of P4P is to motivate providers to deliver services in line with public priorities, it is surprisingly how little has been documented in terms of its impact on provider

motivation and behavior (including changes to working hours, absenteeism, dual practice, retention in rural areas, outmigration and informal charging of patients).

In terms of staff satisfaction, it is to be expected that staff are pleased to receive what is, in almost all cases, additional funding, and indeed a number of reports do find this. However, responses are often nuanced, in that the funding comes with increased pressures and studies do not always differentiate responses to levels of pay with responses to the system for allocating it. In Rwanda some healthcare workers see it as another control mechanism (Kalk et al. 2010). They also complained about inadequate conditions to fulfil the targets, the time taken to comply with new paperwork, and potential damage to the provider-patient relationship. In Haiti, some of the feedback received pertained to the negative incentives embedded in the P4P programme. Participants complained about too much "stress" over meeting institutional targets, mainly from having to make organizational changes (Eichler et al. 2007). In the Nepal Safe Delivery Incentive Programme, the cash payment to staff was reported to have caused conflict and tension (Powell-Jackson et al. 2008). In a district-based scheme in Pakistan, staff were not involved in setting targets, were not well informed of how their performance was assessed and were unhappy about the different systems operating for different types of staff (Witter et al. 2011a).

These pressures can be constructive, or destructive (e.g. if targets are not met due to external constraints). Involvement in setting targets, ability to control the factors which affect those targets, perceived fair processes and transparency in measuring and rewarding them, and an adequate level of funding of targeted actions are all factors which are likely to improve staff responses.

As P4P funding is not in most cases linked to individual performance, and is often paid in contexts where salaries are very low, it is in a number of cases perceived as a simple salary top-up scheme (Toonen et al. 2009; Hawkins 2011; Witter et al. 2011a).

Equity, access and patient payments for care

Reducing patient payments for care is not a primary objective of P4P programmes. However, some authors argue that these payments will allow providers to reduce their charges and will incentivize them to do so, especially in programmes which pay per unit of activity (e.g. a fixed payment per supervised delivery) (Soeters & Kiwanuka 2009). Conversely, there may be a risk with P4P of over-consultation by the middle-classes and a failure to reach the poor. This was found in the pilot scheme in Uganda, where the wealth index of clients treated by the PNFP bonus group increased relative to that of the PNFPs in the control group (Morgan 2010). However, in general, little robust monitoring of equity effects has been undertaken.

Only one study identified looked specifically at equity, by disaggregating changes to household payments in the PBF areas (Soeters and Kiwanuka 2008). Household payments as a proportion of income by the poorest were found to reduce more in intervention districts (by 63.5%, compared to a 21.9% reduction in controls), though it should be noted that payments in PBF areas were at much higher level to start with. For the poor (second quartile), the reverse was found, with a 76.5% reduction in the controls, compared to a 36.2% reduction in intervention district households.

In terms of overall patient payments, in the Democratic Republic of Congo, payments by patients were reported to increase in the intervention group (Soeters and Kiwanuka 2008), while in Burundi, they were reported to decrease (Soeters and Kiwanuka 2009). These differences are most likely explained by the differences in starting levels and also the heterogeneity not only of the intervention but also the support which was provided to the control areas (Witter et al. 2012). In Cambodia, contracting out had a negative effect on out-of-pocket health spending (presumed due to switching

from informal and private providers), while contracting in had no statistically significant effect (Bloom et al. 2007).

In some cases (e.g. Burundi and Rwanda), the PBF scheme has been introduced alongside reduction in financial barriers for users (though fee exemptions or reductions, and demand side payments), which complicates attribution of any distributional changes to PBF.

In relation to geographic equity, there is a risk that areas with greater challenges are progressively marginalized, if it is harder for them to reach targets and so to receive funding. Some form of additional support may be needed. Again, there is very little research into the redistributive effects of P4P across facilities/areas. However, some schemes have adjusted payments to favour hard-to-reach areas – for example, by increasing capitation payments for more remote provinces by 15% in DRC and up to 40% in Burundi.

In relation to results-based aid at the national level, a review found that in some cases poorer countries find it hard to access funding in the first place. This is either because the application process is complex or because a prior performance record is required (Pearson 2010). They recommend that equity is emphasised at all stages of the identification and implementation of RBA/RBF schemes, including through the use of locally identified targets in low income countries (which might imply lower but still challenging targets as opposed to the use of global targets or standards), up front capacity building efforts and technical support, and different (simpler) approval processes.

Unintended consequences

Some authors make observations about side-effects (for example, the increased demand for technical advice at provincial level in Argentina following the Plan Nacer), but this does not provide the same quality of evidence as a study design which proactively seeks to monitor knock-on effects. Only three studies actively looked for unintended effects. In both Zambia and Tanzania there was a concern about the curative nature of the coverage targets and whether this may squeeze out preventive care. However, no conclusive evidence was found to support or refute this concern (Vergeer & Collins 2008; Canavan and Swai 2008). In Cambodia, some negative effects on untargeted services were identified for contracted out districts and positive effects in contracted in districts, but neither were statistically significant (Bloom et al. 2007).

Although the risk of unintended consequences is highlighted for PBF in particular by many authors (Eichler & De 2008, for example), most study designs focus on measuring targeted indicators alone. The opportunity to investigate knock-on effects on other important but untargeted services and indeed on the health system as a whole has not yet been seized.

The main area in which systemic effects have been studied to date is for results-based aid and the role of the global health initiatives. One recent study concluded that ‘many of the schemes reviewed run counter to at least some of the other principles of aid effectiveness (notably alignment with country systems and country ownership). This is partly a feature of the institutions which have taken RBA/RBF forward (e.g. GFATM, GAVI which have a disease specific, sub sectoral focus). Those that bypass government are not aligned, those that involve government may simply be an additional layer of donor interface which add little value. Many of the schemes are narrow which reduces the scope for strengthening the system as a whole and creates risks that they will further fragment the sectors they operate in’ (Pearson 2010).

Health outcomes

The global health initiatives, such as the Global Fund, have estimates of health outcomes, but these cannot be attributed to the performance-based payment mechanism as such. A similar problem arises with the contracting pilot in Cambodia (Bloom et al. 2007). There is as yet no robust evidence of a link between P4P and health outcomes for SRH services, though there is a small body of evidence linking P4P with child health benefits (Peabody 2011, Gertler and Vermeersch 2012).

Costs and cost effectiveness

Magnitude of payments

Overall expenditure on projects, where available, is reported in Table 3.

Per capita costs are reported for a number of the projects, although these obviously cover varying packages of care:

- \$10 per capita per month for the Plan Nacer in Argentina
- In Afghanistan, the contracts ranged from \$3.8 (World Bank) to \$5.22 per capita (EC) per year
- \$2.56 per capita per year in contracting in districts in Cambodia; \$2.94 per capita in contracting out districts (61% and 85% higher respectively than comparison districts)
- \$2 per person per year, plus \$0.4 for administration and overheads, in the DRC CORDAID pilot (and the same for Burundi, but with \$0.6 added for management)
- Euros 0.5 per capita per year in the Zambia and Tanzania CORDAID pilots
- \$0.25 per capita for health worker incentives alone in the World Bank DRC programme (funding service delivery through NGOs)

How these global amounts were established is not clear. Nor is the system for establishing payments per item clear – reports generally provide no information, or state that they were based on discussions with stakeholders.

Per item payments are reported in some cases, where providers are reimbursed per service. For example, in Rwanda, the tariff was \$4.59 per delivery; \$0.18 paid per child preventive visit; \$1.83 for referral of malnourished child; \$0.92 for TT and malaria prophylaxis during ANC; and \$0.09 per ANC visit (\$0.37 for all four visits). In Burundi, payments ranged from \$0.25 per new OPD case to \$10 for TB diagnosis. In the DRC, payments for health centres ranged from \$0.30 per bed-day to \$20 per TB patient successfully treated. Again, how these payments were established is not clear.

Relative size of payments

The relative magnitude of payments (relative to facility revenues and relative to staff salaries) might be expected to reveal something about the power of PBF as a lever. This is not reported in all cases. Where reported, the scale is quite varied:

- In Tanzania, they amounted to 8% of facility income on average (Canavan & Swai 2008).
- In Zambia, authors report that the PBF payments amounted to 17% of facility revenue and were small (but variable) relative to salaries (Vergeer & Collins 2008).
- In the DRC pilot project, the proportion is not stated, but the incentives must have been the major component of funding for the health centres at least, as their overall revenues per person rose from \$0.51 to \$1.04 over the period (Soeters and Kiwanuka 2008).

- In Burundi, the intervention facilities surveyed received 58% of their total revenue from the PBF scheme in 2008 (Soeters and Kiwanuka, 2009).
- In Rwanda, facility funding increased by 22% for PBF group as a whole. On average facilities allocated 77% of the PBF funds to increase personnel compensation, amounting to a 38% increase in staff salaries (Basinga et al. 2010)

Is there evidence on the optimal size of bonus relative to contract size (from the performance-based contracting examples)? In Uganda, the failure to produce significant benefits of the PBF arm was blamed on the size of the bonuses. The maximum performance bonus a facility could receive was 11 percent of its base grant, or roughly between 5 to 7 percent of its total operating revenue. On average most health facilities received bonus payments of less than US\$1,000 per year (Morgan 2010). In Afghanistan, NGOs contracted by three donors could earn a bonus worth 10 percent of the World Bank contract value if they reached or exceeded targets outlined in the contract. In Haiti, NGOs could earn the final 5 percent of their fixed quarterly payment plus an additional 5 percent if all performance targets were achieved. In 2005, this was increased, from 10 to 12 percent. All three cases offered a similar proportion, but with different reported effectiveness (and other important differences, such as different contract values and different underlying financing systems).

In schemes which pay direct to health workers, again, it is very hard to formulate any rules of thumb on the optimal proportion of pay to be derived from P4P payments, given the lack of information and the very varied circumstances of health worker starting position and market opportunities. In a district-based scheme in Pakistan, staff received an average top-up of 29% of pay (\$48 per person per month), but this was felt to be inadequate in relation to the opportunity costs of private practice foregone (Witter et al. 2011a).

The underlying resourcing contexts and health financing systems varied considerably and were not consistently reported. This makes it hard to determine the degree of change which could be plausibly be expected from the additional resources which P4P brought to bear.

Transaction costs

Concerns have been raised about the cost of administering what is a highly labour intensive approach to boosting health service delivery. Donors advocate that administrative costs should be kept within a ceiling (ideally less than 25% of budget costs) (Canavan et al. 2008). This was the level of overheads for the Rwanda programme (25% of total budget). A multi-country study found costs in the range of 15 to 30% of overall expenditure (Toonen et al. 2009). Experience in Rwanda as in other countries shows that fund-holder organizations require 4-7 qualified staff to manage a P4P project with a target population of 300,000 – 700,000 inhabitants (Canavan et al. 2008).

In relation to donor funding, no analysis was identified in relation to the impact of P4P on the transaction costs of aid. There was some expectation that P4P mechanisms – which in theory are supposed to be focused on outcomes, not activities – might help to harmonise aid, and to reduce reporting. However, no assessment of whether this has materialised was identified.

Cost-effectiveness

No studies reported on the cost-effectiveness of their intervention. This is clearly one of the most significant areas of gap in the literature. Governments have to choose within a wide range of options aimed at, for example, boosting their mother and child health indicators, including salary supplements, contracting and other performance management measures, P4P, and demand-side measures (reducing fees, increasing insurance coverage, vouchers, health equity funds etc.). It is

therefore urgent to understand their relative costs and benefits in different contexts and for different designs.

In relation to results-based aid modalities, a recent review concludes that 'value for money will need to be carefully monitored. Financial risk can be shifted to agents but this is likely to have cost implications and may potentially be at the expense of service delivery for the poorest (e.g. the most vulnerable countries are likely to respond worst to CODA)' (Pearson 2010).

Sustainability

The programme in Burundi, after its pilot phase with INGOs in 2006-9, has been scaled up nationwide (Vinard 2011). It is now funded through the Ministry of Health (it absorbs half of the recurrent budget) and by donors (who pay directly to health facilities according to MoH invoices).

In Rwanda, the story is similar – after a pilot phase with INGOs in 2002-6, it has been scaled up nationally. The scale-up was part of larger reforms by the Government towards a decentralized political and fiscal structure. As part of multi-sectoral agreements on outcomes with district mayors, signed in 2007, within the health sector performance incentives would be provided for uptake of health insurance, institutional deliveries, family planning, and use of insecticide-treated bed nets (Kalk et al. 2010; Sekabaraga et al. 2011). The programme is managed through a strong national coordinating body but also decentralised district-based systems (Vinard 2011). The government pays a substantial part of health budget through PBF, which is also externally funded. PBF mechanisms are now channelling more than 50% of the running costs of public facilities in Rwanda and Burundi (Vinard 2011). They remain highly donor-dependent however.

In the DRC, where P4P mechanisms were first used in 2004 in an EU project, there is as yet no decision on its wider adoption.

In Tanzania, despite strong government commitment for a national program using P4P for health workers, sustainability became a challenge early on because of poor planning and lack of support from the donors. The programme was implemented rapidly in 2008-9 but ran into implementation problems. Delays in payment to workers because of the refusal of donors to finance the programme through the donor-supported "health basket" also created problems. In 2010, the government of Tanzania was to enter new talks with the Government of Norway to re-launch the programme addressing its weaknesses (Morgan and Eichler 2009).

In Afghanistan there has been government commitment to contract NGOs to provide basic healthcare services and most services are delivered through this mechanism. The World Bank, USAID and the European Commission have signed performance-based contracts with NGOs. Since there is strong government and donor buy-in, the sustainability of this programme is likely (barring withdrawal of donor support or another civil war).

In smaller, externally funded and managed schemes, such as the district-based scheme in Pakistan, integration and financial sustainability by the local government is a major challenge. The PBI scheme in Battagram, for example, cost only \$0.68 per person per year, but this represented 44% of the annual health budget in the district (Witter et al. 2011a).

In short, the degree of longevity and integration vary – depending on a combination of government buy-in; coherence with other strategies; institutional, social-cultural and technical embeddedness; good working relations with donors; perceived and documented good results; and continued

external support. All on-going P4P programmes documented here continue to be highly dependent on external support.

Observations and lessons learned on design

Defining performance

Performance is clearly defined in a wide variety of ways in these examples:

- Performance in the Pakistan project meant, at an individual level, coming to work reliably.
- For some, such as the IFPS programme in India, it meant a range of activities, including training, planning etc.
- For the large majority of the programmes, it meant maximising specified service outputs (both for schemes aimed at individuals, such as the JSY in India and the SDIP in Nepal, but also those aimed at facilities).
- Some combined these with process indicators (drug supply in Zambia and Tanzania; management indicators in Nicaragua; quality specifications in all of the central African programmes).
- Only one (the Plan Nacer) included an indicator related to equity (coverage of indigenous population).

Despite the talk of focussing on outcomes, no programmes linked payment to outcomes, for the simple reason that these are not within the control of the provider. It is all the more important then that the outputs which are incentivised are evidence-based and of good quality (so that the expected link to health gains is realised).

Level of targeting of payments

The targets for P4P will depend on the objectives of the programme. It is not clear from the case studies presented here that payments to any particular level or target are more effective in principle. Some writers have suggested that payments direct to individuals will be more motivating. However, all payments to organisations have cascaded into some form of incentives for staff. Further, the general literature on health workers and optimal incentive packages suggests that a mix of incentives is likely to be most effective, including non-financial rewards (WHO 2010), which may be less likely to crowd out intrinsic motivation. Channelling resources to facilities – the dominant approach documented in the studies identified here – can allow for more flexible allocation of funds and is more likely to enhance cooperative behaviour.

In a number of cases, P4P funds were also paid to higher level management (district health offices, for example, for supervision) and to lower level community health workers. There is no clear evidence as yet about how effective these different strategies may be, and this is likely to be highly contextual in any case.

Size of payments

In general, the higher payment and the more that providers control the services, the greater would be the expected impact of PBF (Basinga et al. 2010). However, the effects of incentive size are likely to depend on a number of other factors too, including how incentives are used and shared with staff; starting levels of pay; and the costs relating to increasing outputs. Some conclude that small

portion can still motivate change, e.g. 10% of income (Eichler and De 2008). A study in the Philippines (focussed on curative care for children) found showed little difference in terms of quality between a group of physicians who were provided a 15% direct bonus and a second group who received a 5% indirect bonus where facilities were reimbursed against targets. On average, each treatment group saw a 10% increase in quality (Peabody et al. 2011). In Mali, a 40-50% addition to pay was judged necessary to motivate in one project, whereas in Ghana it was decided that a 15% increase on staff take-home revenues would motivate them – these differences being based on the lower level of basic pay in Mali (Toonen and Van der Wal, 2012). In Haiti, a bonus of 10% appears to have motivated providers, whereas in Argentina, a full 40% of payments were linked to performance.

How payments are calculated

It appears that P4P is being used in very different ways in different contexts – in some as a payment mechanism, linked to actual service costs, and in others as marginal top-up to motivate providers and to fund some recurrent costs at facility level. In most cases, facilities remain dependent for a substantial part of their funding on complementary sources ('input-based financing'), not just for investment but also running costs. In some schemes (e.g. Burundi) a mix of approaches has been taken, with some services more highly funded than others (e.g. HIV), which can lead to distortions. In Rwanda, hospitals are given a budget based on the unit costs of planned activities, with marginal top-ups related to quality, which is a more stable system. In DRC, payments were apparently originally related to costs, but these diminished as the budget came under strain, leading to an underfunding of vertical programmes such as TB (Vinard 2011).

There does not yet appear to be a consensus on how P4P payments should be calculated. The budget ceiling is as likely to dictate rates as evidence of how much is needed to motivate or fund change. It is a non-trivial exercise to be able to set payments at the right level to avoid excessive or deficient incentives, given that real costs will vary across actors and over time, and taking into account the often complex funding flows. As coverage changes, marginal costs to providers change, and so payments need to be adjusted, especially for P4P programmes which pay per service. Such systems will also need to build in cost control measures over time – experience in other settings indicates that they are likely to have inflationary tendencies, unless well managed.

Another consideration is transparency. Many systems involve complex weighting of indicators, which make final payments hard for actors to understand.

For payments made direct to staff, these can be made flat rate, per item or as a proportion of salaries. In the Pakistan PBI scheme, payment was made as a proportion of salary, which rewards top staff disproportionately. However, it appeared to be accepted as fair (maybe in part because the decision rule was at least clear) (Witter et al. 2011a).

Number of targets

Is there any learning from these examples about how many targets are optimal? The USAID P4P project in South Sudan had 50, which was seen as too great by project implementers. Rwanda paid for 14 output indicators, but also used a quality scoring system based on a further 13 services (incorporating structural and process dimensions). The trade-off between avoiding distortions (by including a wide range of services) and operating a manageable and comprehensible system (by keeping it simple) requires a difficult judgement.

Payment systems

A mix of payment systems are shown by the examples in this paper. Some are close to traditional contracts, with the main penalty being failure to renew if targets are not met (as was the case in Cambodia, for example). Others operate a fee for episode payment system (the most common approach). Others again offer small bonuses to be won (or lost) if targets are met, with the main payment fixed.

It is generally recognised that there is no perfect payment system, and this is borne out by the range of approaches tried by these different programmes. Payment by service rewards those who are able to operate efficiently (if payments are flat-rate, as in the central African schemes). Payment according to relative improvement rewards those who start with low coverage. Payment according to absolute thresholds (as used in Zambia and Tanzania) rewards high performers. Some of the ongoing impact evaluations are seeking to test different types of targets (see <http://www.rbfhealth.org/project/our-projects> for summary of impact evaluation designs).

Two types of design options for setting targets have been shown to produce disappointing results, according to Eichler and De (2008): (1) a uniform threshold applicable for all P4P participants (for example, everyone must reach 90 percent full immunization coverage) and (2) following a “tournament model,” where those in, say, the top 75th percentile of performance receive the bonus. They recommend setting targets according to the baseline position of each unit (proportionate increases in coverage, with diminishing proportions as coverage rises). This approach is clearly only applicable for preventive interventions and interventions where coverage can be measured.

In most cases, there will be a need for continuous revision of targets to avoid perverse effects. In the Haiti experience, the selection of performance indicators was revised in each phase.

In relation to frequency of payments, it is generally assumed that more frequent payments are better in terms of linking rewards with activities. Most programmes pay monthly or quarterly.

Verification

A critical element of P4P is accountability. Since payments are tied to specific targets being met, some form of independent monitoring and evaluation needs to be established. Many schemes try to engage civil society organisations in third party monitoring but this is not always easy to achieve (Vinard 2011). Ad hoc data collection remains necessary, although this is a costly process and does not necessarily contribute to strengthening the routine health information systems. In Afghanistan, for example, the monitoring and evaluation is contracted out to Johns Hopkins University and the Indian Institute of Health Management Research. The M&E is conducted through nationwide annual household surveys and semi-annual facility-based inspections (Sondorp et al. 2009).

Service or target type

P4P has tended to focus on infectious diseases and on reproductive and child health services, at least in low and middle income countries. As it typically incentivises greater activity, target indicators have focussed on preventive care and on the package of care which is believed to be linked to better outcomes. Quality of care has to be factored in to avoid perverse effects (iatrogenic infections etc.). Services which are easily defined and measured are obviously necessary, and should be prioritised based on the local burden of disease and the areas of underperformance.

In order to avoid undue focus on limited activities, payments which reward provision of a continuum of care are important, particularly in maternal health, where certain aspects such as family planning, post-natal care and neonatal care have tended to be neglected (Morgan et al. 2011).

Most of the programmes have focussed on the primary care sector, or on 'basic health services'. This relates to the desirable characteristics identified above. Paying individuals for performance in a hospital setting is particularly difficult, as quality of care depends crucially on team-work within and between teams. Performance of hospital staff or teams is complex and difficult to measure in a balanced way that captures all dimensions – such as the need for hospitals to continuously prioritise an unpredictable volume of urgent cases while maintaining progress on planned activities (Hawkins 2011).

Stand-alone or linked to demand-side measures

Most countries in these examples are experimenting with a range of supply- and demand-side reforms. These are necessarily very context-specific: no optimal package can be derived. However, a number of countries are combining incentives to providers with reduced access costs for clients (e.g. Burundi and Sierra Leone, which are both removing fees for under-fives and pregnant women, while also introducing P4P). As both stimulate utilisation, a substantial response can be anticipated and should be carefully managed to avoid capacity constraints, quality reduction, and budgetary short-falls.

Another risk, illustrated in countries like Cambodia, is that the plethora of financing mechanisms on supply- and demand-side can end up creating confusion about which cost elements are funded by which channels, what overall resources are being received at facility level, and what the main commissioning tools are (Hawkins 2011).

Purchasing arrangements

One rationale for PBF is that it can offer the opportunity to build a financing institution, independent from political risk, which can mobilize other local and external resources (including insurance premiums), and to adequately involve civil society (as social security systems do in most OECD countries). However, building up an independent purchaser is a challenging juridical exercise, requiring sensitive political compromise between stakeholders (Vinard 2011). In most cases described here, purchasing was either done by external organisations (donors and funders) or by the Ministry of Health. In Rwanda and Burundi, for example, there is no separation of regulator from purchaser, or indeed provider – a system which is more efficient, but leaves providers reliant on the Treasury for their cash flow.

Implementation lessons

It goes without saying that poor implementation will undermine the effectiveness of a P4P – as any other – programme. Failure to mobilise necessary support, delays in negotiating contracts, funding delays, lack of clarity on roles, all of these have been documented in some P4P programmes and have been linked to disappointing results (Morgan 2010, Morgan 2010a). Steps for designing and implementing P4P programmes have been described in a number of reports (for example, Eichler 2006, Toonen and Van der Wal 2012, Soeters 2013).

As with any reform, effective communication with stakeholders is a prerequisite for success. If reforms are too rapid, if stakeholders are not well informed, if details relating to implementation are

not provided, then implementation is likely to be poor. These problems were documented in Tanzania (Morgan and Eichler 2009) and also in DRC, where a study found no evidence of programme effectiveness, which was attributed to a number of factors, including the failure to boost staff income regularly, and the fact that health workers were not well informed about the programme and the performance criteria (Fox et al. 2013).

Application in different contexts

Some commentators have suggested that P4P has worked better in post-conflict settings (Toonen et al. 2009). If this is the case, a number of reasons could be hypothesized – that there is less inertia in the system; that providers have lost some of their intrinsic motivation and are therefore more amenable to financial incentives; that control mechanisms are weak and therefore need to be replaced by other levers; that central funding may have broken down in any case, leaving providers open to market failures etc. P4P has certainly been applied actively in fragile and post-conflict states, and that presumably relates, at least in part, to a vacuum in regular services and donors filling government functions, at least for a transitional period.

On the other hand, if the context is too hostile, then it is unreasonable to expect implementers to achieve significant increases in outputs. This was a point made by implementers in Southern Sudan, who faced penalties for not meeting targets, which they felt were unreasonable given the many challenges of the nascent health system (Morgan 2010a). Problems listed included lack of access by the population, shortages of staff, poor staff pay, and drugs supply problems. These were exacerbated by process issues, such as lack of consultation on targets, targets being measured too frequently, and issues of credibility of baseline data (Morgan 2010a).

A related point was made in relation to the Tanzania programme. ‘Performance-based funding at the district level will, for instance, not fully address the challenges related to poor transport infrastructure, delayed supplies of drugs and equipment from the central level, the shortage of health workers in general, and the shortage of people trained for emergency obstetric care in particular. Nor is there any reason to believe that performance-based funding will improve those aspects of the quality of the service that are difficult to observe for both clients and supervisors’ (Maestad 2007).

Organisational requirements

Quite serious reforms are required to make PBF work: ‘Usually, provider management and accounting systems need to be strengthened, purchasing capacity improved, performance and quality standards established, and adequate provider reporting and information systems introduced to allow for appropriate performance monitoring and transparency’ (Brenzel and Naimoli 2009).

Major areas for capacity building by NGOs within the pilot projects included: (i) strategic planning (ii) cost and revenue analysis (iii) determining client perceptions of quality of services (iv) models of staff organization and utilization and (v) information systems and human resource management (Canavan et al. 2008).

Key skills and tools include contracting (often at several layers); business planning (for providers); monitoring (for purchasers); and regulatory skills³. Practical issues such as whether facilities all have bank accounts are also important factors to take into account.

The link with decentralisation and community participation continues to be debated. Although a degree of local oversight is desirable, there was no documentation in these schemes of increased community participation as a result of PBF (or as a requirement of it). A recent study of the topic found that PBF does not automatically imply more 'voice' from the population, but introduces an interesting complement to health committees (Falisse et al. 2012). Some point out that the main country to take PBF to scale is Rwanda, which has done so with a centralised system (Toonen et al. 2009).

Discussion and some outstanding questions

P4P is an idea which in practice covers a range of modalities - it is hard to discuss them all within the same category. At its borders, it overlaps with other approaches and bodies of literature on contracting, provider payment systems, purchasing and health worker incentives. This paper has discussed the core concept while also reflecting on the diversity of experiences of P4P.

P4P approaches have spread quickly in the past decade, and it is interesting to consider the factors behind this. One is undoubtedly that it has intuitive appeal as a concept. Secondly, it has garnered a high degree of support from a number of influential multi- and bilateral donors (the World Bank, USAID, Government of Norway, and, to a lesser extent AusAid and DFID). It has also found a successful case study in Rwanda. Finally, it meets the needs of an era where the Millennium Development Goals (MDGs) are driving forward a target-oriented approach to health and development. This does however raise the question of how P4P will fare after 2015, when targets and also the resources behind those targets will change.

Some argue that P4P should be evaluated not only on the increase of a few indicators, but much more on its capacity to catalyse reform and to address structural problems (Vinard 2011). As this report has shown, there is little evidence of it transforming health systems to date. In Rwanda, it has been linked with decentralization and other reforms, but the direction of change is not clear – did Rwanda reform because it was introducing PBF, or (as seems more plausible) did it introduce PBF because there was already a wider agenda for change at the political level? The process of its development and introduction will influence how it operates (Witter et al. 2013) and it is important to understand how it affects different institutional actors and the wider political economy, which is another neglected area in the literature.

Other questions relating to health systems remain, including whether the use of P4P mechanisms add coherence to the health financing system in most LMIC countries, or is it adding to their fragmentation. The picture varies, but certainly in some countries it is adding to a complex mix of incentive regimes, with some countries operating a range of small- and large-scale P4P programmes. Cambodia's health sector experience to date, for example, with various devices for linking staff pay to performance have suffered from (i) fragmentation and a lack of a coherent and consistent set of rules governing such practices; (ii) a lack of oversight and monitoring of the implementation of the schemes, particularly at the level of individual staff; and (iii) absence of any rigorous evaluation of the effectiveness of such performance-related pay practices, in spite of widespread "piloting"

³ Some very practical tools such as sample contracts can be found at: <https://nphcda.thenewtechs.com/documents.html>.

(Hawkins 2011). It seems clear that P4P needs to be one approach within the overall health financing strategy – integrated with other tools and used as part of overall strategy and with broader national plans. It will have implications for national pay policies, for example, in other sectors.

It is also important to consider whether P4P is effective at getting funding to the frontline. One of the critiques of traditional budgeting is that funds tend to be absorbed at higher levels and fail to reach front-line providers. It is possible (but not yet investigated) that P4P might be a more effective approach to ensuring that health centres, for example, have adequate operating funds. The benefits and costs should be carefully examined, compared with other approaches (such as the direct facility funding approach piloted in Kenya (Opwora et al. 2009).

Its effect over time is another important area for investigation. Rapid improvements were noted during the start up of some P4P programmes but later payments became integral with health worker remuneration, which may lead to a levelling off or even a decline in performance over time, if there is no variation in payments (Toonen et al. 2009). Countries like the UK which have piloted a range of incentive schemes found that output/target payments tend to be cost inflationary: after achievement of performance targets, staff expect to continue to receive the incentive payment for *maintaining* the target. So the incentive is often incorporated into base pay, requiring further increases in pay for any new performance targets.

It is not always clear whether P4P intended is intended to be a long-term approach (a new funding channel) or a temporary strategy, to change organisational culture. This was not discussed in the papers identified here and yet is a critical question which will determine how it should be operationalized and monitored.

The five-year evaluation of the Global Fund concluded that ‘performance-based financing, a key tenet of the guiding principles, has evolved into a complex and burdensome system that has thus far focused more on project inputs and outputs than on development outcomes, departing from the vision of an outcome-based model’. This has been an issue for most P4P programmes, which may aim to reward outcomes but in fact involve paying for outputs and often quite closely specified processes too.

P4P will work best when there is spare capacity in the health system, such that small additional resources can leverage a large increase in outputs. Where this is not the case – where it motivates more activity but the system is already under strain – then the effects on quality of care are likely to be negative. Few of the studies examined here appeared to start with a clear assessment of what capacity constraints the system faced prior to P4P.

Being contrarian, it could be argued that health care producers need to be less output-oriented and more process-oriented (focussing on the quality of care and on how patients are treated, which is within their control, rather than on results, which are not). The health care system is peculiar in that it has many process goals, as well as outcome goals. Qualities such as empathy and listening skills from the health worker side form an important part of the healing process, but it is not easy to buy these. In higher income settings such as the UK, there is evidence that focus on targets has detracted in some cases from patient-centred care and the patient-clinician relationship (Oxman et al. 2007).

P4P in LMICs appears to be developing under its own dynamic but is perhaps not drawing enough on experience in OECD countries. Evidence from countries like the UK suggest that while the pay for performance framework produced rapid changes in behaviour, particularly with respect to improvements in processes, it was costly, the relation between performance targets and population

health improvements was not clear, and there was evidence that setting targets for some areas reduced performance in other areas of the service (Maynard & Bloor 2010).

An OECD study on performance-related pay (at the individual level) in the public sector was not encouraging. 'Performance pay is an appealing idea, but the experiences reviewed in this study indicate that its implementation is complex and difficult. Previous OECD studies on the impact of performance pay at the managerial level concluded that many of the schemes had failed to satisfy key motivational requirements for effective performance pay, because of design and implementation problems, but also because performance assessment is inherently difficult in the public sector..... Performance measurement in the public sector requires a large element of managerial judgement. The notion of performance itself is complex, owing to the difficulty of finding suitable quantitative indicators and because performance objectives often change with government policy. Many studies have concluded that the impact of PRP on performance is limited, and can in fact be negative' (OECD 2005). A systematic review of economic evaluations of pay for performance in the health sector (which only picked up studies from higher income countries) concluded that evidence was scarce and inconclusive, that P4P efficiency could not be demonstrated and that further research was required (Emmert et al. 2012).

There are few examples of public organisations in the OECD countries having withdrawn their performance-related pay policy (New Zealand being an exception). But the fact that organisations do not withdraw is not necessarily a very good indication of its effectiveness, because the costs of doing so are a deterrent (OECD 2005). Once entrenched, it can be hard to reform.

There has been a consensus for some time about the desirability of increased autonomy for providers, and also about the need for accountability and demonstration of results. Whether paying according to those results is more effective than more general performance management is an important question. P4P is a part of a wider 'managing for results' approach, which emphasises the need for explicit specification of objectives, the measurement of performance against those objectives, and the setting of performance targets. Some of these may link resources to targets, but more in the sense of making targets credible (ensuring resources are there to allow targets to be met, rather than paying for them retrospectively).

The usual direction of reform for individual employees as systems mature involves the consolidation of incentives into basic pay, and encouraging performance through the career system and through promotion of performance management and accountability systems focused on job descriptions/agreements and annual performance reviews.

There are also insights to be gained from wider literature on motivation. P4P is based on the economic agency theory. However, behavioural economics suggests that our behaviour is more complex than rational theories recognise. It emphasises the importance of social relations – of norms, perceptions of fairness and social rewards. This fits with insights from industrial psychology and literature on intrinsic motivation, especially in public service. At the least, they suggest that a combination of material and non-material rewards is likely to be more productive than material alone.

The extent to which financial incentives crowd out or in intrinsic motivation may depend to a large extent on the manner of their introduction – whether they are seen as recognising/rewarding effort, rather than as controlling/indicating a lack of trust in the agent (Myers 2008).

In general, tight performance management is good for poor performers but can demoralise good ones (it can limit their flexibility and creativity) – the knights and knaves hypothesis. This was

supported by a recent study in an NHS hospital, which found that the introduction of high powered incentives and tighter management controls was linked to a reduction in additional hours worked by more altruistic doctors and an increase by those who were less altruistic (Ensor et al. 2009)⁴. The incentive regime therefore needs to reflect the characteristics of the workforce. It could aim to differentiate, but applying different extrinsic incentives to different sections of a workforce in one organisation is likely to increase transactions costs (and possibly labour costs) and cause friction. This raises the question of whether health workers in LMICs are more knightly than knavish, and, indeed, whether they simply lack working conditions to facilitate good practices.

A modelling of incentive regimes to check shirkers without demotivating high performers concluded that the regime should combine two elements – intensive monitoring to persuade the shirking physicians to improve their performance but also non-pecuniary rewards and recognition to counter the deleterious effect of increased supervision for the high performers (Garcia-Prado 2005). Operational research in Zambia also concluded that ‘Non-financial awards are as motivating, if not more motivating, for staff than financial awards and do not generate as much conflict, suspicion, or frustration’ (Furth 2006). Studies of the effects of P4P at the individual level need to take account of selection effects over time (the kinds of workers who are attracted to work under these regimes) as well as cross-sectional motivation and effort effects (Gerhart & Fang 2013).

At the international level, P4P (and its affiliate of ‘cash on delivery’ aid) is sometimes presented as offering a new and liberating of doing business between donors and governments. Conditionality based originally on project goals and then on macro/policy demands is now replaced with a focus on outcomes (with governments free to determine how these are reached). However, some question the extent to which ‘Value for Money’ (VFM) agenda equates to good development practice and encourages national leadership and accountability⁵. Is it about buying results for donors, or is it indeed, as some argue, related to proving the purpose of aid to domestic audiences in developed countries⁶?

The compatibility of P4P approaches with the Paris Declaration also needs consideration. In relation to global health initiatives, governments often have a limited ability to negotiate on targets and in some cases (e.g. the Global Fund) parallel operating systems are required.

Donor policy-conditionality (such as budget support, linked to poverty-reduction goals) is a form of performance-related pay, even if not normally bracketed under that. The difference is that it is often linked to intermediate actions, such as government commitments, and leaves more flexibility to take into account factors which may have influenced non-performance (Pearson 2010).

Conclusions

P4P has spread quickly in the health sector in low- and middle-income countries over the past decade. It has been deployed in relation to two groups of services in particular – control of infectious diseases and sexual, reproductive and child health services. This paper focuses on experiences of the latter, though many of the lessons will be shared across the two groups. The paper has demonstrated the range of approaches which use a performance-based component, including global health initiatives (though these are not specifically focussed on SHR services), performance-based

⁴ Management reforms in the mid-2000s in the hospital were found to have improved productivity, with a small increase in costs (from 2% to 3% of overall expenditure being absorbed by management).

⁵ See <http://www.thebrokeronline.eu/Blogs/Busan-High-Level-Forum/Value-for-money-or-Results-Obsession-Disorder> for example

⁶ See <http://www.owen.org/blog/3275>, for example

contracting of NGOs, performance-based funding of local government, facility funding, and health worker incentive schemes.

The field is changing fast, so not all experiences can be captured, but based on a literature review some of the main documented programmes were described (25 in total). Evidence of their impact was assessed, based on a conceptual framework of the ultimate and intermediate goals of P4P.

It is clear that P4P can have a positive influence on health outputs, at least in some contexts, and in the short term. However, there are a number of important areas which remain to be better understood. These include:

- the long-term impact of P4P, if P4P is intended as a long-term financing mechanism as opposed to a short-term behaviour modifier
- its health systems impact, particularly in terms of:
 - integration/fragmentation of financing channels
 - effectiveness of purchasing arrangements
 - information systems
 - pay policy
 - non-targeted services (and how adverse effects can be controlled)
 - capacity building
 - governance arrangements
- its impact on health outcomes, which study designs have not yet been able to reveal
- its equity effects and impact on patient payments
- analysis of the cost-effectiveness of the approach, particularly related to alternative methods of achieving similar goals (wider performance management tools; increasing provider autonomy; direct facility funding; pay reforms)
- its sustainability, given the high degree of donor dependence to date (financial and technical) and also the focus on short-term MDG goals, which will change after 2015
- its strengths and weaknesses as a mechanism for international aid, especially in terms of consistency with the Paris Declaration and its impact on aid harmonisation and transactions costs
- how to increase its efficiency – particularly reducing the high overhead costs recorded in most schemes to date
- its effects on health workers' motivation and behaviour (in the LMIC setting), and how to maintain motivation over longer periods of time
- a better understanding of the contextual factors favouring the use of P4P, including in relation to the process of its introduction, and to the necessary systems capacity prior to introduction
- learning on design issues, such as how to design payment systems to minimise inflationary tendencies and promote quality of care, how to effectively incorporate non-monetary rewards, as well as on the magnitude of payments required to motivate in different settings

The overall evidence base remains weak, though it is growing. Schemes in LMICs should learn from and feed into lessons generated in higher income settings and in related literature sets on aid effectiveness, performance management, contracting, provider payment systems and health worker incentives and motivation. As P4P programmes mature, it is likely that they will be seen less as a stand-alone strategy and more as part of a range of purchasing and performance management approaches. Development and monitoring of P4P should also take a broader perspective, starting from a sound understanding of context and systems.

References and bibliography

- Amara, M. 2011, Designing of Sierra Leone simple performance based financing Scheme (PBF). Presentation to PBF workshop, Bujumbura, February 2011.
- Basinga, P., Gertler, P., Binagwaho, A., Soucat, A., Sturdy, J., & Vermeersch, C. 2010, Paying primary health centres for performance in Rwanda, Policy research working paper 5190; Washington, D.C.: World Bank.
- Basinga, P., Gertler, P., Binagwaho, A., Soucat, A., Sturdy, J., & Vermeersch, C. 2011, Effect on maternal and child health services in Rwanda of payment to primary health-care providers for performance: an impact evaluation, *The Lancet*, vol. 377, pp. 1421-1428.
- Bloom, E., Bhushan, I., Clingingsmith, D., Hong, R., King, E., Loevinsohn, B., Kremer, M., Schwartz, J.B. 2007, Contracting for Health: Evidence from Cambodia
<http://www.brookings.edu/views/papers/kremer/20060720cambodia.pdf>
- Brennan, R.J., Rogers, D., Sarcar, S., Vergeer, P. 2010, Identifying Indicators for Performance Based Contracting (PBC) is Key: The Case of Liberia. Results Based Financing for Health, Washington D.C: World Bank.
- Brenzel, L. 2009, Taking Stock: World Bank Experience with Results-Based Financing (RBF) in Health; World Bank: Washington D.C.
- Brenzel, L. and Naimoli, J., 2009, Health RBF in Practice: ten institutional challenges. Washington, D.C.: World Bank: Technical Brief.
- Busogoro, J.F. , Beith A. 2010, Pay-for-performance for Improved Health in Burundi. Bethesda, Maryland: HealthSystems 20/20 project, Abt Associates Inc.
- Canavan, A. & Swai, G. 2008, Payment for Performance (P4P) Evaluation: Tanzania Country Report for Cordaid. Amsterdam: KIT.
- Canavan, A., Toonen, J., Elovainio, R. 2008, Performance Based Financing: An International Review of the Literature, KIT Development Policy & Practice, Amsterdam: KIT.
- Chee, G., His N., Carlson K., Chankova S., Taylor P. 2007, Evaluation of the first five years of GAVI immunisation services support funding. Prepared for the GAVI Alliance. Bethesda, MD: Abt Associates Inc.
- Christianson, J., Leatherman, S., & Sutherland, K. 2007, *Financial incentives, healthcare providers and quality improvements: a review of the evidence*, The Health Foundation, London.
- Custers, T., Hurley, J. Klazinga, N. and Brown, A. 2008, Selecting effective incentive structures in health care: A decision framework to support health care purchasers in finding the right incentives to drive performance, *BMC Health Services Research*, vol. 8: 66.
- Dagur, V., Senauer, K., Switlick-Prose K. July 2010, Paying for Performance: The Janani Suraksha Yojana Program in India. Bethesda, Maryland: Health Systems 20/20 project, Abt Associates Inc.
- Deci, E., Koestner, R. and Ryan, M. 1999, A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation, *Psychological Bulletin*, Vol. 125, No. 6, 627-6.

- De Walque, D., Gertler, P., Arredondo, S., Kwan, A., Vermeersch, C., Bizimana, J., Binagwaho, A., Condo, J. 2013, Using Provider Performance Incentives to Increase HIV Testing and Counseling Services in Rwanda. Washington, DC: World Bank. Policy Research Working Paper 6364.
- Eichler R. 2006, Can "Pay for Performance" Increase Utilization by the Poor and Improve the Quality of Health Services? Discussion paper for the first meeting of the Working Group on Performance-Based Incentives. Washington D.C.: Center for Global Development.
- Eichler, R. and De, S. 2008, Paying for Performance in Health: A Guide to Developing the Blueprint. Bethesda, MD: Health Systems 20/20, Abt Associates Inc.
- Eichler, R., Auxila, P., Antoine, U., Desmangles, B. 2007, Performance-Based Incentives for Health: Six Years of Results from Supply-Side Programs in Haiti, Working paper No. 121. Washington D.C.: Center for Global Development.
- Eichler, R., Glassman, A., 2008, Health Systems Strengthening Via Performance-based Aid: Creating Incentives to Perform and To Measure Results, Global Health Financing Initiative Working Paper 3, Brookings Global Economy and Development Series, Washington D.C.: Brookings.
- Eldridge, C. & Palmer, N. 2009, Performance-based payment: some reflections on the discourse, evidence and unanswered questions, *Health Policy and Planning*, vol. 24, no. 3, pp. 160-166.
- El-Saharty, S., El-Hayatmy, M., Prose, KS., Eichler, R. 2010, P4P for Improved Health in Egypt, Bethesda, MD: Health Systems 20/20, Abt Associates Inc.
- Emmert, M., Eijkenaar, F., Kemter, H., Esslinger, A., Schoffski, O. 2012, Economic evaluation of pay-for-performance in health care: a systematic review. *European Journal of Health Economics*, vol, 13; pp. 755–767.
- Ensor, T., Kilby, A., Myers, J., Hay, R., 2009, Management reforms and performance in a London Teaching Hospital. Oxford: Oxford Policy Institute.
- Falisse, J-B., Meessen, B., Ndayishimiye, J., Bossuyt, M. 2012, Community participation and voice mechanisms under performance-based financing schemes in Burundi. *Tropical Medicine and International Health*, vol. 17, no 5, pp 674–682.
- Fox, S., Witter, S., Wylde, E., Mafuta, E., Lievens, T. 2013, Paying for performance in a fragmented, fragile state: reflections from Katanga Province, Democratic Republic of Congo. *Health Policy and Planning*. doi: 10.1093/heapol/czs138.
- Furth R. 2006. Zambia Pilot Study of Performance-based Incentives. Operations Research Results. Published for the U.S. Agency for International Development Agency (USAID) by the Quality Assurance Project (QAP).
- Garcia-Prado, A. 2005, Sweetening the carrot: motivating public physicians for better performance, , WPS3772. Washington, D.C.: World Bank.
- Gerhart, B. & Fang, M. 2013, Pay for (individual) performance: issues, claims, evidence and the role of sorting effects. *Human Resource Management Review*
<http://dx.doi.org/10.1016/j.hrmr.2013.08.010>
- Gertler, P. and Vermeersch, C. 2012, Using performance incentives to improve health outcomes. Washington, DC: World Bank Policy Research Working Paper 6100.

Global Fund to fight AIDs, Tuberculosis and Malaria, 2009, Five year evaluation synthesis report. Geneva: GFATM.

Gonzales, G., Eichler, R., Beith, A., 2009, Pay for Performance for Women's Health Teams and Pregnant Women in the Philippines. Bethesda, MD: Health Systems 2020.

Harris, M. 2011, Brazil's Family Health Programme: A cost effective success that higher income countries could learn from, Editorial, *British Medical Journal* 341:c4945.

Hawkins, L., 2011, Synthesis assessment of medium-term issues and options for supply and demand-side initiatives for improving quality and access to health services, supported under HSP 2. Draft paper for Cambodia mid-term review.

Hecht, R., Batson, A., Brenzel, L. 2004, Making Health Care Accountable: The New Focus on Performance Based Funding of Health Services in Developing Countries, in *Health and Development: Why Investing in Health is Critical for Achieving Economic Development Goals*, Finance & development, Washington DC: International Monetary Fund.

Jack, W., 2003, Contracting for Health Services: An evaluation of recent reforms in Nicaragua. *Health Policy and Planning*; 18(2): 195-204.

Johannes, L., Mullen, P., Okwero, P., Schneidman, M. 2008, Performance based financing in health: The experience of three projects in Africa, Note No. 19, OBAApproaches. Washington D.C.: Global Partnership on Output-based Aid/The World Bank.

Kalk, A., Paul, F., Grabosch, E. 2010, Paying for Performance in Rwanda: Does it Pay Off?, *Tropical Medicine and International Health*, 15(2): 182–190.

Lim, S., Dandona, L., Hoisington, J., James, S., Hogan, M., Gakidou, E. 2010, India's Janani Suraksha Yojana, A Conditional Cash Transfer Programme to Increase Births in Health Facilities: An Impact Evaluation. *The Lancet*, 375(9730): 2009-2023.

Lu C., Michaud C., Gakidou E., Khan K., Murray C. 2006, Effect of the Global Alliance for Vaccines and Immunisation on diphtheria, tetanus, and pertussis vaccine coverage: an independent assessment. *Lancet*; 368(9541):1088-1095.

Macq, J., Chiem J-C. 2011, Looking at the effects of performance-based financing through a complex adaptive systems lens. *WHO Bulletin*; vol. 89(9): pp. 699-700.

Macro International, 2009, The Five-Year Evaluation of the Global Fund to Fight AIDS, Tuberculosis, and Malaria Synthesis of Study Areas 1, 2 and 3. Washington D.C.: Macro International.

Maestad, O. 2007, Rewarding Safe Motherhood: How can Performance-Based Funding Reduce Maternal and Newborn Mortality in Tanzania? Bergen: Chr. Michelsen Institute.

Maynard, A. & Bloor, K. 2010, Will financial incentives and penalties improve hospital care?, *British Medical Journal*, vol. 340, no. 88.

Marquand, D., 2004, Decline of the Public: the Hollowing Out of Citizenship. Cambridge: Polity Press.

Meessen, B., Musango, L., Kashala, JPI., Lemlin, J. 2006, Reviewing institutions of rural health centres: the Performance Initiative in Butare, Rwanda, *Tropical Medicine and International Health*, 11(8): 1303–1317.

Meessen, B., Soucat, A., & Sekabaraga, C. 2011, Performance-based financing: just a donor fad or a catalyst towards comprehensive health-care reform? *Bulletin of the World Health Organization*, vol. 89, no. 153, p. 156.

Meuwissen L., Gorter A., Kester A., Knottnerus J. 2006, Can a Comprehensive Voucher Programme Prompt Changes in Doctors' Knowledge, Attitudes and Practices Related to Sexual and Reproductive Health Care for Adolescents? A case study from Latin America, *Tropical Medicine and International Health*, 11(6):889-98.

Morgan, L. 2010, Some days are better than others: Lessons Learnt from Uganda's First Results Based Financing Pilot, RBF Feature story, Washington D.C.: The World Bank.

Morgan, L. 2010a, A contract too far? Will performance-based contracting (really) work in Southern Sudan, RBF Feature Story, Washington, D.C.: The World Bank.

Morgan, L. 2011, "When the Hustle Gets Rough: Making Performance-Based Contracting Work in Liberia", RBF Feature Story, Washington DC: The World Bank.

Morgan, L., Beith, A., Eichler, R., 2011, Performance-Based Incentives for Maternal Health: Taking Stock of Current Programs and Future Potentials. Bethesda, MD: Health Systems 20/20 project, Abt Associates Inc.

Morgan, L., Eichler, R. 2009, Pay for Performance in Tanzania. Bethesda. Bethesda, MD: Health Systems 20/20 project, Abt Associates Inc.

Myers, J., 2008, Public Service Motivation' and performance incentives: a literature review. Oxford: Oxford Policy Institute.

Musgrove, P. 2011, Rewards for good performance or results: a short glossary. Washington, D.C.: World Bank.

Naimoli, J. and Vergeer, P. (eds.) 2010, A Series of Snapshots of Experiences in Verifying Performance Linked to Financial Incentives for Results-Based Financing (RBF) Programs from Selected Countries; Results based financing for health verification at a glance. Washington D.C.: The World Bank.

OECD, 2005, Performance-related Pay Policies for Government Employees. Paris: OECD.

Opwora, A., Kabare, M., Molyneux, S., & Goodman, C., 2009, Financing Kenyan health centres and dispensaries: exploring the implementation and effects of Direct Facility Funding in Kenya. Nairobi: CREHS Policy Brief.

Oxman A., Lavis J., Fretheim A. 2007, Use of evidence in WHO recommendations, *Lancet*, Vol. 369, pp. 1883-9.

Oxman A., Fretheim A. 2008, An overview of research on the effects of results-based financing. Report No. 16. Oslo: Norwegian Knowledge Centre for the Health Services.

Peabody, J., Shimkhada, R., Quimbo, S., Florentino, J., Bacate, M., McCulloch, C., & Solon, O. 2011, "Financial incentives and measurement improved physicians' quality of care In the Philippines", *Health Affairs*, vol. 30, no. 4, pp. 773-781.

Pearson, M., 2010, Results based aid and results based financing: What are they? Have they delivered results? London: HLSP Institute.

Perrin, B. 2013, Evaluation of payment by results: current approaches, future needs. Report of a study commissioned for the Department for International Development, London.

http://www.oecd.org/derec/unitedkingdom/DFID_PaymentbyResults.pdf

Petersen, L., Woodard, L., Urech, T., Daw, C., Sookanan, S. 2006, Does Pay-for-Performance Improve the Quality of Health Care? *Annals of Internal Medicine*, 145:265-272.

- Powell-Jackson, T., Neupane, B., Tiwari, S., Morrison, J., & Costello, A. 2008, Final report of the evaluation of the Safe Delivery Incentive Programme. London: DFID.
- Rowan, M. 2009, Performance-Based Disbursement: Innovations in FP Project. Delhi: Futures Group International.
- Savedoff, W. 2010, *Basic economics of results-based financing in health*. Bath, Maine: Social Insight.
- Sekabaraga, C., Soucat, A., Diop, F., Martin, G. 2010, Innovative Financing for Health in Rwanda: A Report of Successful Reforms, in *Yes Africa Can: success stories from a dynamic continent*, Punam Chuhan-Pole, P. & Angwafo, M. (eds.); Washington D.C.: The World Bank.
- Sekabaraga, C., Diop, F., Soucat, A. 2011, Can innovative health financing policies increase access to MDG-related services? Evidence from Rwanda, *Health Policy and Planning*; 26 :ii52–ii62.
- Sheffler, R. 2010, Pay For Performance (P4P) programs in health services: what is the evidence? World Health Report 2010 Background Paper 31. Geneva : WHO.
- Soeters R. & Kimakuka C., 2008, Résultats de l'Enquête Ménage, l'Enquête Qualité, et l'Enquête Infirmiers Titulaires. Pour le programme Achat de Performance dans les Zones de Santé du District Sanitaire Nord du Sud Kivu. The Hague: CORDAID.
- Soeters R. & Kimakuka C., 2009, Rapport étude d'évaluation, Achat de Performance Burundi. The Hague: CORDAID.
- Soeters, R., Musango, L., & Meessen, B. 2005, Comparison of two output based schemes in Butare and Cyangugu provinces with two control provinces in Rwanda. Kigali: GBPOA, World Bank, and Ministry of Health Rwanda.
- Soeters, R., Peerenboom, PB., Mushagalusa, P., Kimanuka, C. 2011, Performance-Based Financing Experiment Improved Health Care in the Democratic Republic Of Congo, *Health Affairs*, 30:81518-1527.
- Soeters R. 2013, PBF in Action: Theory and Instruments. Course Guide - Performance Based Financing (PBF). The Hague: CORDAID-SINA.
- Sondorp, E., Palmer, N., Strong, L. and Wali, A. 2009, Afghanistan: Paying NGOs for Performance in a Post-conflict Setting; Washington D.C., Center for Global Development.
- Ssengooba F., McPake B., Palmer N. 2012, Why performance-based contracting failed in Uganda: an "open-box" evaluation of a complex health system intervention. *Social Science & Medicine*; pp. 377-83.
- Toonen, J., Canavan, A., Vergeer, P., & Elovainio, R. 2009, Performance Based Financing: a synthesis report. Amsterdam: KIT, in collaboration with Cordaid and WHO.
- Toonen, J. and Van der Wal, B. 2012, Results based financing in healthcare: developing an RBF approach for healthcare in different contexts: the cases of Mali and Ghana. Amsterdam: KIT.
- UNFPA, 2009, Concurrent Assessment of Janani Suraksha Yojana (JSY) in Selected States: Bihar, Madhya Pradesh, Orissa, Rajasthan, Uttar Pradesh, New Delhi, India: United Nations Population Fund.
- USAID, 2010, Performance Based Incentives: Primer for USAID Missions, Washington D.C.: United States Agency for International Development.

- Vergeer, P. & Collins, C. 2008, Payment for Performance (P4P) Evaluation: Zambia Country Report, Amsterdam: Cordaid/KIT.
- Vermeersch C., Rothenbuhler E., Sturdy J. 2012, Impact evaluation toolkit: measuring the impact of results-based financing on maternal and child health. Washington, D.C.: World Bank.
- Vinard, P., 2011, Lesson learning and documenting health system development and the role of performance based financing experience in Burundi, DR Congo, Rwanda and South Sudan. Amsterdam: HealthNet TPO/ALTER.
- WHO, 2010, Increasing access to health workers in remote and rural areas through improved retention: global policy recommendations. Geneva: World Health Organisation.
- WHO, 2011, Performance based grants for reproductive health in the Philippines. Policy Brief, Department of Reproductive Health and Research; Manilla, Philippines. Manilla: World Health Organization and ATENEO de Manilla University Graduate School of Business.
- William, J. 2003, Contracting for Health Services: An evaluation of recent reforms in Nicaragua, *Health Policy and Planning*, 18(2): 195-204.
- Witter, S. 2013, Demand-side financing for strengthening delivery of sexual and reproductive health services: evidence synthesis paper. Washington, D.C.: World Bank.
- Witter, S., Toonen, J. Meessen, B., Kagubare, J., Fritsche, G. and Vaughan, K. 2013, Performance-based financing as a health system reform– mapping the key dimensions for monitoring and evaluation. *BMC Health Services Research*, 13:36.
- Witter, S., Fretheim, A., Kessy, F., & Lindahl, A. 2012, Paying for performance to improve the delivery of health interventions in low and middle-income countries, *Cochrane Library*, issue 2.
- Witter, S., Khadka, S., Nath, H., & Tiwari, S. 2011b, The national free delivery policy in Nepal: early evidence of its effects on health facilities, *Health Policy and Planning*, vol. 26, p. ii84-ii91.
- Witter, S., Zulfikar, T., Javeed, S., Khan, A., & Bari, A. 2011a, Paying health workers for performance in Battagram district, Pakistan, *Human Resources for Health*, vol. 9, no. 23.
- World Bank, 2007, Performance based contracting to improve health services in post-conflict situations: DRC. Washington D.C.: The World Bank.
- Zeng, W., Cros, M., Wright, K., Shepard, D. 2012, Impact of performance-based financing on primary health care services in Haiti, *Health Policy and Planning*, vol. 10.

Table 1A Description of P4P schemes

Programme	Year implemented	Scale of implementation	Description of P4P intervention	Services targeted	Authors
PAYMENTS TO GOVERNMENTS					
Global Fund to fight AIDS, Tuberculosis and Malaria (GFATM)	2001 onwards	Global - multiple countries (136)	Performance based grants to countries through public-private partnerships.	Health services related to prevention and treatment of malaria, TB, and HIV/AIDS	Macro International 2009
Global Alliance on Vaccine Initiative (GAVI) -HSS window	2005 onwards (HSS window)	Global - multiple countries (commitments to 44; disbursements to 36 so far for HSS)	Payments to governments aimed at Health System Strengthening. Governments develop proposals and present to GAVI. Funding is linked to targeted outputs.	Immunizations and related child and maternal care	HLSP 2009; www.gavi.org
PAYMENTS TO MID LEVEL ORGANIZATIONS (NGOs, local government)					
Family Health programme, Brazil	1994 onwards	National	Municipalities face negative incentives if targets are not met - their future funding is reduced and remedial measures are put in place to improve performance. Later pilot with performance bonuses of up to 15% if targets met	Basic health care services including maternal and child healthcare.	Hecht et al. 2004; Harris 2010
Plan Nacer in Argentina	2004 onwards	National	Payments made from central government to provinces, and provinces pay providers. Providers reimbursed a monthly base fee (60%) on the basis of FFS for agreed services. This is determined by the number of eligible enrollees in Plan Nacer. Remaining 40% provided as a "complementary transfer", determined by the achievement of stated targets for ten output and outcome health indicators.	MNCH package includes 80 services to pregnant women and mothers (up to 45 days after delivery), and to children under the age of 6 not covered by health insurance. 10 indicators for performance are: (i) timely inclusion of women for ANC , (ii) Effective neonatal/delivery care (Apgar Score), (iii) Effective prenatal care and premature birth prevention (weight above 2.5kg), (iv) Quality prenatal and delivery care (mothers immunised and tested for STD), (v) Medical auditing maternal and infant deaths, (vi) Immunization coverage (measles), (vii) Sexual and reproductive health care, (viii) Well child care (<1 year), (ix) Well child care (1-6 years old), (x) Inclusion of Indigenous Population.	Naimoli & Vergeer 2010; Eichler & Glassman 2008

Programme	Year	Scale of implementation	Description of P4P intervention	Services targeted	Authors
Innovations in Family Planning Services, Uttar Pradesh India	1992 onwards	State of Uttar Pradesh	Performance based disbursement of funding from USAID to Indian registered “society,” the State Innovations in FP Services Project Agency (SIFPSA). SIFPSA in turn contracts public sector and NGOs for provision of services (this is done on cost reimbursement and does not include PBF elements).	Overall focus on improving access to and utilization of services. Since inception, program has expanded services targeted over 3 phases: Phase 1 (1994-2002): focus on strengthening government hospitals, district action plans, clinical trainings, IEC activities, and contraceptive social marketing; Phase 2 (2004-09): focus on larger reproductive health services in addition to FP; Phase 3 (2009-2012): focus on HSS, evaluating PPPs, training and human capacity building, and behavior change communication	USAID 2010
Output-based payments in Haiti	1999 onwards	Initially 3 NGOs, and later expanded to 25 NGOs by 2005, providing basic health services to 2.7 million people	Contracting out of services to NGOs - USAID-funded project initiated in 1999	Targets focused on service delivery included: ANC (at least 3 visits), full immunization by age 1, reduced discontinuation of oral and injectable FP methods, post natal visits, assisted deliveries, children weighed and enrolled in nutritional recuperation. Additional management targets included commodities supply management, timely reporting, following guidelines for financial management, human resources management and essentials drugs logistics, strengthening organizational structure, implementation of management audit recommendations, and use of CORE (Cost and Revenue Analysis Tool)	Eichler et al. 2007
Contracting of Health Services Pilot Project, Cambodia	1998-2003	5 districts - 3 contracting in and 2 contracting out	Performance based contracting, with two variants (contracting in and out). Contracting in districts were expected to work within the government structure and use their personnel, supply chain, etc. Contracting out districts had full authority over their management, hirings, procurement.	Delivery of a minimum package of services for maternal and child health: childhood immunization, administration of vitamin A to children, antenatal care for pregnant women, child delivery by a trained professional, delivery in a health facility, the knowledge and use of birth control, and use of public facilities when seeking curative care	Bloom et al. 2007

Programme	Year	Scale of implementation	Description of P4P intervention	Services targeted	Authors
Performance-based contracting in the DRC	2002 onwards	85 health zones covering 10 million people (~15% of the total population), at least in expanded phase since 2008	PBC through IDA-financed multi-sectoral project with 10 international and national church-based NGOs, who channel resources and technical resources to the health zones	Family planning, maternal and child health, outpatient utilization	Johannes et al. 2008
Paying NGOs for performance in Afghanistan	2003 onwards	8 provinces with NGO contracts and 3 with contracts with MoPH	Performance-based contracting of NGOs to provide basic services. World Bank contracts NGOs in three provinces, via Ministry of Public Health. USAID also has contracts with NGOs, but negative incentives (payment is withheld if targets are not met). Proposed changes from 2006 onwards with incentive being extension of contract for an additional 2.5 years. European Commission contracts with NGOs and sets targets but no PBF element	Basic package of services (including MNCH)	Sondorp et al. 2009
Pay for performance in South Sudan	2005 onwards	No information given	Performance based contracts with three lead agencies (international organisations)	Basic health services including maternal and child health. Maternal health indicators include ANC, assisted deliveries, PNC	Morgan 2011
Performance based contracting in Liberia	2009 onwards	105 health facilities in 7 counties funded under the 5-year RBHS project;	Performance based contracting with NGOs: Rebuilding Basic Health Services programme contracts NGOs which provide management support to health facilities. Performance incentives initially included a 6% payment only but later revised to include in-kind payments as well.	Basic health care package including maternal health: family planning, facility based deliveries, ante-natal and post-natal care, HIV counselling and treatment	Brennan et al. 2010; Morgan 2011
PAYMENTS TO FACILITIES					

Programme	Year	Scale of implementation	Description of P4P intervention	Services targeted	Authors
Pay for Performance in Rwanda	2006 onwards (with pilot initiated earlier)	80 facilities as part of national scale up of services initiated in 2006	Pay for performance (P4P) - payments made directly to primary level health facilities according to targeted services delivered	MNCH related services, with performance measured against 14 indicators related to service delivery, and weighted for quality. Payment varies for different services. For the health sector as a whole, the President has signed contracts with district mayors, focusing on the following performance indicators: (a) the number of people adhering to health insurance; (b) the number of institutional deliveries; (c) the number of women using family planning methods; and (d) the use of insecticide-treated bed nets.	Basinga et al 2011; Basinga et al. 2010; Sekabaraga et al 2011; Kalk et al 2010
Performance-based financing in DRC	2006 onwards	2 districts in DRC: 39 health centres and 4 hospitals	Performance-based contracting with health facilities. Autonomous health facilities managers invited to submit business plans on a quarterly basis to purchasing authority. Funds used for recruiting and motivating staff, social marketing, rehabilitating infrastructure, developing subcontracts with private providers and purchasing drugs. Facility managers had the authority to negotiate user fees with communities	Participating health centres received subsidies for 16 indicators, including use of oral or injectable contraceptives by women, facility-based births, outpatient visits, number of bed days, and full immunization of children by age 1 year. Hospitals received subsidies for 22 general indicators and 8 HIV/AIDS indicators.	Soeters et al. 2011
Pay for Performance in Burundi	2006 onwards	Piloted in 2006 in 3 provinces, with national scale up beginning in 2009 on a roll out basis (initially covering 6 additional provinces)	PBF payments to facilities in selected districts - piloted initially with support from INGOs from 2006, and then scaled up from 2009. Quantitative and qualitative indicators are used. Quality is monitored through service-specific composite indicators. Pilot programmes assessed quality on a quarterly basis, with a bonus of up to 15% of score on quantitative results during the same time period. Under the scale up, the MoH is raising this bonus to 25%.	Basic health services, including outpatient visits; family planning; and maternal and child healthcare (among others)	Busogoro & Beith 2010; Toonen et al. 2009

Programme	Year	Scale of implementation	Description of P4P intervention	Services targeted	Authors
CORDAID Pay for Performance pilot project in Tanzania	2006-2008	64 church health facilities in five dioceses, comprising of 13 hospitals, 12 health centres and 39 dispensaries supported by Cordaid (evaluation limited to 3 of 5 diocese)	Pay for performance pilot, managed by CORDAID, funding selected diocese - part fixed budget upfront, part in relation to hitting targets for five core indicators	Outpatient and inpatient utilization of services; supervised deliveries; new VCT clients; and drug management (also measure ANC)	Canavan & Swai 2008
Performance-based contracting in Uganda, 2003-6	2003-2006	Experiment in 5 districts covering 118 facilities (68 PNFPs)	Performance-based contracting to PNFP facilities. Two different interventions: treatment group B: base grant from government but with freedom on how to spend it; treatment group C was also awarded bonuses, in addition to base grant and freedom on how to spend it.	Outpatient utilization, maternal and child health, family planning	Morgan 2010
Performance-based Contracting Pilot, Nicaragua	2000 onwards	6 hospitals - pilot programme	Performance contracts with hospitals	There are 4 categories under which bonus is provided, which are: service targets, management, quality and organization of services	Jack 2003
Pay for Performance, Egypt	2001 onwards	5 governorates	Incentives given to providers for services delivered against set targets.	Primary healthcare, including maternal and child care	El-Saharty et al. 2010
Pay for Performance Pilot in Zambia	2007 onwards	5 mission hospitals and later 3 mission health centres	Mission facilities paid for meeting targets (50% of funding is fixed; 50% conditional on targets). Four targets used, each with equal weight and set at same level for all facilities	Inpatient turnover; facility based deliveries; VCT user rate; drug supply	Vergeer & Collins 2008
Performance-based Grants for Reproductive Health in the Philippines (PBG1)	2008 onwards	Unclear. Brief discusses examples from 4 provinces.	PBG-1 are incentivised grants aimed at funding FP and RH programmes by local governments.	Pregnancy related care and family planning	WHO 2011
PAYMENTS TO HEALTH CARE WORKERS					
Pay for Performance in Tanzania	2008-2010	National	Performance-based incentive paid to health workers against achieving specific results	Maternal and child health (MDGs 4 and 5), which for maternal health includes facility based births and intermittent preventative	Morgan and Eichler 2009

Programme	Year	Scale of implementation	Description of P4P intervention	Services targeted	Authors
				treatment 2 for pregnant women.	
Performance-based Grants for Reproductive Health in the Philippines (PBG2)	2008 onwards	Unclear, but appears to be on wide scale across provinces. Brief only discusses examples from 4 provinces.	PBG-2 are direct performance-based grants aimed at TBAs, CHW, midwives to direct women to facilities for ANC and birth. PBG-2 also include demand side incentives.	Pregnancy related care and family planning	WHO 2011
Cambodia Midwifery Allowance	2007 onwards	National	Incentives to midwives for facility based births	Maternal and child health, with a focus on facility based deliveries	Hawkins 2011; Murakami 2009
Battagram P4P project, Pakistan, Save the Children	2008-2010	District level intervention in Battagram (North Western Frontier Province)	Performance-based incentives (PBI) to government employed health facility workers. These health workers covered under the project were eligible to receive an additional 20-35% of their pay based on performance. Health workers directly employed by Save the Children were not eligible for the bonus, but received higher payments.	Basic health services including maternal and child health. Maternal health included ANC, TT2, assisted deliveries, PNC	Witter et al. 2011
Safe Delivery Incentive Program, Nepal	2005 onwards	National	The SDIP is a CCT programme with a PBF component for health workers. The incentive, worth NRs. 300, is provided to the health team for each delivery they assist in a public health institution or at the woman's home.	Home and facility-based supervised deliveries	Powell-Jackson et al. 2008; Ensor et al. 2009
Janani Soraksha Yojana	2005 onwards	National (though UNFPA study looks at 5 states: Bihar, Madhya Pradesh, Orissa, Rajasthan, Uttar Pradesh)	Janani Suraksha Yojana (JSY) - provider payments for institutional births; payments made to community health workers. Incentive payments for women (CCT) and ASHAs (community health workers); additional payments for administrative costs and IEC.	MNCH: At least 3 ANC visits, institutional deliveries, immunization of newborns, postnatal checkup, and counselling to encourage breastfeeding - however in practical terms, the trigger is institutional delivery (institutional births are defined as births at government facilities and private facilities accredited under the programme)	Lim et al 2010; UNFPA 2009 (assessment of JSY in 5 States)

Table 2B Description of P4P schemes (continued)

Programme	Who received P4P payments?	Payment methods	Other accompanying components	Summary of evaluation/study design	Authors
				Quality of evidence	
PAYMENTS TO GOVERNMENTS					
Global Fund to fight AIDS, Tuberculosis and Malaria (GFATM)	The evaluation found that while the GFATM is expected to enter into partnerships with implementing agencies, there were few contractually binding partnerships. Instead the GFATM relied on a "friendship model" with affirmations from committed partners. Implementing agencies could include private or public sector or NGOs.	Disbursements are made periodically, conditional on performance results.	Other global programmes such as the U.S. President's plan for AIDS Relief (PEPFAR), and other partners, including the World Bank's Multi-Country HIV/AIDS programme (MAP) and the U.S. President's Malaria Initiative have also helped to focus efforts on the three diseases	Three separate evaluations performed using a mix of methods and focusing on 18 countries, including primary data collection through district comprehensive assessments; review of secondary data such as Demographic and Health Survey results and country health information system data; quantitative analyses to assess grant performance; review of Global Fund documentation and a broader base of literature; and qualitative analyses of focused interviews with Global Fund Board Members, Secretariat Staff, implementers and partners at the global and country levels. The evaluation was conducted on a very short timeline. Evaluation relates to GFATM as a whole - not just the PBF element.	Macro International 2009
Global Alliance on Vaccine Initiative (GAVI) -HSS window	National governments	Disbursements are made periodically, conditional on performance results.	Three other windows for funding for immunizations and vaccines and CSO support also exist to which countries can apply for funding. In addition, there are other larger donors (especially the World Bank) that are supporting HSS. In comparison, GAVI HSS support is smaller.	Evaluation is based on 21 country case studies. Since HSS was only introduced recently, the window of time was not sufficient enough to capture impact. While mention is made of maternal care, specifics are lacking. Moreover, the evaluation states that current it would be difficult to separate out the impact of GAVI HSS in the outcomes indicators	HLSP 2009; www.gavi.alliance.org
PAYMENTS TO MID LEVEL ORGANIZATIONS (NGOs, local government)					
Family Health programme, Brazil	Municipalities receive funding from federal government through the provincial governments. This is a public health sector programme	Healthcare expenditure is federally mandated, with contributions from the regional and national coffers	No information given	Summary of scheme; editorial	Hecht et al. 2004; Harris 2010

Programme	Who received P4P payments?	Payment methods	Other accompanying components	Summary of evaluation/study design	Authors
				Quality of evidence	
Plan Nacer in Argentina	Facilities (both public and private). Staff can use up to 50% of payments for staff.	At the Government level: The World Bank provides lending for Plan Nacer. Future lending from World Bank for the programme is contingent upon meeting certain milestones during lending period. At the local level: Upon verification of results, transfers made from National Ministry of Health to Provincial Health Insurance Unit which pays the providers	Plan Nacer has a health insurance component to encourage demand side uptake of services.	These are descriptive reports, not impact evaluations	Niamoli & Vergeer 2010; Eichler & Glassman 2008
Innovations in Family Planning Services, Uttar Pradesh India	SIFPSA receives payments from USAID. SIFPSA then pays its contractors.	Payments to SIFPSA are made when agreed upon benchmarks are achieved. Benchmarks are all costed and funds are released only after the benchmark is achieved. The majority of the measures are process-oriented.	Since 2005, JSY has been operating nationally, which has a strong CCT (demand side) component. It also has a supply side component with payments to health workers through facilities.	This is not an evaluation – just a short description of the programme.	USAID 2010
Output-based payments in Haiti	Payments made from USAID to NGOs via MSH (Management Sciences for Health - a US based international NGO). Some part of bonus could be passed on to staff	This evolved over time with the setting of targets. 1999-2001: 95% of a negotiated budget was paid in fixed quarterly sums and an additional 10% was conditional on results (NGOs could be paid up to 105%); 2002-04: fixed quarterly payments up to 95% of a negotiated budget, with 5% of the award fee (referred to as the “withhold”) based on achieving performance on management indicators and the other 5% of the award fee (referred to as a “bonus”) linked service indicators. In 2005, this changed again: NGOs can earn up to 106% if all targets are met. PBF associated with all aspects including signing contract and submission of reports.	Not discussed, but technical assistance, increased autonomy, flexibility and reduced reporting are clearly all important changes which accompanied the scheme.	Project conducted baseline survey through contracting an independent firm for the catchment area for the 3 NGOs initially contracted in 1999 using cluster sampling. Data were collected on immunization, ORS, ANC visits, discontinuation rates of oral and injectable contraceptives. Additionally, data on waiting times was gathered by measuring waiting times in a sample of institutions at different intervals. NGOs were not selected randomly, but rather those were inducted into the program that were perceived to be capable of meeting the requirements of the project. The study measures trend against baseline. There was no control group against which the performance was measured. Panel regressions aimed to isolate both NGO specific effects and contract period effects that may contribute to improved results. However, the lack of control group, the ancillary components and the switch from a 100% reimbursement to a 95% fixed costs model all make attribution of effects to PBF hard. The study is a CGD working paper.	Eichler et al. 2007

Programme	Who received P4P payments?	Payment methods	Other accompanying components	Summary of evaluation/study design	Authors
				Quality of evidence	
Contracting of Health Services Pilot Project, Cambodia	Government paid the contractors. Contractors were responsible for district and lower level payments. In the two contracting-out districts health workers were paid higher fixed salaries than previously paid with incentive of dismissal (negative incentive) for poor performance. In contracting-in districts the incentive consisted of a fixed supplement to staff members' government salaries plus a performance-based bonus. In the Preaeng district of Prey Veng province performance contracts were set up with health facilities and their management, who in turn set up contracts with staff. Here, it is reported that staff members received a guaranteed supplement of 55% of their government salaries plus a 30% performance bonus and a 15% punctuality bonus.	Contracting-out districts received their funds directly from the ADB after the Ministry of Health made a payment request. The contracting-in districts received the management fee portion of their contract budget in the same manner. Operating funds and supplies were provided to the contracting-in and comparison districts through normal government channels.	No information given	Randomised trial. Measured against baseline and control. Baseline in 1997 with full follow up in 2003 consisting of survey on perceptions of the quality of care at government facilities; survey of the 143 health centres in the project area (no baseline); and administrative data on public expenditures during the project years compiled from Ministry of Health records (in 2004). Sample size was 3700 households consisting of 20,000 individuals.	Bloom et al. 2007
Performance-based contracting in the DRC	At the national level NGOs receive money through an IDA financed project. Then NGOs pay health zones. 15% of the budget is reserved for health worker incentive payments	No information given	No information given	Not an impact evaluation - World Bank OBA brief	Johannes et al. 2008

Programme	Who received P4P payments?	Payment methods	Other accompanying components	Summary of evaluation/study design	Authors
				Quality of evidence	
Paying NGOs for performance in Afghanistan	World Bank pays the MoPH through the MoF. The MoPH pay the NGOs. One contract per province. USAID and EC pay NGOs directly.	For World Bank project, bonuses are provided in stages, with 1 percent of the contract value payable for at least a 10 percent increase from the baseline for specified indicators. The final 5 percent bonus is paid at the end of the contract. Overall bonus is limited to 10% of contract value. Relative flexibility in how the budget is spent as long as National Salary Policy and specifications of the basic package of health services (what services, staffing patterns, and ratios of facility to population) are met.	None mentioned but considerable overall investment in rebuilding the health sector and beyond post-conflict	This is a case study assessment of the services provided in Afghanistan. It is a descriptive study and not an impact evaluation.	Sondorp et al. 2009
Pay for performance in South Sudan	Lead agencies/INGOS implementing the contracts	Lead agencies are paid every 6 months, with 70% being paid upon submission of the report, and 30% upon verification of results by the MOH. If targets are not met, the MOH has thirty days to negotiate with lead agencies on how to overcome the obstacles. However, there is no clause in the contracts that says that payment will not be made.	No information given	This is a feature story - no evaluation as yet	Morgan 2011
Performance based contracting in Liberia	Accredited NGOs	No information given	Incentives are also provided for other services including child health. The Ministry of Health and Social Welfare is also implementing a similar programme of direct contracting with counties and NGOs.	Briefs describing process of development, not impact	Brennan et al. 2010; Morgan 2011

Programme	Who received P4P payments?	Payment methods	Other accompanying components	Summary of evaluation/study design	Authors
				Quality of evidence	
PAYMENTS TO FACILITIES					
Pay for Performance in Rwanda	Facilities at district level providing basic care, including health centres and district hospitals; both public and private (faith-based) with a 60-40 split. Funds could be used at their own discretion. Most was provided to staff as incentives (roughly the funding ratio was 2:1 for incentives to staff versus running costs).	Quarterly remuneration against a given set of indicators for each health centre and for each district hospital for services delivered. Additional remunerations against "several mechanisms assuring the quality of the service delivered". Financial incentives linked to service type, ranging from US\$ 4.59 for institutional delivery to US\$0.09 for 1st ANC visit.	Overall Rwandan context: fiscal decentralization of health sector and development of health insurance 'mutuelles', both contributed to improvements in MNCH outcome. In addition to PBF funding, funding for AIDS from PEPFAR contributed to Rwanda's 4-fold increase in the health budget. Increase in workers' salaries; and widespread distribution of bed nets also contributed to the improvements in health outcomes.	Study design: treatment/control comparison: 80 facilities which were being incorporated into the programme taken as the treatment group and 86 as control group (total 166 facilities studied). Control group would continue to receive input-based financing for the next 23 months until national roll out of P4P was completed. Two surveys conducted (one at baseline, and one after 25 months). Data collected through facility questionnaires. Additionally household survey on 2158 households with children under age 5 also conducted (13 HH selected randomly from each catchment area). Maternal baseline characteristics taken from facility survey and child baseline from HH surveys. Possible limitations include: recall bias/error by interviewees; original randomised design compromised by prior implementation in some districts; not clear if the incentive effect of prenatal care extend to other services; does not show impact on health outcomes; no comparison of what the effect would have been if payments were made to individual practitioners instead of facilities (Basinga et al 2011).	Basinga et al 2011; Basinga et al. 2010; Sekabaraga et al 2011; Kalk et al 2010
Performance-based financing in DRC	39 health centres and 4 hospitals through contracting with a purchasing authority. In addition, health centre managers further signed 22 subcontracts with private clinics in their catchment areas for better coverage. Staff benefited indirectly through increase in facility revenues.	In 2006 an autonomous financing authority/purchasing authority was established under the PBF programme. All payments are made through this authority which signed contracts with district health centres and hospitals. Monthly subsidies are paid to participating facilities and could vary between \$200 and \$4,000 based on performance against indicators, with up to a 15% bonus for a score of 100%. Health facilities in remote areas received an additional 15% bonus due to their isolation.	Investment in districts differed according to the NGO which was supporting them. Fee regimes also differed across the participating districts/controls, and there were other significant contextual differences. Participating facilities received a range of technical support, including in developing business plans.	Baseline was developed through a stratified household cluster survey conducted in November 2005 (n= 240 households) in the two districts participating in the performance-based financing experiment and in two control districts (n= 200 households) prior to intervention. Post intervention assessment was conducted in 2008 to measure the same quantitative health service outputs and 6 indicators of patients' perceptions of quality. The post intervention assessment also included an evaluation of 53 indicators of health centre quality using logistic regression. To note is that the sample size is small, and is further compromised by cluster sampling.	Soeters et al. 2011

Programme	Who received P4P payments?	Payment methods	Other accompanying components	Summary of evaluation/study design	Authors
				Quality of evidence	
Pay for Performance in Burundi	Payments received by health facilities based on performance against quantitative indicators: health centres receive US\$700–1,500 per month, and district hospitals receive US\$3,500–5,000 per month. Facilities management have autonomy in allocating payments to two categories: staff or service quality improvements. However, staff financial incentives payments cannot be more than 50% of each payment.	<p>Pilot: Agence d’Achat de Performance (AAPs), autonomous NGO established bodies, with funding directly from the Ministry of Finance, are responsible for contracting health facilities. Under the HealthNet TPO model contracting was done between AAP, health facility and the Provincial Steering Committees, multisectoral bodies with representation from health facilities and sometimes led by staff from the provincial administration. In the CORDAID areas, contracts between health facilities and AAPs were negotiated and signed directly.</p> <p>Scale up: Provincial committees for verification and validation (CPVV) replace AAPs in the scale up. These are public-private entities, responsible for contract negotiation and signature, and data verification and validation. These will be complemented by the Provincial Health Management Teams (akin to the Provincial Steering Committee).</p>	A national P4P technical unit known as the Cellule Technique Nationale has been established to define the broader P4P strategy and to coordinate the programme. It is composed of the different stakeholders.	This is a case study providing an assessment of the P4P programme overall. It is a descriptive study and does not provide an impact evaluation. It reports on an unpublished evaluation of the pilot.	Busogoro & Beith 2010; Toonen et al. 2009

Programme	Who received P4P payments?	Payment methods	Other accompanying components	Summary of evaluation/study design	Authors
				Quality of evidence	
CORDAID Pay for Performance pilot project in Tanzania	Health facilities managed by the Dioceses via the Diocesan Health Offices (DHOs). Cordaid sets guidelines for use of bonuses: 50% maximum for staff motivation; equipment, drugs and supplies to a maximum of 30%; infrastructure max 20%; running cost (including maintenance and communication) to a maximum of 10%. The District Health Offices are eligible for 25% of the performance bonus allocation	Progress against pre-set targets is measured on a 6 monthly basis. Payment is made upon verification of results. Payments are made to the DHOs which then pay the health facilities. The annual allocation to health facilities is set at 50% as guaranteed base fund; and 50% earmarked as bonus allocation.	CORDAID has worked with the dioceses for decades.	This is a mainly qualitative study, conducted over a 3 week period, with data collected mainly through interviews with stakeholders at central and district levels and visits to sites (P4P mission-based DHOs and non P4P government facilities). Cordaid supports 5 dioceses of which 3 were selected for the study. Criteria for selection were: (i) remote populations with limited resources, and (ii) dioceses that were accessible (by air). A total of 18 health facilities were visited. Information was gathered through (i) health staff and management interviews; (ii) study of HMIS to extrapolate data (2005-2007); (iii) client satisfaction interviews with randomly selected health facility users (at these facilities); (iv) staff motivation questionnaire followed by focus group discussions; and (v) interviews with district and diocesan representatives, and community representatives where available. Limitations are presented in the paper. Main issues with quality of HMIS and financial data and there is no rigorous evaluation of the data. Additionally, only a few interviews could be conducted with community representatives (patient/user side perspective is limited). This is a report prepared by Cordaid on their programme in Tanzania and covers the period 2005-07	Canavan & Swai 2008
Performance-based contracting in Uganda, 2003-6	PNFPs	No information given	None mentioned. However, in the health sector, other changes were taking place, including an increase in the salaries in the public sector; which led to a shift in the movement of health workers from PNFPs to public sector.	Quasi-experimental design of study. 2.5 year study with 3 rounds of surveys. Two treatment groups compared with one control group. Control group consisted of public, private, and PNFP facilities and was subject to pre-existing financial arrangements. Treatment group B received base grant with autonomy on spending. Treatment group C received base grant with autonomy to spend as well as bonuses if self selected output targets were achieved. World Bank - RBF Feature Story	Morgan 2010

Programme	Who received P4P payments?	Payment methods	Other accompanying components	Summary of evaluation/study design	Authors
				Quality of evidence	
Performance-based Contracting Pilot, Nicaragua	Hospitals sign yearly contracts with the Ministry of Health (MINSa). MINSa pays the hospitals and it is up to the hospitals how to disburse the bonus (which could be as a employee bonus via salaries; or apply it towards improvements to the hospital).	Hospitals are scored on performance, and bonus provided on a schedule where services and management each can earn up to 20% of the total bonus, while quality and organization each can earn up to 30% of the total bonus.	None mentioned.	Policy evaluation. There is no impact evaluation on services. Focus is more on the political economy	Jack 2003
Pay for Performance, Egypt	Service providers (public and private) that are accredited and linked to the Family Health programme.	The Family Health Fund (autonomous body) contracts and pays public, private and NGO health service providers. When it is verified that a health care facility meets targets, FHF makes a cash payment to the facility manager, who then distributes the incentives to the staff involved in attaining the target. Audits are done on a quarterly basis and it takes up to 2 months after the end of quarter to receive the bonus payment	There is a strong CCT programme as well as health insurance programmes which cover different pockets of population (public employees, school aged children, etc)	This is a descriptive case study focusing on the design of the P4P scheme. It does not evaluate the utilization effect or health outcomes due to P4P. However, given the accompanying components, it would be difficult to tease out the effects of supply side incentives only.	El-Saharty et al. 2010
Pay for Performance Pilot in Zambia	Payments made to facilities but Cordaid specified that 40-60% of payments could go to staff. Rest for infrastructure, drugs, supplies, and running costs (with guidelines for allocation to different categories). Maximum of 20% on top to district health office for management	Base payments of Zambian Kwacha (ZMK) 90,000,000 as fixed funding at the start of the year, with a similar amount available to each hospital if 100% of the targets were achieved. For health centres the amount was ZMK 50,000,000 each. Bonus paid every 6 months.	Independent of this project, in kind demand-side incentives offered in some places such as mother kits. 15,000 Euros given to HCs and 30,000 Euros to hospitals to allow them to rehabilitate or buy supplies at start of project. Authors note need for more technical support and capacity building in implementing project.	Cordaid review of its programme. Based on interviews and the collection and analysis of health and financial data. Semi-structured interviews with government officials at national, provincial and district level and donors such as the World Bank and CHAZ. The set up and effects of PBF were studied in four PBF supported mission hospitals (St. Paul's, Kasaba, Lubwe and Minga) and one Rural Health Centre (Muzeyi) Discussions, interviews and data collected in Mbereshi and Petauke hospital as well as Chiparamba Rural Health Centre, were used for comparison with the PBF supported facilities. - This study explored health data from 2004-2007. The evaluators have not been able to compare this to conditions prior to PBF implementation and it can thus not be determined whether changes are due to PBF or have always been the case. Too early for an impact evaluation.	Vergeer & Collins 2008

Programme	Who received P4P payments?	Payment methods	Other accompanying components	Summary of evaluation/study design	Authors
				Quality of evidence	
Performance-based Grants for Reproductive Health in the Philippines (PBG1)	PBG 1 directed at local health facilities (PBG 2 discussed below)	No information given	Another variant is PBG 2 (see below)	WHO policy brief based on a rapid assessment.	WHO 2011
PAYMENTS TO HEALTH CARE WORKERS					
Pay for Performance in Tanzania	Health facilities receive bonuses which are supposed to be divided among health team members equally, with each worker having the ability to earn up to of 200,000 Tanzanian shillings (Tsh) annually. At the hospital level payments are made to directly to the hospital with autonomy to allocate as long as staff is consulted.	Bonuses are channeled from the government through the Council Medical Officer of Health Accounts to facility bank accounts opened at the dispensary and health centre level. Council Health Management Team (CHMT) bonuses are linked to the facilities they manage. CHMTs report to Regional Health Management Team (RHMT). In order to qualify for a bonus, CHMTs and RHMTs must complete their HMIS reporting in a timely manner. CHMTs qualify for 50% of their bonus when 50% or more of their health facilities reach their targets. Another 50% of the payment comes from timely reporting to the RHMTs. RHMTs are paid 50% of their payment for their timely reporting in HMIS and 50% payment for performance for 50% or more of health facilities in the region meeting their targets. Payment is annual.	None mentioned	This is a case study based on an assessment of existing documentation and interviews. A number of problems with design and implementation are noted. For example, the verification of performance was carried out by district and regional supervisors, whose own bonuses depended on the performance of the facilities, causing a conflict of interest. Communication of the policy to district teams was also poor	Morgan and Eichler 2009
Performance-based Grants for Reproductive Health in the Philippines (PBG2)	PBG 2 aimed at TBAs, CHWs, and midwives (the women's health teams). PBG 1 (discussed above)	PBG 2 payments vary: Under one method (e.g. Sorsogon Province) Peso 1000 (US\$ 22) received by local women's health team for every facility based birth. Division of payment among the TBA (60%), midwife (20%) and CHW (20%); under	As mentioned PBG 2 has a demand side incentive for the pregnant women worth Peso 500. Another variant is PBG 1 directed at facilities	WHO policy brief based on a rapid assessment.	WHO 2011

Programme	Who received P4P payments?	Payment methods	Other accompanying components	Summary of evaluation/study design	Authors
				Quality of evidence	
		another method (Surigao del Sur), the Peso 1000 is divided in a 60:40 ratio among the women health team and the health facility team: delivering doctor (10%), attending nurse (10%), health facility (20%). Each member of the Women Health team receives 20% of the Peso 600.			
Cambodia Midwifery Allowance	Payment to midwives by the government: US\$ 15 per live birth in a health center or health post; US\$ 10 per delivery in a regional or national hospital. Payments also provided to Village Health Support Groups and TBAs for referring women to facilities		Health equity funds and community based health insurance on the demand side to encourage utilization of services, especially among the poor	Case study data; not an impact evaluation	Hawkins 2011; Murakami 2009
Battagram P4P project, Pakistan, Save the Children	Health workers employed by the government in public health facilities were provided bonuses by Save the Children. Districts staff and staff working on vertical programmes also received bonuses, although these were fixed (35% of salary).	Payments were made directly to the health workers' bank accounts by Save the Children	Save the Children was managing the health facilities at the district level, of which PBI was a small part. The district health budget from the provincial government was also channelled through the NGO, which had the mandate to (i) organise and manage the healthcare services; (ii) procure and supply medicines; (iii) implement HMIS; and (iv) monitor and supervise the health system.	Mix of qualitative and quantitative methods employed. The district was divided into 4 hubs under the project. For this analysis, health facilities were chosen from each hub, with an additional stratification against performance: very good, good, satisfactory and poor - one under each category from each hub. Quantitative analysis was conducted using HMIS data, financial records, monthly progress reports, supervisory and performance scores of facilities, and project documents from 2007 to mid-2010. Qualitative data was collected through 11 key informant interviews with stakeholders (Save the Children, World Bank, provincial and district offices, and one local association). At the facility level, in-depth interviews were held with 7 managers and other staff working at 4 facilities (three basic health units and one rural health centre). In addition, 11 focus group discussions were held with staff (male and female) and community members (male and female).	Witter et al. 2011

Programme	Who received P4P payments?	Payment methods	Other accompanying components	Summary of evaluation/study design	Authors
				Quality of evidence	
Safe Delivery Incentive Program, Nepal	Health workers who could be medical doctor, staff nurse, auxiliary nurse midwife, health assistant, auxiliary health worker or maternal and child health worker. The deliveries have to be either at a public facility or at the woman's home	Funds are provided to health institutions through the District Public Health Office.	Strong CCT component aimed at women and health facilities were reimbursed the cost of the facility-based delivery. 48% of the budget for the SDIP went to mothers; 47% to service providers and 5% to health institutions themselves.	Evaluation was conducted using data from 10 districts. Two types of evaluations were conducted: process and impact, both of which employed a mix of quantitative and qualitative analysis. Data was collected through surveys of health institutions and women (50 health institutions were surveyed); HMIS, SDIP and Emergency Obstetrics Care monitoring systems and the Mother and Infant Research Activities (MIRA) community surveillance system. Data analysis was conducted using regression analysis, interrupted time series and propensity score matching. In addition, in-depth interviews and focus group discussions were also conducted at the community level. The evaluation of the programme was conducted for the Government of Nepal. Most of the results relate to the demand-side payments and so are not reported here. The programme was amalgamated into the Aama programme (including nation-wide fee exemption) from 2009 onward.	Powell-Jackson et al. 2008; Ensor et al. 2009
Janani Soraksha Yojana	At the district level payments made to ASHA or community health workers with accreditation with the trigger being facility based delivery. Payment is Rs. 600 for rural areas and Rs. 200 for urban areas per institutional delivery.	This is a Central Government-supported scheme. Payment is made from government to service providers through State Governments. State/District level: State prepares state and district budget based on costs for JSY which is funded by the national government. The Ministry of Health and Family Welfare provides funds to the State Health Society which is responsible for its disbursement to the districts. CHW Payments: ASHA are paid directly by either the Medical Officers at PHCs after they bring women to facility for delivery or directly by the ANMs.	JSY is mainly a CCT scheme with a PBF component. CCTS are provided to women for institutional deliveries, ANC and PNC visits.	Uses two rounds of DLHS (2002-04 and 2007-09) - nationwide district level data; methods used include exact matching, with vs without analysis and D in D with logistic regression with state and district fixed effects. Peer reviewed journal article (Lim et al. 2010). Assessment is based on mixed methods; quantitative data collected from 1200 mothers (delivered in the past 1 year) from rural areas; questionnaire for community leaders, checklist questionnaires for CHW, ASHAs, ANMs, medical officers of PHCs, CHCs, and government hospitals (n=50); in depth interviews with district hospitals, state and district nodal officers. (UNFPA India)	Lim et al 2010; UNFPA 2009 (assessment of JSY in 5 States)

Table 3A: Impact of P4P schemes

Name of programme	EVIDENCE OF IMPACT ON:				Authors
	Utilisation of services/coverage of Programmes	Quality of care	Health outcomes	Organization of care/range of services	
PAYMENTS TO GOVERNMENTS					
Global Fund to fight AIDS, Tuberculosis and Malaria	Significant increase in the number of sites delivering HIV interventions since 2004: in most countries sites providing testing and counselling has doubled between 2004 and 2007; PMTCT is now offered in at least 1/4th of health facilities in all countries, even though the number of sites is below one per 1,000 pregnant women in all countries, except Zambia (2.2 per 1,000).	Sacrifice of quality to achieve target outputs reported in some countries	Estimated that in the 14 countries with a generalized epidemic over 570,000 life years were added from the use of ART between 2003 and 2007. In the same period, it was estimated that the number of infections averted due to PMTCT amounted to over 16,000.	Not analysed, though the additional funds have expanded the range of services available in many countries	Macro International 2009
PAYMENTS TO MID LEVEL ORGANIZATIONS					
Plan Nacer in Argentina	No information given	Although not mentioned, generally indicators have shown improvement	No information given	No information given	Niamoli & Vergeer 2010; Eichler & Glassman 2008
Output based Payments in Haiti	Panel regressions show that (a) immunization coverage increased between 13 and 24 percentage points; (b) births by SBA increased from 17 to 27 percentage points. Changes in utilization of ANC and post-natal care were not significant.	Not measured empirically - anecdotal evidence from NGOs that emphasis was skewed towards meeting targets versus quality	No information given	Within the NGOs, there is evidence of capacity being built, but no discussion of wider impact on organisation of services/health system	Eichler et al. 2007

Name of programme	EVIDENCE OF IMPACT ON:				Authors
	Utilisation of services/coverage of Programmes	Quality of care	Health outcomes	Organization of care/range of services	
Contracting of Health Services Pilot Project, Cambodia	Contracting in caused a 36 percentage point increase in ANC visits (statistically significant); contracting out caused a significant increase in vitamin A receipt in children (42 percentage points); contracting in and out both lead to significant increase in use of public health services for curative care consultation (18 and 29 percentage points), and in facility based deliveries (18 and 30 percentage points). Immunizations also increased substantially but were only significant under contracting in. Care seeking behavior: Household members were about 3.6 and 5.4 percentage points more likely to consult a public provider under contracting-in and contracting-out, respectively, compared to a comparison group baseline of 0.6%. This effect is statistically significant.	Significant impact on probability of scheduled staff being present. This increased by 50 percentage points for contracting in and 79 percentage points for contracting out. Significant increase in supervision visits from 2.5 to 5.7 over past 3 months. This was not significant for contracting in. Overall, greater improvement in management of services under contracting out compared to contracting in. Perceived quality of care by users was negative for both variants, but not statistically significant.	Contracting out reduced the chance of an individual reporting they were sick in the past month (significant at 5% under randomization inference) and reduced the incidence of diarrhea in children under 5 (significant at 10% under randomization inference).	With contracting-in, more likely to have round the clock services available at health centres (increase of 83 percentage points is statistically significant). The contracting-out effect is 47 percentage points but not significant. No significant impact on providing delivery services due to contracting in or out	Bloom et al. 2007
Performance-based contracting in the DRC	In the covered areas, outpatient consultations have increased from 0.06 per capital at baseline to 0.30 in 2007. Measles immunization coverage increased from 25% to 92%; assisted deliveries from 25 % to 74 %.	No information given	No information given	No information given	Johannes et al. 2008
Paying NGOs for performance in Afghanistan	Overall coverage of services increased during 2003-06 in all 34 provinces. Proportion of facility to population improved from 1 to 34,000 to 1 to 20,000 during this period. Increase in immunizations (DPT3); facility based deliveries have doubled; 300% increase in female health workers in 8 provinces; ANC visits increased from 45 to 75% in the 3 MoPH provinces	No information given	No information given	No information given	Sondorp et al. 2009
Performance based contracting in Liberia	Some utilization results presented: 81% increase in facility-based deliveries; 52% increase in couple-years of family planning protection; 134 % increase in pregnant women receiving a second dose of intermittent preventive treatment of malaria (IPT2)	No information given	No information given	No information given	Morgan 2011

Name of programme	EVIDENCE OF IMPACT ON:				Authors
	Utilisation of services/coverage of Programmes	Quality of care	Health outcomes	Organization of care/range of services	
PAYMENTS TO FACILITIES					
Pay for Performance in Rwanda	Compared to control group: a 23% increase in the number of institutional deliveries; 56% increase in no. of preventative visits for children under 24 months; 132% increase in no. of visits for children 24 to 59 months; no impact on women completing 4 ANC visits or child immunization (Basinga et al. 2011)	An increase of 0.157 SD (95%CI: 0.026-0.289) in prenatal quality as measured by compliance with Rwandan prenatal care clinical practice guidelines.	Not measured in this study.	No information given	Basinga et al 2011; Basinga et al. 2010; Sekabaraga et al 2011; Kalk et al 2010
Performance-based Financing in the Democratic Republic of Congo (DRC)	Significant improvement compared to control group for knowledge of HIV/AIDS and use of modern health facilities or pharmacy (difference of 10 percentage points each, significant at 5% and 10 % respectively). Other indicators, including use of family planning, showed improvement, but not significant compared to control group	Perceived quality of care saw improvement compared to control group: availability of drugs (difference of 37 percentage points between treatment and control group), quality (15 percentage points difference), respect for patients (12 percentage points difference); the professionally determined performance indicators for the health centre was also significantly better for treatment vs. control group (65% vs. 39%) as was the indicator for qualified personnel (65% vs. 54%).	No information given	No information given	Soeters et al. 2011
Pay for Performance in Burundi	Evaluation results of pilot show increase in uptake of services, higher child immunizations and facility based deliveries by 50 to 60% compared to baseline.	Quality of services improved	No information given	Providers develop their own business plans as well as contract with individual health workers, and other input providers, which has lead to greater clarity on expectations and roles.	Busogoro & Beith 2010; Toonen et al. 2009
Pay for Performance in Tanzania	No difference observed in P4P and non P4P utilization rates.	Inconclusive	No information given	No impact on range of services recorded.	Canavan & Swai 2008

Name of programme	EVIDENCE OF IMPACT ON:				Authors
	Utilisation of services/coverage of Programmes	Quality of care	Health outcomes	Organization of care/range of services	
Performance based contracting in Uganda	No gain for PBF group compared to control groups. 22 of 23 facilities met at least one of 3 chosen targets. Group B (greater autonomy) performed better.	Based on exit polls, data show deterioration of quality of services (attitude of staff) among the bonus group.	No information given	Based on exit polls, availability of medicines worsened in the bonus group	Morgan 2010
Pay for Performance Pilot in Zambia	Limited improvements in access to curative care, in both the PBF and non-PBF facilities	No information given	No information given	No information given	Collins & Vergeer 2008
Performance-based Grants for Reproductive Health in the Philippines (PBG1)	General improvement reported in facility based births. For example, in Infugao, facility based births increased from 31% to 50% between 2007 and 2008.	No information given	No information given	No information given	WHO 2011
PAYMENTS TO HEALTH CARE WORKERS					
Performance-based Grants for Reproductive Health in the Philippines (PBG2)	General improvement in facility based births reported. However, specific effects of PBG 2on utilization not presented.	No information given	No information given	No information given	WHO 2011

Table 4B: Impact of P4P schemes (continued)

Name of programme	EVIDENCE OF IMPACT ON:				Authors
	Patient payments	Equity	Provider motivation	Unintended consequences/ comments on results	
PAYMENTS TO GOVERNMENTS					
Global Fund to fight AIDS, Tuberculosis and Malaria	No information given	No evidence of changes in coverage differences between disadvantaged groups and those who are better off. All countries would benefit from expanding services to make services more accessible to people in underserved areas. Although gender equity is a guiding principle, only 44-55% of countries had indicators to measure this. The Monitoring & Evaluation Toolkit does not yet include gender-specific indicators and inclusion would benefit measurement.	No information given	Not discussed here, though a number of studies have examined the impact of the GFATM on health systems, with mixed results.	Macro International 2009
PAYMENTS TO MID LEVEL ORGANIZATIONS					
Plan Nacer in Argentina	No information given	Not analysed, but the focus of the scheme is on enrollment of poor women and children	No information given	No information given	Niamoli & Vergeer 2010; Eichler & Glassman 2008
Output based Payments in Haiti	No information given	No information given	Overall satisfaction - system is better than previous reimbursement-based one in terms of workload for NGOs. Anecdotal evidence presented in the paper (feedback from NGOS) suggests too much "stress" over meeting institutional targets, mainly from having to make organizational changes and for individuals in the organization having to change the way they work. Loss of income if a target is not met, is also frustrating for the NGOs, especially when targets may not be met due to other actors (e.g. not receiving supplies in time from another vendor).	No information given	Eichler et al. 2007

Name of programme	EVIDENCE OF IMPACT ON:				Authors
	Patient payments	Equity	Provider motivation	Unintended consequences/ comments on results	
Contracting of Health Services Pilot Project, Cambodia	No significant impact recorded	No information given	No information given	Contracting in had a 21 percentage point improvement in knowledge of AIDS risk factors (statistically significant). In aggregate, the effects were not statistically significant for non-targeted indicators. These include treatment of diarrhea in children, the number of antenatal services (excluding a blood pressure check, which was targeted), whether individuals report that an outreach team has visited the village in the previous four weeks, whether a mother breastfed a newborn within six hours of birth, whether a mother gave a newborn water in the first month of life, and knowledge of AIDS risk factors.	Bloom et al. 2007
Performance-based contracting in the DRC	The reported average consultation fee in project areas has declined from US\$4 to US\$2.	No information given	No information given	No information given	Johannes et al. 2008
PAYMENTS TO FACILITIES					

Name of programme	EVIDENCE OF IMPACT ON:				Authors
	Patient payments	Equity	Provider motivation	Unintended consequences/ comments on results	
Pay for Performance in Rwanda	Patients covered simultaneously under health insurance (mutuelles), which helped lower their costs: Slight decrease in direct health spending in real terms during 2000-05: Adjusted median spending per episode declined from RwF 555 to RwF 419. The decline happened for all socio-economic groups. For the poorest, median expenditures per episode dropped from RwF 348 to RwF 281. Payments decreased for all categories except transport. Payment at the point of delivery decreased by 35% for consultations, and by 30% for drugs.	General change over period (not PBF-specific): utilization of services increased from 10.7% to 17.4% among the poorest quintile. Assisted birth deliveries increased from 12.1% to 42.7% , the largest increase for assisted deliveries. Under 5 utilization of modern health services doubled from 7% to 18 %. Distance to facility was seen as lowering access	Facilities provided funding based on performance against indicators. Overall satisfaction with payments. However, interviews suggest some problems at individual provider level, which may have consequences for quality of care: "50% of health workers considered P4P a control mechanism than as a supportive system. Only 24% believed that P4P had improved the management, and 32% did not think it was useful. In addition, 64% of staff felt that management support to their professional, personal and psychological needs was insufficient. With only 1/3rd of all positions within the Rwandan health sector actually filled, the P4P approach was frequently described as putting additional stress on a system already overstretched. 72% per cent of medical staff reported to regularly work supplementary hours and to feel constantly tired because of the workload" (Kalk et al. 2010). The caveat in these qualitative results is the small sample which is restricted to one district hospital.	In the qualitative work, two phenomena described: (1) Gaming of the system, with providers focusing on producing results against indicators as opposed to providing needed care. Interviews seem to suggest that long term care is often neglected, and other RH related issues such as morbidity ignored, because they are not part of the incentives/indicators package. (2) Crowding out, as intrinsic motivation to help others is replaced by extrinsic motivation associated with incentives. Also, intermediate indicators may not necessarily be appropriate for the outcome measured.	Basinga et al 2011; Basinga et al. 2010; Sekabaraga et al 2011; Kalk et al 2010
Performance-based Financing in the Democratic Republic of Congo (DRC)	Patient out of pocket payments increased 45% in treatment group between 2005 and 2008. The annual per capita revenues from patient user fees increased by 25 percent between 2005 and 2008	The increase in health spending did not affect the poorest 25% of the households in the participating districts. HH survey shows that health spending in this group declined by 14%, while that of the relatively wealthy proportion of the sample increased.	Not presented	Not presented	Soeters et al. 2011

Name of programme	EVIDENCE OF IMPACT ON:				Authors
	Patient payments	Equity	Provider motivation	Unintended consequences/ comments on results	
Pay for Performance in Burundi	No information given	No information given	Pilot results indicate that staffs were retained in PBF facilities, and staffs from other areas were migrating into the health facilities in PBF zones, attracted by the incentives. Involvement in developing business plans for health facilities also seems to have had a positive effect on provider motivation	No information given	Busogoro & Beith 2010; Toonen et al. 2009
Pay for Performance in Tanzania	No information given	No information given	The fact that the contracts were signed between Cordaid and the diocese, albeit maintaining the purchaser-provider split essential for PBF, proved to be a major disadvantage in instilling responsibility for results and ownership of the performance indicators at health facility level who were often not involved or aware of contract negotiations and agreements but responsible for its results. The health facilities managers highlighted dissatisfaction with several of the indicators selected by Cordaid and the corresponding targets set. Overall staff satisfaction was found to be similar across PBF and non-PBF facilities. Intrinsic motivation factors emerged as most important	P4P related drugs are dispensed/managed differently or sparingly to avoid stock-outs. Authors comment that it may cause focus on curative care, as there were no preventive indicators or indicators linked to quality of care	Canavan & Swai 2008
Performance based contracting in Uganda	Based on exit polls, payments were perceived to increase (worsen) in the bonus group compared to control	The wealth index of clients treated by the PNFP bonus group increased indicating that they were serving wealthier clients compared to the control group	Autonomy in financial decision making appears to have had a positive influence on service provision.	The experiment was unable to improve performance. This may have been due to design of scheme (small bonus size), poor information management systems which were not improved, and a movement from PNFPs to public sector of health workers which affected the capacity of PNFP missions (external environment influences)	Morgan 2010

Name of programme	EVIDENCE OF IMPACT ON:				Authors
	Patient payments	Equity	Provider motivation	Unintended consequences/ comments on results	
Performance based Contracting Pilot, Nicaragua	No information given	No information given	No information given	There is no incentive to perform better once the 100% threshold for bonus is reached.	Jack 2003
Pay for Performance Pilot in Zambia	No information given	Not established, but authors note that no emphasis within PBF scheme on pro-poor measures (e.g. no focus on remote areas, which are more in need)	The expenditure ceilings set by Cordaid (40-60% for staff motivation, 20-30% for equipment and medical supplies, 20-30% for infrastructure and 10-30% for running costs) were found to confine health managers autonomy in decision making to improve performance. Increase in workload is not accompanied with an increase in staffing.	Increasing the inpatient turnover rate especially appeared questionable for health centres which are to focus mainly on preventative and promotive health care; managers felt it led to a neglect of PHC. TB detection rates were investigated to see if they fell in PBF areas - there was no clear pattern. Various confounding factors, such as higher revenues in mission sector and removal of user fees in 2006.	Collins & Vergeer 2008
Performance-based Grants for Reproductive Health in the Philippines (PBG1)	No information given	No information given	No information given	These are not consequences, but design and communication issues: health workers under PBG 1 were not aware of the bonus in many cases. PBG 1 also had issues with delay in release of funds.	WHO 2011
PAYMENTS TO HEALTH CARE WORKERS					
Performance-based Grants for Reproductive Health in the Philippines (PBG2)	No information given	No information given	PBG 2 incentive of Peso 200 not enough to encourage TBAs to send women to facilities (they can earn between 1000 and 1500 pesos for home births). Providers and mothers also reported delays in payments	No information given	WHO 2011

Table 5 Cost and sustainability of P4P schemes

Programme	Overall cost of scheme	Assessment of cost-efficiency	Sustainability	Authors
PAYMENTS TO GOVERNMENTS				
Global Fund to fight AIDS, Tuberculosis and Malaria	Initiated in 2001 with US\$ 1 billion in pledges. The Global Fund is estimated to have contributed to the rapid expansion of programming addressing HIV/AIDS, tuberculosis, and malaria through more than 550 grants, and it is estimated that significant amounts of its grants are allocated to key health systems elements (35% of about US\$4 billion of approved financing by 2008)	In the 18 evaluation countries for HIV, it was estimated that HIV funding increased rapidly with 18% of the funding coming from the Global Fund. The evaluation states that "increased funding has led to better access to care, including rapid increases in intervention uptake and notable survival benefits through ARV treatment".	The GFATM has received significant support from donors and is expected to be sustained in the near future. However, a fall in pledges in 2011 reflects the difficult international economic climate.	Macro International 2009
GAVI HSS Window	Since inception in 2005 US\$ 524 million committed to 44 countries and disbursed US\$ 255 million to 36 countries.	No information given	GAVI is a global alliance of donors including the World Bank, UNICEF, WHO, Gates Foundation and others. Its main focus is on immunizations and vaccinations. HSS was introduced as a separate window of funding in 2005 and has seen rapid scale up. There is support for the alliance to continue its functions since the additional funding has been useful in supporting countries for providing better services related to immunizations and vaccines.	HLSP 2009; www.gavialliance.org ; Chee 2007
PAYMENTS TO MID LEVEL ORGANIZATIONS				
Plan Nacer in Argentina	Not mentioned. However, payment to providers is costed at \$10 per person per month.	No information given	This is a national programme, funded by the World Bank. Continuity of funding from the Bank is contingent on meeting specific criteria in the implementation of Plan Nacer.	Niamoli & Vergeer 2010; Eichler & Glassman 2008

Programme	Overall cost of scheme	Assessment of cost-efficiency	Sustainability	Authors
Contracting of Health Services Pilot Project, Cambodia	Total spending was \$2.56 per capita in contracting in districts, 61% higher than the \$1.59 per capita spent in comparison districts. Contracting out districts spent \$2.94 per capita, 85% higher than comparison (significant difference). There are no significant changes in total health spending.	No information given	This was a pilot programme, initiated in mid 1999 and lasted through 2003	Bloom et al. 2007
Performance-based contracting in the DRC	US\$ 150 million provided through IDA for this phase (2008 onwards, when project expanded to 89 zones)	No information given	Ongoing, but sustainability not assessed	Johannes et al. 2008
Paying NGOs for performance in Afghanistan	US\$ 155 million (World Bank, USAID and European commission combined)	No information given	The Government is committed to an output-based approach and the three main actors - World Bank, USAID, and European Commission - are interested in continuing with it for the foreseeable future.	Sondorp et al. 2009
PAYMENTS TO FACILITIES				
Pay for Performance in Rwanda	The P4P budget in Rwanda grew from US\$ 200,000 in 2002 to US\$ 9.3 million in 2007. Transfers from the centre to the districts for provision of health services increased from Rwf 1.3 billion in 2005, to Rwf 7.1 billion in 2006 and Rwf 9.7 billion in 2007. In 2007, the Treasury transferred US\$ 1.8 per capita for provision of basic health services to the districts.	No information given	The first pilots were launched in 2002. By 2005, the Government of Rwanda had committed to scale up, which began nationally in 2005-06. There is considerable buy-in to the programme from the Government and donors. The programme is funded through the World Bank's RBF programme, GFATM and bilateral donors as well as the government's budget. "In 2005, donor per capita expenditures were estimated at US\$15 out of a total of US\$34, about half of which came through earmarked financing for HIV/AIDS (World Bank and MOH Rwanda 2010) and the other half through budget support".	Basinga et al 2011; Basinga et al. 2010; Sekabaraga et al 2011; Kalk et al 2010
PAYMENTS TO FACILITIES				

Programme	Overall cost of scheme	Assessment of cost-efficiency	Sustainability	Authors
Performance-based Financing in the Democratic Republic of Congo (DRC)	Budget was \$2 per person per year, plus \$0.4 for administration and overheads. No information on overall expenditure was given. However, for the treatment group the average monthly revenue per capita was calculated at \$1.04 compared to \$0.45 in the control health centres.	No information given	There is reported to be strong buy-in from the Government. However, this was externally funded, and managed by CORDAID.	Soeters et al. 2011; Soeters & Kiwanuka, 2008
Pay for Performance in Burundi	Overall cost not provided. However, the GoB, with support from the World Bank, will cover US\$ 67 million for the programme.	No information given	Strong Government commitment for scale up. Donor support also has provided a boost to the programme (pilot and then scale up), with the Dutch government and the EU having supported HealthNet TPO and Cordaid (pilot on P4P). During scale up, others such as the Swiss and Belgian Development Cooperation Agencies, GAVI Alliance, and the World Bank have become increasingly involved as financiers. It is likely that the programme will be sustained	Busogoro & Beith 2010
Pay for Performance in Tanzania (CORDAID)	Budget set at 0.5 Euros per capita; amounted to 8% of facility income on average.	No information given	Externally funded pilot. However, there is strong support for performance based financing in the health sector, especially with the aim of achieving the MDGs 4 and 5.	Canavan & Swai 2008
Performance based contracting in Uganda	US\$ 300,000 total budget	No information given	The programme was supported by World Bank, CIDA, USAID. Results were not positive because of which interest has waned. There is work underway to revive interest in the intervention.	Morgan 2010
Performance based Contracting Pilot, Nicaragua	No information given	No information given	This was a pilot - however generally there is support for this. Both the World Bank and IADB support the country's Health Sector Strategy which aims to improve administration, financing and management.	Jack 2003
Pay for Performance Pilot in Zambia	Budget set at 0.5 Euros per capita. Amounted to 17% of facility revenue. Amounts small relative to salaries (but variable by facility).	No information given	Pilot project - not scaled up.	Collins & Vergeer 2008

Programme	Overall cost of scheme	Assessment of cost-efficiency	Sustainability	Authors
Performance-based Grants for Reproductive Health in the Philippines (PBG1)	No information given	No information given	PBG has government interest for learning from and improving interventions like it. How this materialises remains to be seen.	WHO 2011
PAYMENTS TO HEALTH CARE WORKERS				
Pay for Performance, Tanzania	US\$ 29 million in the first year budgeted for performance incentives	No information given	The programme has government interest, but the initial attempt could not be sustained: the programme was launched nationally, but there were implementation problems: (i) government did not want to roll out the programme gradually, instead implemented it simultaneously across the country; (ii) there was not adequate time for training and communicating the programme to the facilities and health workers; (iii) donors did not have confidence in the programme and strongly favored a roll out and refused to pay for health workers' bonuses. As of 2010, there was still strong government interest in the programme, and it had initiated new discussions with donors on how to package and roll out the programme.	Morgan & Eichler, 2009
Performance-based Grants for Reproductive Health in the Philippines (PBG2)	No information given	No information given	PBG has government interest for learning from and improving interventions like it.	WHO 2011