report

DFID Department for International Development

WORKING PAPER 39
January 2013

EVALUATION OF PAYMENT BY RESULTS (PBR): CURRENT APPROACHES, FUTURE NEEDS

Report of a study commissioned by the Department for International Development

Evaluation of Payment by Results (PBR):

Current Approaches, Future Needs

Final Report

Burt Perrin
Independent consultant

January 2013

Table of Contents

Executive Summary

Background and purpose

S1. Payment by results (PBR) is a form of aid financing that makes payments contingent on independent verification of results. PBR is part of a wider UK government agenda and several other government departments are piloting this approach.

S2. PBR as defined by the Department for International Development (DFID) has three key elements:

- Payments based on results;
- Recipient discretion – i.e. the recipient has space to decide *how* results are achieved; and
- Verification of results as the trigger for disbursement.

S3. It is important to recognise, however, that there is no consistent use of terminology, and various forms of PBR frequently go under other monikers.

S4. This study had three main objectives:

- To identify and synthesise evidence, to the extent possible, from evaluations of PBR approaches in development.
- To provide an analytical critique of the quality of existing evaluations.
- To provide guidance for approaches, including evaluation questions and methods, to future evaluations of PBR programmes.

Evidence base for PBR approaches

S5. With a very few exceptions, almost all research and evaluation studies of PBR have been in the health sector. Almost all the studies are of Results-Based Finance (RBF) initiatives (incentives to service provider organisations and individuals) rather than of Results-Based Aid (RBA) to governments.

S6. The importance of an outcome (or results) orientation, focusing on the actual benefits arising rather than on inputs and services provided, is largely uncontested. Nevertheless, the evidence regarding the potential of incentives to change professional practice is weak.

S7. Perhaps the most optimistic conclusion that can be drawn from available evidence is that contracting out may increase access and use of health services in the short term rather than broader health outcomes. Unintended effects are quite possible, and there is limited evidence to date to date that PBR approaches offer value-added compared to other modalities.

S8. Actual implementation of PBR approaches has encountered significant challenges and difficulties. There has been limited attention to some basic questions about PBR approaches, including the mechanisms by which incentives may work or not, cost effectiveness, comparison with other potential approaches, impact on equity, and sustainability.

S9. What does emerge strongly from the evidence base is that PBR needs to be implemented as part of a package that includes other forms of supports and services. The underlying complexity of each intervention presents a serious challenge to implementation and evaluation, inhibiting meaningful generalisation without identification of the specific mechanisms at play.

Quality of research and evaluation studies on PBR

S10. Practically all the "evaluations" identified may be more appropriately described as research, carried out by people who identify themselves as researchers. There is limited awareness or attention in these studies to generally accepted evaluation standards and quality criteria, including the Organisation for Economic Co-operation and Development (OECD)/Development Assistance Committee (DAC) Evaluation Criteria and the Evaluation Quality Standards, to which DFID along with other bilateral aid agencies have agreed. In particular, there has been limited attention to the five DAC evaluation criteria, except to objectives achievement (effectiveness).

S11. The research quality of studies that have been identified has been considered poor by major systemic reviews and critiques. It has proven to be very difficult in practice to implement and to maintain the fidelity of sophisticated experimental or quasi-experimental designs, resulting in biases that the reviews say compromise the ability to draw conclusions from the data obtained. There are, however, alternative evaluation approaches that are particularly suited to evaluation of complex situations involving multiple factors, and where it may be more meaningful to document "contribution" rather than linear cause-and-effect.

S12. Research designs used to date generally are very weak regarding external validity (or generalizability), and have given limited consideration to the ability to generalise or explain how findings can be applied in other situations or contexts. In order to be able to apply or adapt findings from one setting or situation to another, it is essential to be able to understand the mechanisms, the 'hows' and 'whys' through which a PBR approach has (or has not) resulted in changes. But to date, there has been limited attention to this consideration.

Guidance for future evaluations of PBR

S13. A further objective, and the most important priority identified for this study, was to provide ideas for future evaluation of PBR.

S14. *Questions for evaluation*. A basic principle for meaningful evaluation is to start first, before considering potential methodological options, by identifying *the questions* that need to be addressed. This is a key principle highlighted throughout the evaluation literature, and in the DAC *Standards* for evaluation.

S15. Arguably the most important question for evaluation is to identify the mechanisms and sets of circumstances under which PBR approaches can make a positive difference. There are a range of other important questions about PBR identified in the text, such as cost effectiveness and comparison with other potential approaches and strategies, appropriate size and nature of incentives, and exploration of unintended effects and how these can be minimized.

S16. There is a particular need for evaluation to explore and to describe the *process* by which PBR initiatives are implemented in practice, and the reasons why changes from the original conception may be needed.

S17. To date, there has been scant evaluation attention to sectors other than health, and hardly any evaluations of RBA initiatives with incentives aimed at governments directly. However, DFID is in the process of initiating pilots to explore these considerations.

S18. *Methodological implications and alternatives.* There is no single method that is "best" or that should be given preference. In general, evaluation approaches should be used that can best inform policy and programming decisions on a timely and cost-effective basis. A mixed method approach should be utilised. In all cases, articulation of the theory of change can aid in identifying evaluability, indicating what types of questions can be evaluated at given points in time, and serving as a basis for choosing the most appropriate evaluation design. Given the complex context in which PBR schemes work, always in combination with other factors, it may be more appropriate to use a contribution analysis approach rather than linear cause-and-effect.

S19. There are significant opportunities for theory-based approaches to evaluation that can identify and document the mechanisms at play. In particular, a realist evaluation approach that seeks to identify what works for whom in what circumstances, may be particularly suited to evaluation of PBR schemes.

Recommendations

S20. Implications and recommendations for policy and programme

- There is a need for some healthy scepticism, with recognition that the value of PBR is, at least as of yet, unproved.
- In common with all aid modalities, one should embark upon a PBR approach only after considering its potential impact and cost effectiveness in comparison to other possible strategies.
- Potential unintended effects are likely. They should be anticipated and articulated at the design stage, and monitored on an ongoing basis.
- Each application of PBR should be tailored to the particular situation, recognising that one model is unlikely to be appropriate equally across the board.
- Consideration should be given to the meaning of a 'hands off' approach. At a minimum, it is appropriate to insist upon adherence to ethical guidelines and standards. A hands-off approach should not be seen as a barrier to independent evaluation.
- Given the challenges identified to implementation of PBR, programmes should include an internal M&E capability.

S21. Implications and recommendations for evaluation

- Evaluation should start by identifying priority questions that can best inform policy and programming decisions, and only then consider potential methodologies that can best address these questions in a timely and cost-effective manner.
- The most important question for evaluation to address should not be "does it work?" but, rather, should be to "identify the mechanisms and sets of circumstances under which PBR approaches can most likely result in behavioural change leading to long-term impact".
- Evaluation should explore unintended consequences of incentive approaches, identifying when these are most likely to occur and when these may offset expected benefits.
- Other potential evaluation questions that should be considered include (but are not limited to): cost effectiveness and comparison with other potential approaches and strategies, appropriate size and nature of incentives, sustainability, and equity.
- There is a particular need for evaluation to explore and to describe the process by which PBR initiatives are implemented in practice, and the reasons why changes from the original conception may be needed.
- A mixed method approach should be taken, involving both quantitative and qualitative methods. In all cases, the theory of change should be articulated.
- Priority should be given to methods that can provide explanation. In this regard, theory based and in particular realist evaluation approaches should be given special consideration.
- Given the complex context in which PBR schemes work, always in combination with other factors, it may be more appropriate to use a contribution analysis approach rather than aim, unrealistically, to identify linear cause-and-effect.

Abbreviations

CCTs	Conditional Cash Transfers
DAC	Development Assistance Committee
DFID	Department for International Development
GAVI	Global Alliance for Vaccines and Immunisation
NGOs	Non-Governmental Organisations
Norad	Norwegian Agency for Development Cooperation
OECD	Organisation for Economic Co-operation and Development
PBR	Payment by Results
RBA	Results-Based Aid
RBF	Results-Based Finance
RCT	Randomised Control Trials

1 Introduction

1.1 Background and purpose

1.1 The purpose of this study was to identify and synthesise evidence, to the extent possible, from evaluations of Payment by Results (PBR) programmes in development settings, to analyse evidence gaps and challenges in evaluating these types of programmes, and to provide guidance for future evaluation of PBR schemes.

1.2 This approach is part of a wider UK government agenda that is being piloted by various government departments, and is closely linked to establishing value for money of expenditures on development aid. DFID has recently initiated three pilot evaluations of PBR approaches, which are led by the commissioning country offices. Further policy work on this approach is currently under development.

1.3 Payment by results (PBR) is a form of aid financing that makes payments contingent on independent verification of results. PBR is part of a wider UK government agenda and several other government departments are piloting this approach.

1.4 PBR as defined by DFID has three key elements:

- Payments based on results;
- Recipient discretion – i.e. the recipient has space to decide *how* results are achieved; and
- Verification of results as the trigger for disbursement.

1.5 DFID further distinguishes two basic types of PBR approaches:

- Results-Based Aid (RBA) – payments from funders to partner governments.
- Results-Based Finance (RBF) – payments from a funder or government to service providers (which could be an organisation and/or individual service providers).

1.6 It is important to recognise that internationally, and in the research literature, there is no consensus about use of terminology. "Results-based financing" is often (but not always) used as a generic term to describe any approach where funding is linked in any way to performance. Other common terms and monikers, all referring to payment based in some way upon achievement of outputs/outcomes, that are often used interchangeably include: Payment for Performance (P4P), Cash on Delivery (COD), Performance-based Financing (PBF), Performance-based Payments (PBP), Performance-based Incentives (PBI), and Performance-based contracting (PBC). Often, PBR is equated with contracting for services, to NGOs and/or private contractors with at least some results component. Nevertheless, there frequently are variations in approaches within and across these various terms.

1.7 Sometimes there is a distinction between supply and demand side approaches, with the former involving incentives for service enablers/providers (that may include public, non profit, and/or private organisations) and the latter to beneficiaries (e.g. communities, families, individuals). Individual practitioners (e.g. health or community workers) generally, but not always, are considered part of the supply side. The PBR

literature frequently intermingles consideration of incentives for both providers and recipients, including conditional cash transfers (CCTs).

1.8 An internal Norad document (Olsen, 2011) makes the following observation:

> With the lack of one commonly accepted definition, terms like *Result-based Financing*, *Performance-based Financing* and *Pay-for-Performance (P4P)* are often used inter-changeably. The Working Group on Performance-Based Incentives suggests the following working definition for P4P: "Transfer of money or material goods conditional on taking a measurable action or achieving a predetermined performance target."

1.9 As discussed in more detail in the text below, even the "same" types of approach in practice are frequently defined, and implemented, in many different ways. For example, there are many different formulas for defining and verifying "results", who is to receive the financial incentive, and under what conditions. In some but not in other cases, payments are not made until after verification, which may or may not involve government or institutional information systems. The financial incentive provided can vary widely, from a minor bonus to the sole means of funding. A range of complementary activities and supports invariably accompany most incentive schemes. Even implementation of the "same" scheme in practice may vary widely from setting to setting. Indeed, as is the case with General Budget Support (GBS), it is fair to say that every application of PBR, irrespective of its label, is unique. These are important considerations to bear in mind when considering interpretation of some of the findings from specific research studies.

Definition of evaluation and criteria for assessing quality

1.10 This study follows the Organisation for Economic Cooperation and Development/Development Assistance Committee (OECD/DAC) definitions, description of evaluation criteria and evaluation quality standards, to which DFID has agreed.

1.11 OECD/DAC defines evaluation as follows:

> "Evaluation is the systematic and objective assessment of an on-going or completed project, programme or policy, its design, implementation and results…. Evaluation also refers to the process of determining the worth or significance of an activity, policy or program."

1.12 There are many definitions and descriptions of evaluation in the (extensive) evaluation literature and texts, with a variety of other published and internationally accepted statements of standards and evaluation principles[1]. Evaluation can use a wide variety of both quantitative and qualitative methods, providing comprehensive information about

[1] OECD/DAC definitions, principles, criteria and standards are available in various documents that are summarized in: OECD DAC Network on Development Evaluation. *Evaluating Development Co-operation: Summary of Key Norms and Standards* www.oecd.org/dac/evaluationnetwork). Other widely recognised standards and principles include the US-based Joint Committee Standards on Evaluation, as well as related sets of standards by the African Evaluation Association, and statements of evaluation principles by various sources such as the UK Evaluation Society and others.

a wide variety of potential questions such as what is taking place, why, and whether it is appropriate or not. Evaluation potentially can address a wide range of questions and considerations such as: the nature of an intervention and how it is being implemented, identification of reasons and mechanisms for what is taking place, needs analysis, as well as documenting outcomes and impacts resulting from what was done. In particular, evaluation is intended to be *practical*, providing guidance for future policy and programme directions that can range from minor fine-tuning in the short term to consideration of very different policy directions.

1.13 OECD/DAC further sets out Principles for evaluation and identifies five main evaluation criteria for evaluation of development programmes and projects (relevance, effectiveness, efficiency, impact, and sustainability). DAC documentation identifies more specific questions under each of these criteria. These are listed on the table on page 20 and discussed in Section 3.1 of this report.

1.14 The recent DAC *Evaluation Quality Standards* (EQS), endorsed by members of DAC including DFID, identify the key pillars needed for a quality evaluation process. These form a basis for assessing the quality of existing completed evaluations of relevant PBR approaches and in providing guidelines for future evaluation approaches.

1.15 Evaluation approaches considered in this study are not limited to any particular methodological approach, but involve credible, well executed methods that are independent of those undertaking the implementation. This means *inter alia* that scoping exercises, one-off surveys, and related undertakings would not fall under the generally accepted definition of evaluation and thus are beyond the scope of this study.

1.2 Focus, method, and scope

1.16 This report addresses the following considerations, in subsequent sections of this report:

- The evidence base regarding the effectiveness and efficiency of PBR approaches in development.
- An analytical critique of the quality of existing evaluations.
- Guidance for approaches, including evaluation questions and methods, to future evaluations of PBR programmes.

1.17 This study was carried out through a review of relevant evaluations that could be identified, reviews and syntheses, critiques, and other relevant documentation. The major means for identifying completed evaluations was through contacts with key informants who were expected by DFID to be familiar with relevant evaluations in this area. DFID wrote directly to all other aid agencies on the OECD/DAC Evaluation Network mailing list, requesting their assistance in identifying relevant evaluations that they had undertaken or commissioned or were aware of.

1.18 A variety of other key informants known to be working or otherwise familiar with PBR approaches were also contacted. Many of these were identified by DFID (staff from Evaluation Department as well as from the Improving Aid Impact Team). A number of other individuals in a position to be aware of any relevant evaluation studies were also contacted, along with follow up with others identified during the course of

the data gathering process. In addition, a number of databases were also searched to identify any relevant recent evaluations. Annex 1 provides a list of key informants contacted (excluding DFID staff) as well as databases that were searched.

1.19 Identification of relevant evaluation and research studies was not at all straightforward and required a major effort, involving numerous and sometimes repeated contacts with partners and key informants as well as a search of various databases and other sources.

1.20 **Limitations.** Only a very limited amount of time and resources was allotted for this exercise. The terms of reference also specifically indicated that this report should be short (maximum 30 pages). This report intentionally was expected to provide a general overview of the current situation with respect to evaluation of PBR initiatives and with priority on identifying implications for future evaluation of PBR approaches. It takes into consideration, but does not cite or specifically refer to, each individual evaluation or article identified in Annex 2. Despite responses from key informants who collectively would be expected to be familiar with the work in this area, along with supplemental searches, this study was not intended to be fully comprehensive. Thus there may well be some other relevant evaluations that were not identified – although some very recent reviews were considered that likely would have been able to identify the most significant and relevant studies.

1.21 The scope of this project encompasses completed evaluations[2] of PBR approaches with respect to any development domain and undertaken by any actor. There are a wide variety of results-oriented approaches that fall outside the scope of this study. Similarly, conditional cash transfers (CCTs) to beneficiaries represent a particular modality that also falls outside the scope of this study, as there is a relatively strong understanding of how these instruments can be evaluated and where the evidence gaps are.

1.22 As Annex 2 suggests, only a limited number of completed evaluations of PBR schemes could be identified. Nevertheless, this review uncovered a number of recent reviews and syntheses, as well as various critiques. These in turn have considered, critically assessed, and synthesized evidence from a wide range of additional evaluations in addition to those specifically identified, as well as the quality of this evidence. As discussed below, generally a fairly consistent picture emerges from these various sources of information.

[2] With a few exceptions, notably the most recent evaluation plans for three DFID-sponsored pilots.

2 Evidence base for PBR approaches

Key messages:

- With very few exceptions, almost all research and evaluation studies of PBR have been in the health sector.
- There is limited evidence that PBR approaches offer value-added vis-à-vis other modalities.
- PBR needs to be implemented as part of a package that includes other forms of supports and services. The underlying complexity of each intervention presents a serious challenge to implementation and evaluation, inhibiting meaningful generalisation without identification of the specific mechanisms at play.
- There has been limited attention to some basic questions about PBR approaches, including the mechanisms by which incentives work, cost effectiveness, comparison with other potential approaches, impact on equity, and sustainability.
- Unintended negative effects are quite possible.

2.1 Mapping of evaluations

2.1 Annex 2 lists evaluations, syntheses and reviews that could be identified through the process described above.

2.2 A generally homogeneous picture of these evaluations emerges:

- Almost all the evaluations are with respect to initiatives in the health sector, in particular with respect to utilisation of health services. There was one evaluation of a project incentivising local utilities in Indonesia to provide water and sanitation access, and a couple considering education.[3]
- Almost all the studies are of RBF initiatives (incentives to service provider organisations and individuals) rather than to governments. A major exception is the evaluations and critiques of the GAVI Foundation's Immunisation Services Support Programme (ISS), as well as the evaluation of the Global Fund.
- Almost all evaluations are impact evaluations, using mainly experimental approaches or in some cases quasi-experimental designs (e.g. before-after, comparison groups, and occasionally some other designs).
- As discussed in more detail in Section 3.1, virtually all the studies, and most of the reviews, were undertaken by those who view themselves as researchers rather than as evaluators, with limited reference to the evaluation literature, including to evaluation principles and standards. Most were carried out by health researchers with the studies published in medical or health journals.

2.3 This study identified a fair number of recent and generally quite informative and credible reviews and syntheses, along with various critiques of the evaluation work that has been undertaken to date. These reviews generally assess the quality of various individual studies and provide syntheses of trends and findings across specific evaluation studies, as well as gaps and implications arising, taking into account the strengths and

[3] An exception is a World Bank review of impact evaluations of social safety nets. But the focus of this review is outside the scope of this study, for example considering evaluation of CCTs and other supports including workfare, school feeding, and educational fee waivers.

limitations of the evidence. These and other characteristics of the reviews mean that reviews and syntheses generally are considered more credible than individual studies.

2.2 Status of the evidence regarding effectiveness of PBR approaches

Do PBR approaches "work"?

2.4 The importance of an outcome (or results) orientation, focusing on the actual benefits arising rather than on inputs and services provided, for all public services including development aid, is well recognised. By paying for actual, verifiable results achieved, PBR has been promoted as a means of ensuring the achievement of results. But is there evidence to support this assumption in practice?

2.5 Overall, a generally consistent picture emerges from existing research and evaluation and in particular from comprehensive reviews and critiques that have explored the strengths and limitations of methodologies that have been used. Overall, the evidence regarding the potential impact of incentives to change professional practice is weak. As Vähämäki, Schmidt, and Molander (2011) indicated, in a major review supported by SIDA of PBR along with other results-oriented approaches:

> "The basic idea . . . seems uncontested…. However, the management practice, per se, has encountered severe challenges and difficulties in its implementation, and is being questioned by practitioners and researchers."[4]

2.6 Perhaps the most optimistic conclusion that can be drawn from available evidence is that contracting out may increase access and use of health services (e.g. Lagarde and Palmer, 2009). Oxman and Fretheim (2009a) indicated that financial incentives targeting recipients of health care and individual healthcare professionals may be effective in the short run for simple and distinct, well-defined behavioural goals. They add that there is less evidence that financial incentives can sustain long-term changes. There have been some but limited attempts to look at the potential contribution of incentive approaches to health outcomes beyond utilisation of health services. But in a very recent Cochrane Collaboration review by the same group of researchers (Witter, Fretheim, Kessy, and Lindahl, 2012), they came to a more humble conclusion, that the current evidence base is too weak to conclude that PBR approaches have any significant impact.

2.7 The existing research base is generally quite consistent. Any potential results are with respect to access to services rather than to health (or other) outcomes, and are short

[4] Examples of difficulties in implementation identified in various studies include: defining appropriate and desired outcomes; being clear and realistic about expectations and communicating this across the organization, developing meaningful indicators; providing the necessary technical supports, expertise, and resources so that appropriate action is possible; priority setting and appropriate allocation of resources; providing appropriate motivation for staff who would need to take action (who may or may not be eligible for financial rewards based upon performance); developing, implementing and updating appropriate administrative and systems mechanisms at both central and local levels; developing and using feedback systems to monitor progress; flexibility in making necessary modifications, ensuring that attention and resources are not diverted from other important needs; roles and actions of key actors and partners; showing flexibility when conditions, beneficiary needs, and other considerations are not as expected or when unforeseen difficulties (e.g. transport to rural areas) are encountered; etc.

term in nature and may, possibly, displace attention to other needs. While there is the occasional study that suggests a positive outcome of a PBR approach, the validity of these conclusions has been challenged by others, and there is at least one case where the control group apparently performed better than the group receiving the incentives. It is also well known that publication bias means that research indicating lack of effect of these (or any other) schemes are not so likely to be published.

2.8 For example, a recently published evaluation (Basinga et al., 2011) of a PBR scheme in Rwanda that provides financial payments based upon a variety of primary maternal and child outcomes (e.g. utilisation of prenatal care, rates of child visits, vaccinations rates, institutional deliveries) has been receiving considerable attention, with the potential of this approach being exported to neighbouring countries. However, the validity and conclusions of this study have come under question, for example with the recent Cochrane Collaboration review (Witter et al., 2012) pointing to methodological limitations, e.g. some control districts were found to have existing pay for performance schemes, with "discrepancies in the explanations about how districts were allocated to intervention or control in different reports of the study." Another evaluation in Rwanda (Kalk, 2010) used in-depth interviews with administrators, doctors, nurses, and patients to explore the actual effects of the incentives; these included considerable side effects such as gaming. Others (e.g. Ireland, 2011) point to limitations of scaling up from the Basinga et al. study given the unique context in Rwanda, with extensive government and donor attention and resources, extensive attention to health sector reform and innovation as well as attention to various MDGs.

2.9 However, the simple answer to the question "does PBR work?" is that this is the wrong question. As Eldridge and Palmer (2008), for example, have indicated:

> "PBP is a popular term, but one that needs more careful definition. The current enthusiasm for it raises a number of unanswered questions. The existing literature would be helped greatly by case studies detailing the potential advantages and disadvantages of PBP, determining the most influential factors for success, and recording experiences in field situations in some detail."

2.10 Annex 3 provides a table, taken from Witter et al., 2012, indicating the outcome and impact measures used in the studies included in this Cochrane Collaboration systematic review. Most "outcome" measures (e.g. provision of various forms of primary care, preventative, and other health care services; rates of vaccinations; attendance rates for antenatal care; utilisation of preventative care services by children)[5] are primarily measures of service utilisation and generally refer to what are defined as outputs, rather than outcomes, within the evaluation community.[6]

[5] This applies not just to measures identified in this review, but also in almost all the research studies. Some other frequent examples of "outputs" for which compensation is provided in various PBR schemes include: payment for delivery of a basic package of health services at the primary care facility level (e.g. the Afghanistan study included in the Eldridge and Palmer review), bonuses paid to NGO based on achievement of performance targets including management capacity and health output targets (Haiti), etc.

[6] In a very few studies, such as one included in the Lagarde and Palmer review, there are some examples of what may be considered as outcome indicators, such as update of vitamin A for children, % women with a live child aged 6 to 23 months who currently use a contraception method.

Nature and actual implementation of PBR schemes

2.11 A key finding of this study is that implementation of PBR schemes, in practice, is not so straightforward and has encountered severe challenges and difficulties. As a result, each application is unique, and often quite different from how initially envisioned or conceptualised. For example, as Lagarde and Palmer (2009) observed, specifically with respect to contracting out (just a subset of broader PBR possibilities):

> "Unlike more simple, clinical interventions, there is an array of factors that can influence a strategy like contracting. First, the very label 'contracting out' can mean different things in different settings and in fact even in the studies included in this review it involves different elements. What is actually implemented is the result of the role and decisions of key actors and interactions at quite decentralised levels, therefore implementation issues need to be looked at very carefully"

2.12 This represents more than just a design or implementation challenge, but goes to the essence by which incentives can be expected to "work". A major finding and key theme in the reviews and evaluations identified is that PBR approaches never operate alone, but as part of a package of increased funding, technical support, training, new management structures and monitoring systems, and often in the context of a significant reform effort. As Oxman and Fretheim (2009b) have indicated:

> "It is not possible to disentangle the effects of financial incentives as one element of RBF schemes, and there is very limited evidence of RBF per se having an effect. RBF schemes can have unintended effects. ...For RBF to be effective, it must be part of an appropriate package of interventions, and technical capacity or support must be available."

2.13 Thus PBR schemes, by their very nature, need to be implemented in conjunction with other forms of supports and actions. This has implications both for the design and implementation of future PBR schemes and how they should be evaluated. Research models used to date for the most part have been unable to deal with complex relationships, such as when PBR schemes are implemented together with other programme elements, in particular increased funding and technical support. As Section 4.3 indicates, there are however models of evaluation that can cope with complexity and that may strive to document "contribution" rather than inappropriately seek uni-dimensional "cause-and-effect" and attempt to isolate the impact of an intervention that can only work in given circumstances and in combination with other interventions and factors.

2.14 But this is more than an evaluation consideration. As Pearson (2010, 2011) has observed: "Many of the new initiatives are designed to work on a project basis, and seem to disregard the complexity of results management as such. There is seldom only one solution on how to achieve the best results." Others have also reinforced this point. The available evidence suggests that it is naïve and inappropriate to think that PBRs can "work" without also taking into account needed contextual supports and other factors. PBR schemes assume that change is within the control of the agent. But this is not always the case, and when there are barriers preventing even the most motivated individual (or organisation) from taking needed actions, desired change will not take place, irrespective of the incentives provided.

2.15 Most of the PBR schemes that have been researched have been implemented on a pilot basis, inevitably representing atypical and unsustainable situations. Usually these pilots receive significant attention and resources that likely would not be sustainable on an ongoing basis.[7] Staff and others involved in pilot operations, frequently including the intended beneficiaries, know that they are engaged in something special and may be more committed, and often the staff involved in implementing pilots have more expertise and commitment than those engaged in regular operations. Thus approaches tried out on a pilot basis do not necessarily lend themselves to scaling up. As Toonen et al. (2009) indicate: "Scaling up requires new institutional arrangements at both central and local level which has implications for compatibility with existing structures and for sustainable funding of transaction costs." They observe that this has major implications for government and other systems and management, and for ongoing funding. Other studies and reviews come to similar conclusions.[8]

2.16 Toonen et al. (2009) also observe a paradox behind some PBF approaches. Contracting out (with incentives) is being promoted due to perceived government limitations in capacity to deliver services directly. However, for contracting to work on an ongoing basis, there is a need for strong managerial capacity, along with other forms of institutional changes, systems support, and the like.

Some questions about PBR as yet without answers

2.17 Thus far, the major focus of research has been on the elusive question: does PBR work or not. Nevertheless, there are many other questions that can affect decisions about deciding to embark on a PBR strategy or not and what form it should take. The following is a list of some important considerations regarding PBR that have received scant attention to date and that have been identified by the reviews and critiques as important and relevant. These are taken up again in Section 4.1 of this report.

- Nature of the implementation of the PBR scheme, and the mechanisms whereby incentives "work" or not.
- Level and nature of incentives required to change behaviour.
- Efficiency and cost effectiveness of PBR approaches.
- Comparisons with other programmatic approaches.
- Effects on quality and broader health (and other) outcomes.
- Positive and/or negative effects of PBR approaches on equity.
- Interactions with the broader health (and other) systems.
- Impact of financial incentives on intrinsic motivation and sense of responsibility and health care and community service workers.
- Long-term effects and sustainability.

[7] For example, a preliminary evaluation of the ongoing Health Results Innovation Trust Fund (HRITF) initiative of the World Bank (Martinez et al., 2012) indicates that the median cost per project is $12M, with a range from $1-20M.

[8] Sections 3.2 and 4.3 provide further discussion of the limitations of generalizing from pilot studies together with some ideas of working around these, such as through using a realist evaluation approach that can identify mechanisms that might apply in other similar contexts.

Unintended effects

2.18 A recurring theme in the various reviews and critiques concern the very real potential of unintended negative effects arising from PBR approaches. They point out that only a few of the studies undertaken have explored unintended effects. For example, the recent Cochrane Collaboration review (Witter et al., 2012) indicated that: "Only two studies [of the nine that met the criteria for inclusion] reported on unintended effects – in both studies the authors voiced concerns about the curative nature of the coverage targets and whether this may squeeze out preventive care." Even though few studies looked at this explicitly, nevertheless many of the reviews and some of the evaluations have been able to identify strong evidence of unintended effects.

2.19 Vähämäki et al. (2011) and Pearson (2011) are among many others who have observed that an unrealistic focus on indicators can distort performance towards activities that are most easily measured and achieved in the short-term (e.g. outputs or quick fixes), diverting attention and funding from more important, and more enduring, outcomes. Oxman and Fretheim (2009a) similarly have discussed how "results-based financing can have undesirable effects, including motivating unintended behaviours, distortions (ignoring important tasks that are not rewarded with incentives), gaming (improving or cheating on reporting rather than improving performance), widening the resource gap between rich and poor, and dependency on financial incentives". Other studies document how the wrong indicator can lead to distortions or misrepresentation of actual performance, or neglect of other needs (such as less attention to quality [Macro International, 2009 evaluation of the Global Fund] or neglecting patients in intensive care [Kalk et al., 2010]).

2.20 Findings from PBR evaluations and reviews about potential unintended effects are consistent with those from the general literature on performance indicators[9] This literature indicates that distortions and the potential of perverse effects are greatest when a very limited number of indicators are used where the measure in practice then becomes the objective and when there are strong pressures to engage in gaming and distortions, such as financial rewards for meeting targets. Nevertheless, some PBR proponents suggest the fewer the indicators the better.

2.21 Questions have been raised in the reviews (e.g. Vähämäki et al., 2011; Pearson, 2011) about dangers in particular of a hands-off approach. Concern has been raised about the compatibility of such an approach with other DFID principles and development objectives, such as: consistency with Paris Declaration principles, attention to equity and rights, responding to actual needs irrespective of agreed-upon targets (which may change over the course of time or in response to situations beyond anyone's direct control such as droughts, epidemics, etc.), avoiding corruption, developing and supporting local capacity, local ownership, partnerships and engagement of the community, addressing other donor and government priorities, etc.

[9] E.g. see Bemelmans-Videc et al., 2007.

3 Critique of research and evaluation studies on PBR

Key messages:

- Practically all the "evaluations" identified may be more appropriately described as research. While there has been extensive use of sophisticated research designs, there is limited attention in these studies to generally accepted evaluation standards and quality criteria. There has been limited attention to the five generally accepted DAC evaluation criteria, except to objectives achievement.
- The research quality of studies that have been identified has been considered poor by major systemic reviews and critiques, with challenges both to their internal validity (integrity of the findings) and external validity (ability to generalise). There are, however, some interesting exceptions.

3.1 Conformity with generally accepted evaluation criteria and standards

3.1 Practically all the "evaluations" identified, including those discussed in the various reviews, may be more appropriately described as research than evaluation studies, undertaken by experts who consider themselves researchers rather than evaluators. Most of the studies have been published in medical or health rather than in evaluation journals, with scarcely any reference to the evaluation literature. The vast majority have focused almost exclusively on impact, mainly employing black box designs that have given little attention to the circumstances under which the schemes that were researched were undertaken or factors responsible for success or failure. There is a heavy emphasis on technical aspects of the research with a primary focus on internal rather than on external validity.

3.2 Given the above, almost all the research to date has not been undertaken with any specific consideration, or apparent awareness, of generally accepted principles and standards for evaluation. Nevertheless, there *are* standards and principles to which DFID in particular has agreed, in particular the DAC evaluation criteria and the Evaluation Quality Standards (EQS). Thus it can be appropriate to apply these criteria and standards against evaluations done to date, and as a framework for identifying gaps and limitations in the body of evidence to date and in particular for identifying priorities for future evaluation undertakings.

DAC Evaluation Criteria

3.3 The following table provides an overview of how the evidence identified[10], collectively, respond to the five DAC evaluation criteria

DAC Evaluation Criteria

DAC criteria for evaluating development assistance[11]	*How addressed by available evidence*
Relevance: The extent to which the objectives of a development intervention are consistent with beneficiaries' requirements, country needs, global priorities and partners' and donors' policies. Note: Retrospectively, the question of relevance often becomes a question as to whether the objectives of an intervention or its design are still appropriate given changed circumstances.	• Generally the relevance and appropriateness of the intervention and its objectives are assumed rather than questioned. Some of the reviews, however, do suggest that the results being rewarded are short term in nature and not linked to health or other outcomes.
Effectiveness: The extent to which the development intervention's objectives were achieved, or are expected to be achieved, taking into account their relative importance. In addition to identification of objective achievements, considerations regarding effectiveness also include identification of the major factors influencing the achievement or non-achievement of the objectives.	• The main focus of almost all the research has been on achievement of objectives, frequently on just a very small number of quantitative, easy-to-measure indicators. • With predominantly "black box" research designs used, there has, however, been negligible consideration to the second question regarding factors influencing the achievement or non achievement of the objectives.
Efficiency: A measure of how economically resources/ inputs (funds, expertise, time, etc.) are converted to results. This generally requires comparing alternative approaches to achieving the same outputs, to see whether the most efficient process has been adopted.	• Almost no attention to considerations of efficiency, cost effectiveness, or of alternatives.
Impact: Positive and negative, primary and secondary long-term effects produced by a development intervention, directly or indirectly, intended or unintended. Consideration of impact should also include the positive and negative impact of external factors, such as changes in terms of trade and financial conditions.	• Most research has focused on changes in service utilisation rather than in secondary or longer-term outcomes or impacts. • There has been negligible attention to unintended or negative effects, although these have been identified in a number of reviews. • There has been limited attention to the contribution of other factors besides the incentive mechanism (e.g. technical assistance, government policies and reform initiatives, level of resources).
Sustainability: The continuation of benefits from a development intervention after major development assistance has been completed. The probability of continued long-term benefits. The resilience to risk of the net benefit flows over time.	• There has been limited attention in research to date on sustainability. Some reviews indicate that many of the PBR projects studies are in atypical situations that are not likely to be continued or scaled up, and also question the sustainability of PBR approaches.

[10] These include, as well, evaluations identified and considered in reviews and syntheses.

[11] Definitions are taken from the OECD/DAC *Glossary of Key Terms in Evaluation and Results Based Management*, with some supplementary explanations from the OECD/DAC *Evaluating Development Co-Operation: Summary of Key Norms and Standards.* 2nd Ed. 2011.

3.4 As this table suggests, the main focus of the research work to date has been on establishing a causal relationship between the intervention (some variant of PBR) and its stated objective (generally change in utilisation of a specified health service).[12] The primary methodology that has been used is a counterfactual design, generally employing a Randomised Control Trial (RCT) design (in some cases, various forms of quasi-experimental designs have been used, including use of non-random comparison groups and before-and-after designs). This design, while potentially strong at establishing cause-and-effect relationships in specific circumstances, is not intended to address other considerations and evaluation questions as suggested by the DAC evaluation criteria.

DAC Evaluation Quality Standards

3.5 The Standards[13] (typically referred to as EQS) "identify the key pillars needed for a quality development evaluation process and product." They are intended as a guide to good practice, for use by DAC members (including DFID) as well as potentially by others. They are intended to be applied sensibly and adapted to specific contexts, rather than to be used as a manual or guide. In this context, they can serve as a basis for suggesting questions that can be asked about the quality of evaluation undertakings.

3.6 It is beyond the scope of this study to do a complete assessment of research studies identified against the 37 different standards that form part of the EQS. The following are, however, some observations regarding some selective standards under the four EQS categories of standards:

1. Overarching considerations

3.7 There is some, but limited, evidence of a *partnership approach* being systematically considered early in the evaluation process. Often, PBR schemes and/or evaluation may be imposed upon stakeholders (or made a condition of funding, which amounts to the same thing), with some but not other stakeholder groups given an opportunity to be consulted in the planning of the evaluation. When outside research experts have been chosen to implement "rigorous" designs, there may be little opportunity for meaningful contributions of partners to the evaluation process. In some cases, local agencies or researchers may be involved in the process that can assist with capacity development of development partners. There are examples (such as various interventions in Rwanda) where the national and/or sub-national governments may be engaged in the process. But this is not always the case. Evaluations also vary considerably with respect to their *quality control* mechanisms. As discussed and illustrated below, the systemic reviews and syntheses identified significant shortcomings in the methodologies of many studies.

[12] This is frequently labelled as "impact evaluation", where impact is defined as identifying a causal relationship between an intervention and a consequence at any level. This is, however, considerably narrower than the DAC definitions of impact, and, in general, with the literature on outcome-oriented approaches that require taking into consideration the entire results chain and, in particular, the impact on the livelihoods of beneficiaries and communities.

[13] See: OECD/DAC. *Quality Standards for Development Evaluation* for a complete list of the standards and a description (e.g. in *Evaluating Development Co-Operation: Summary of Key Norms and Standards*. 2nd Ed. 2011).

2. Purpose, planning and design

3.8 One of the standards specifies that: "the development intervention being evaluated is clearly defined, including a description of the intervention logic." This is a particular area of weakness of most of the studies identified, and remarked upon in many of the reviews and syntheses. There is limited consideration to identifying and describing how the PBR intervention was *actually* implemented in practice, which often was different from how it was initially envisioned, (in minor or in major respects) such that conclusions about attribution may be greatly compromised. There is very limited attention to the presence of other factors that may contribute to (or detract from) the achievement of results, and only a couple of examples of meaningful theory-of-change (attribution or programme logic) models could be identified[14]. The theory-of-change approach is generally recognised in the evaluation literature as a basic planning tool that can, for example, help identify the expected interaction of an incentives scheme with other elements and interventions and indicate how these are expected to lead to long-term impact (with intermediate steps also identified)[15].

3.9 The EQS specify that stakeholders should be involved early in the evaluation process and be given an opportunity to contribute to evaluation design. The EQS also specify that evaluation objectives should be translated into specific evaluation questions that can then inform the development of the methodology. Given the uniformity of the methodological approach used in most of the evaluations identified and the lack of research attention to many important questions identified elsewhere in this report, it could be that in some cases the general methodological design and evaluation questions were determined beforehand, with limited involvement of all groups of stakeholders.

3.10 In addition, the EQS indicate that the evaluation plan should address cross-cutting issues such as gender equality, human rights and the environment. While some of the evaluations do look at gender differences, as indicated in the preceding section, there has been limited attention to considerations regarding equity. There is little reference to environmental considerations, and some of the reviews and critiques (e.g. Vähämäki, 2011) indicate that one consequence of a hands-off approach could be lack of attention to such considerations such as these. Ethical questions (e.g. Ssengooba et al., 2012) also are sometimes raised about randomised approaches.

3.11 One of the EQS standards concerns the adequacy of resources provided for the evaluation itself. In most cases, budgetary information is not available. However, in some cases, it appears that budgets are well in excess of that available for many other forms of evaluation, which places an onus on demonstrating value for money spent on these exercises[16]. The time frame to implement many of these evaluations often spans several years, which may not be out of line with large-scale research efforts, but is considerably longer than for most policy-related evaluations, and beyond most short to medium term planning and programming cycles.

[14] E.g. see the model developed by Victora et al. (2010) on p. 26.
[15] See, for example, a recently released report by DFID: Isabel Vogel. (2012*). Review of the use of 'Theory of Change' in international development*.
[16] For example, see Note 7 regarding the budget available for the HRITF pilots and evaluations.

3. Implementation and reporting

3.12 Generally, the research teams undertaking evaluations that have been identified appear to be independent of the stakeholders, although this does not always appear to be the case or possible to ascertain.

3.13 One cluster of standards under this general category concerns issues such as: clarity of analysis, context of the intervention, acknowledgement of changes and limitations, intervention logic, incorporation of stakeholder comments, etc. Generally there is considerable (often very considerable) attention to treatment of data – but as many of the reviews point out, with inadequate consideration of how the intervention was implemented, challenges to the integrity of the design emerge (e.g. when control or comparison groups are compromised or other confounds come about through the project implementation process). There are few examples that present stakeholder comments or show how these were taken into consideration.

3.14 The EQS standard regarding the evaluation report says that it should be in an appropriate form and understandable given its intended audience. It is not always clear who the intended audience is for many of the publications identified. Other researchers often seem to represent the main user group – but with the implicit understanding that this information would also be of interest to policymakers. Some of the research reports are highly technical and may be difficult for non researchers to fully assimilate. However, a surprising number of even technical reports have quite readable succinct summaries highlighting implications for policymakers as well as for researchers.[17] These would be more useful however, as discussed elsewhere, if the research questions would have explored the reasons for the given findings about impact.

4. Follow up, use, and learning

3.15 This category includes three standards addressing various aspects of evaluation use: timeliness, relevance, and use of evaluation; dissemination; and systematic response to and follow up on recommendations. A key factor that often distinguishes evaluation from research is a prerequisite that evaluation focuses specifically on utility, whereas research is often carried out for long-term knowledge generation. Given the research nature of almost all the studies that could be identified, and the limited range of questions that these have addressed, the actual utility and awareness of what makes for an utilisation-focused approach of much of the research are unclear. Nevertheless, it should be acknowledged that much of the research that has been carried out can (and indeed has) at a minimum served to stimulate debate and consideration regarding the appropriateness of PBR and related questions.

3.2 Research methodologies employed: validity considerations

Internal validity considerations

3.16 A major focus of systemic reviews concerns the research integrity, quality and validity of the methods used and evidence presented from the research studies that have been

[17] For example, the two recent Cochrane Collaboration systemic reviews (Witter et al., 2012; Lagarde and Palmer, 2009), are quite readable, and include plain language summaries.

considered. It is useful to distinguish between internal validity considerations (whether or not the specific intervention did make a difference in the particular context it was tried) and external validity (generalizability of findings, such as to other settings, populations, treatment and measurement variables, or even to the same setting at a different time period). In other words, internal validity concerns whether or not the research can indicate that application of a particular PBR approach made a difference in the specific situation and set of circumstances where it was tried. External validity refers to whether there is evidence that these findings might be applicable in a different setting or set of circumstances. In order to be able to make use of research and evaluation, both internal and external validity are essential.

3.17 Systemic reviews typically first identify a wide range of potential studies, and then rule out of consideration all but a very few that meet basic criteria of quality identified in advance.[18] Even after this winnowing process, systemic reviews have been very critical of the overall poor quality of research concerning the effectiveness of PBR approaches. For example, a very recent (2012) Cochrane Collaboration review undertaken by Witter et al. concluded that: "The overall quality of current evidence is poor … Only one study [of the nine that met the inclusion criteria] was assessed as having low risk of bias." Just two of these nine studies looked for unintended effects, and just one considered equity. Only one of these studies looked at any form of health outcomes, with the research mainly looking at changes in utilisation or coverage of health services.

3.18 Other reviews and critiques have come to similar conclusions. For example a slightly earlier (2009) Cochrane Collaboration review by Lagarde et al., focusing specifically on contracting for services, concluded that: "The poor quality of the studies included in this review suggests that so far there have been few attempts to try to evaluate the effects of contracting out health services." Just three studies met the criteria for inclusion in this review, but nevertheless: "Each of the studies presented methodological weaknesses in their analysis, design, or both."

3.19 This review, in common with many others, also identified practical constraints that precluded meaningful randomisation in practice. For example, Lagarde and Palmer refer to a study from Cambodia that was designed as a cluster-randomised controlled trial which, however, was limited by several flaws that are likely to have biased estimates of effects. In this case, only a portion of the randomised districts chosen could be included due to too few bids, numbers of clusters chosen were too limited leading to non equivalence of intervention and control areas, and "financial resources available for contracted districts were 85% greater than those of control districts". This example also illustrates how the manner in which the intervention is implemented can also affect the research design.

3.20 This is a common theme. Other reviews raise questions about the integrity of the application of designs in the studies that they have examined. It has proven to be very difficult in practice to implement and to maintain the fidelity of "rigorous" designs, resulting in biases and compromising the ability to draw conclusions from the data obtained. Numerous other confounds are also commonplace, some resulting from the

[18] Generally based upon the research design used, with RCTs given preference. This can lead to the failure of systemic reviews to take into account research employing other methodologies. However, independent searches for other evaluations only turned up a very small number of other evaluations.

extreme difficulties of maintaining the fidelity of the experimental and control groups, but also some being due to other factors. As discussed earlier, incentivisation schemes rarely can be or are introduced independent of other measures. This means that it is impossible to attribute results to the incentive mechanism rather than to the typical package or services and support. Other confounds may include, but are not limited to, additional profile and attention given to the project, technical support, and resources. As Pearson (2011) for example has observed, a key issue "is how to disentangle the effects caused by the specifics of a scheme - through its results focus - from the effects due simply to the additional funding attached."

3.21 None of the research studies identified employed (nor could they) double blind designs (where neither the service provider nor recipient knows if they are in the experimental or control group). People (managers, health and other professionals, beneficiaries, families, community members, and others) can and do act differently when they know that they are (or are not) receiving special treatment, such as the use of financial incentives.

3.22 As discussed in other sections of this report, a major challenge identified from the reviews and research studies to date, indeed arguably one of the major limitations, is the limited consideration given to documenting the *actual* nature of the intervention itself, that is what occurs in practice as opposed to what was initially envisioned. Impact evaluation, and establishing attribution is not just about identifying impact, but about documenting a causal link to an intervention. When the nature of the intervention is not clear, this becomes problematic, resulting in data that are difficult or impossible to interpret meaningfully.

3.23 There clearly is a role for randomised impact evaluation designs. However, a key lesson from the research that has been done to date is that it is deceptively challenging in practice to implement these designs in a way that can avoid challenges that may compromise the validity of findings. Martin Ravallion (2012) has observed that: "Experiments are rarely so clean in practice, such that a number of assumptions are needed to draw valid inferences about the experimental population, let alone valid policy inferences - including assumptions that are not required by non-experimental studies."

External validity considerations

3.24 Research designs used to date generally are very weak regarding external validity, and have given limited consideration to the ability to generalise or to how findings can be applied in other situations or contexts. In order to be able to apply or adapt findings from one setting or situation to another, it is essential to be able to understand the mechanisms, the 'hows' and 'whys' through which a PBR approach has (or has not) resulted in changes. It is also essential to identify key contextual factors, including the role of other factors that may also need to be in place for a PBR approach to succeed. Black box approaches, the predominant research design that has been used, have paid scant attention to date to documenting or providing understanding of *process* – attention which is highlighted as a major need by various reviews and researchers, including those who are strong advocates of experimental methods. As indicated earlier, there has been limited use of theory-of-change approaches that can also help to identify the mechanisms at play.

3.25 As previously indicated, most of the PBR initiatives that have been researched have been of pilot projects. Pilots can be very appropriate for trying out new approaches before widespread implementation. They may be used effectively in identifying what may be possible to achieve under ideal conditions. But as previously discussed, pilots are atypical from more normal or regular operations (for example, generally pilots benefit from higher-than-normal financial resources and other forms of support). This limits the ability to draw conclusions, often about attribution and in particular about applicability in more normal situations – unless the evaluation design can identify the mechanisms at play contributing to identified impact and the extent to which they may apply in other contexts. Without such information that largely has been missing from research to date, "scaling up" is problematic.

3.26 Many of the research studies that have been identified have employed questionable counterfactuals. As Victora et al. (2010) have observed: "Traditional designs, which compare areas with and without a given programme, are no longer relevant at a time when many programmes are being scaled up in virtually every district in the world." Rather than using the status quo as the counterfactual, it would be more meaningful to compare various forms of PBR (such as those with differing levels or natures of incentives, with different results indicators etc.), as well as with different types of interventions that might also be used with the intent of enhancing performance and impact.

3.27 In the scientific research tradition, a series of research studies are expected over the long term to contribute to long-term theory testing, through replication, exploration of alternative approaches and questions arising in early research, and ultimately to contribute to knowledge generation. While there clearly is a role for this form of research, after some 20 years of testing out PBR approaches[19] and at least ten years of research using sophisticated designs examining these schemes, there as yet is little concrete evidence that can aid in policy work and planning of current and future PBR approaches. Nevertheless, most of the research studies and systemic reviews call for more research, to a large extent applying the same types of methods that have been used to date. At a minimum, to be able to provide more timely guidance for policy and programmatic work in this area, it would seem appropriate to use a wider variety of evaluation models, designs, and techniques, including those that may be able to focus on those questions of interest that have not been well addressed to date.

3.28 It is necessary to bear in mind that there is no perfect method. Every method has strengths and limitations. It is generally recognised in the evaluation literature that a mix of different approaches can help provide complementarity and triangulation. Yet, to date, there has been a highly restricted range of evaluation designs and approaches used with respect to PBR schemes. Alternative evaluation approaches are discussed in the following section of this report.

[19] E.g. Brenzel (2009), in a World Bank report, indicates that the Bank has supported results-oriented operations in health for some 20 years.

4 Guidance for future evaluations of PBR

Key messages:

- Evaluation should start by identifying priority questions, and then consider potential methodologies that can be applied.
- Arguably the most important question for evaluation is to identify the mechanisms and sets of circumstances under which PBR approaches may make a positive difference.
- There are a range of other important questions about PBR identified in the text, such as cost effectiveness and comparison with other potential approaches and strategies, appropriate size and nature of incentives, and exploration of unintended negative effects and how these can be minimized.
- There is a particular need for evaluation to explore and to describe the process by which PBR initiatives are implemented in practice, and the reasons why changes from the original conception may be needed. Methods can range from M&E to more comprehensive ad hoc evaluation studies.
- A mixed method approach should be utilised. In all cases, articulation of the theory of change can aid in identifying evaluability, indicating what types of questions can be evaluated at given points in time, and serving as a basis for choosing the most appropriate evaluation design. Given the complex context in which PBR schemes work, always in combination with other factors, it may be more appropriate to use a contribution analysis approach rather than linear cause-and-effect.
- There are significant opportunities for theory-based approaches to evaluation that can identify and document the mechanisms at play. In particular, a realist evaluation approach that seeks to identify what works for whom in what circumstances, seems particularly suited to evaluation of PBR schemes.

4.1 How should future evaluations of PBR schemes be approached? There are many suggestions in the various reviews and critiques of PBR approaches, some quite specific, identifying implications for future research or evaluation based upon what has been attempted to date. This section provides some general guidance, drawing largely from these sources and the discussions above. However, these ideas are not intended to represent a workplan or handbook for evaluation of PBR, but rather to suggest possibilities and considerations that should be taken into account in future work in this area.

4.1 Start with the (right) questions

4.2 A basic principle for meaningful evaluation is to start first by identifying *the questions* that need to be addressed. This is a key principle highlighted throughout the evaluation literature, and in the DAC *Standards* for evaluation. Choice of specific methodological designs and approaches should follow, not lead, those questions and information needs of highest priority and potential application.

4.3 As a starting point, one should consider the **full range of potential questions** that evaluation may be able to assist with and not just focus primarily on impact, as has largely been the situation to date. The DAC criteria (relevance, effectiveness, impact, efficiency, and sustainability) can serve as a useful guide to identify objectives and questions for evaluation. Impact, in particular short-term impact that characterises almost all the research/evaluation to date with respect to PBR, is an important question, but just one of many possible questions that can be asked about PBR approaches. As discussed above, information about impact in terms of achieving objectives without information about the factors responsible and the circumstances that would be needed to reproduce this effect is not very useful or actionable. In many cases, it may be best to determine the relevance of a PBR approach and factors

associated with meaningful implementation before undertaking a costly impact evaluation.

4.4 The following are questions for evaluation of PBR that as of yet have not received sufficient (or any) attention. Many of these questions follow the discussion in the above sections discussing what is already known or not about the effectiveness of PBR schemes. As identified in Section 2.2, the various reviews and critiques consistently identify many of these questions for future research and evaluation. This of course is not intended to represent a complete list of all possible or relevant questions that can be asked about specific PBR approaches, or about PBR schemes more generally.

Outstanding questions for evaluation identified by Pearson (2011)

- What settings are best suited to results-based approaches?
- What else must take place together with the results-based approach, as part of a package of approaches?
- What factors are responsible when results are not achieved (e.g. structural barriers associated with working in poor countries and other external factors beyond the control of the agent)?
- To what extent is there consistency with the principles of aid effectiveness, such as Paris Declaration principles?
- What are the incentives, and, costs, faced by various agents?
- Is there any evidence that benefits are, or are likely to be, sustainable?
- What is the impact on equity?
- What are the risks faced by results based approaches (e.g. to what extent are reported results real)?
- Are the targets used contributing to the desired outcomes?
- How can one build up the systems and promote a supportive results-oriented culture?

4.5 The following discussion highlights what arguably are the five most pressing questions concerning PBR, and then indicates some other questions, also of importance, that have been identified by reviews and critiques of research to date.

Five key evaluation questions about PBR

1. Under what sets of circumstances is an incentive approach appropriate? This is a much more important, and useful, question than to ask merely if PBR "works" or not. As Macq et al. (2011) put it:

> "Rather than searching for the impossible proof of whether PBF [Performance-based Financing] works or not, we should instead try to learn useful lessons from experiences. … The focus of PBF assessment should be on "why" and "how" the intervention works."

4.6 Almost certainly, some types of PBR may "work" in some sets of circumstances, but not in others. This suggests a shift in emphasis, with less attention to "does PBR work?" and more on identifying the circumstances under which it might be a more effective and appropriate strategy to apply than alternatives. *This is very much a question of impact, not merely of process.*

4.7 The need for such research and evaluation arises strongly from the experiences to date and has been highlighted in various reviews and critiques. For example, one of the conclusions of the most recent Cochrane Collaboration review (Witter et al., 2012) is the need to "uncover the mechanisms by which the intervention may or may not

work, and to probe the motivational effects which are intended to be at the core of the intervention." Another review by Eldridge (2008) concluded that: "The existing literature would be helped greatly by case studies detailing the potential advantages and disadvantages of PBP, determining the most influential factors for success, and recording experiences in field situations in some detail."

4.8 Indeed, one might well argue that it is premature to attempt an impact evaluation until one has first identified the actual mechanisms by which incentives can work and the conditions needed for this to take place. For example, an underlying assumption of PBR approaches is that financial incentives are motivating. But motivating for whom? Under what circumstances? How does this complement or detract from competing motivations (intrinsic motivation, sense of professional responsibility, perceived importance to provide other services that are viewed as more essential than those specifically contracted for)?

4.9 A related consideration concerns identifying and describing the context, including identification of enabling and inhibiting factors – in the general environment as well as other forms of technical assistance, management support, and the like, as these factors may determine if there is any possibility for the intervention to work. Clearly, the context in which contracting out is implemented and the design features of the interventions are likely to greatly influence the chances for success (e.g., as expressed by Liu et al. 2008). Witter et al. (2012) have observed:

> "Paying providers for performance is clearly premised on the assumption that … a change in behavior on the provider side is required. If, however, the barriers are more connected with demand-side factors (such as low affordability of services), then paying for performance for providers alone will not be effective."

Developing an understanding of the conditions by which PBR approaches may work represents the most pressing need for research and evaluation attention.

2. What is the optimal size and nature of the incentive required? What does this depend upon? How much of an incentive is needed in order to significantly influence behaviour? What are the differences in effects from incentives provided to organisations from those provided directly to individuals?

4.10 There is very little information about this very important question. Too small an incentive may not be sufficient to change behaviour; too much may represent payment for something that would have occurred anyway (with the potential of unintended effects). Large payments may also be more likely to lead to gaming and distortion. There is some speculation in the literature regarding the size of incentives, but without actual evidence. For example, if the financial incentive represents a small bonus on top of a health professional's base salary, this may be less of a motivating factor than when the salary of a health professional is very low (as is frequently the case, even with highly trained doctors, in many low-income countries) and the incentive represents a major portion of total remuneration.

4.11 Similar considerations may apply when the incentive is paid to an institution or organisation (e.g. health care organisation, a government unit, an NGO or private sector provider) rather than to an individual practitioner. Some writers have suggested

that the incentive may have limited effect if the organisation can or also does receive funding from other sources. Another consideration concerns modality of payment, with some PBR approaches (the Cash on Demand model as proposed by the Center for Global Development is most explicit about this) withholding all payment until after the desired results have been documented. How does this affect organisations with limited resources, which do not have reserves to cover up-front costs or to bear the risks involved? Also, to what extent does it matter if the financial incentive is paid to an institution/organisation instead of to an individual practitioner (or group of practitioners)?

4.12 The literature indicates that the ability to obtain results depends upon many factors, many of which are beyond the control of the provider (individual or organisation). In situations such as this, no amount of incentive may be able to influence behaviour. Is it possible to identify those situations where external barriers such as these are most likely?

4.13 The optimal size and form of incentives relate to the questions posed above: an understanding regarding how incentives actually work in practice and when they may or may not be sufficiently motivating to induce changes in behaviour. It may be helpful to explore blockages and barriers to improved performance, and the extent to which incentives may be able to help overcome these. It may also be helpful to ask if a change in the size or nature of the incentive, or different and less costly alternatives, could have produced similar outcomes.

3. What is the cost effectiveness of PBR approaches? There has been almost no attention to this, or to other considerations of efficiency. Examples of potential evaluation questions may include: What are the full costs of PBR approaches? To what extent do the benefits justify these costs? Are there ways in which the costs can be minimised without affecting outcomes?

4. More attention to potential unintended effects. As discussed earlier, there is considerable evidence indicating that negative, unintended effects of incentive approaches *are* very possible. When organisations and individuals are rewarded for providing certain results, it is natural that they may game the system to enhance these indicators, which may include neglecting other pressing needs. There, however, has been limited attention to identifying the circumstances when these are most likely to occur, under what conditions negative effects might outweigh intended benefits, and in identifying ways in which unintended negative effects can be minimised. Given the strong evidence about the potential of unintended effects, it may even be considered irresponsible to proceed with implementation of a PBR approach without, at a minimum, building in identification of potential unintended effects into the programme and evaluation design.

5. More meaningful comparison with alternative models of enhancing performance, including comparison among the effectiveness of alternative PBR approaches as well as with other alternative models (i.e. without the use of incentives) for enhancing results. Most of the impact evaluations undertaken have used as the counterfactual comparison of a particular PBR approach with the status quo (often described as input-based financing). Yet as has been identified, there are a variety of potential models of PBR. There are also many other potential means of enhancing performance. As many writers have observed, the development landscape is changing rapidly. "Traditional" approaches of even a couple of years ago may be well out of date by the time an impact evaluation is completed, if not much sooner. Research designs

based upon such comparisons, in particular assuming that little will change in the control or comparison groups, may be lacking in relevance. This consideration is closely related to that of cost effectiveness.

Some other important evaluation questions

4.14 The following are other potential evaluation questions concerning PBR approaches, largely unaddressed by research to date and identified in the literature.

4.15 **What is the actual nature of the intervention?** The research undertaken to date highlights the importance of identifying the extent to which the intervention was implemented as planned and expected, and reasons for changes. While this may seem obvious, much of the research indicates that little is actually known about the actual intervention and how it has been implemented in practice – essential information both to be able to understand what was done, and to provide for attribution. This is a recurring theme in the reviews and critiques assessed. For example, the Norad evaluation of the Health Results Innovation Trust Fund (HRITF) impact evaluations to date highlight this as a major need:

"Impact evaluation is expected to measure the with and without situation through a set of indicators, but many things may not go exactly as planned or as assumed that may have an impact on the final results and that may not affect all intervention and control sites in the same way. Documentation is expected to capture processes and unexpected changes that might affect final results"

4.16 What is the **impact on equity?** This includes, but goes beyond, considerations related to gender equality. Do PBR approaches assist those most in need? Or do the benefits go to those easiest to reach? Some studies (e.g. Pearson, 2011; Witter et al., 2012) suggest that lack of a pro-poor approach may reduce focus on those most in need, such as those living in remote areas.

4.17 To what extent, and under what circumstances, are **PBR approaches compatible with other development principles and desired outcomes**, such as those in the Paris Declaration? There are suggestions in the literature that PBR may not support national ownership, a partnership approach, and use of national systems (e.g. when separate/parallel verification systems are established), and may not be compatible with harmonisation if the PBR approach is not integrated with approaches of other donors.

Research agenda: A question of institutional strengthening? (some questions posed by Canavan et al., 2008)

- What are the effects on health system and does PBF have implications for wider health systems performance? What are the unexpected effects or outcomes of PBF?
- Does the PBF approach really change the behaviour of institutions and individuals or are we going to see a drop of performance (to previous level) if we erase the incentives?
- If there is evidence of a sustained behavioural change, could a phase out strategy be possible and could there be a switch to other financing mechanisms?
- Should PBF be seen as a permanent way of financing/organizing a health system?
- Should the subsidy structure (e.g. with escalating subsidies) be different for activities with 'natural' different coverage rates such as the high achievement in Rwanda in terms of EPI (Expanded Program on Immunization) and the very low coverage with family planning?

4.18 **To what extent do PBR approaches lead to improvements in quality?** To date, most PBR evaluations have looked at changes in quantity, such as the numbers of people who receive given services, partly because these generally are easier to measure.

4.19 **What are the effects of incentives on the broader health (or other institutional) systems?** To what extent are these positive or negative? For example, does this lead to more focus on outcomes or on needs that have been neglected? Or are other needs that are not incentivised neglected? To what extent do incentives increase overall motivation and commitment, or take the place of intrinsic motivation and professional commitment? Are there certain institutional contexts where incentives may be more (or less) appropriate or effective?

4.20 Perhaps most importantly of all, **what are the long-term effects of incentive approaches?** As indicated previously, to date the most optimistic conclusions from existing research is that, perhaps, a change in short-term behaviours may result. There is insufficient evidence indicating if any such changes can be sustained over a longer period of time, or if these can lead to more significant changes in the lives of communities and people. As indicated above, most approaches to PBR that have been tried thus far are of a demonstration or pilot nature, with additional attention and resources that likely would not be possible on an ongoing basis. What can be learned from this experience with respect to the potential for scaling up or using such techniques in different types of situations?

4.2 Some key gaps

RBA (incentives for governments)

4.21 Almost all the research attention to date has gone to consideration of RBF types of initiatives, with incentives to service providers. There has been very limited attention to approaches providing incentives to governments directly (except at the sub-national level, where the unit in question is a service provider). Yet this may be of greater potential importance and has been identified as a major priority for DFID.

4.22 A major exception is the evaluation of the GAVI Immunisation Programme (ISS). This programme provided for cash payments to countries based upon the number of children immunized with DTP (Diphtheria, Tetanus, Pertussis vaccine). While the two evaluations that have been undertaken of this initiative suggest that the ISS programme overall has had a significant impact in increasing the numbers of children immunised, they point out that the incentive aspect represents just one component of a significant programme with many other aspects, including technical support and often extensive funding.[20] An evaluation of the Global Fund (Macro International Inc., 2009) also found that use of the incentive mechanism was inextricably linked with other aspects of what was mainly a large-scale funding programme.

4.23 There seem to be various reasons why RBA approaches thus far have received limited evaluation attention. One reason is that there have been far fewer interventions of this nature. A key factor may be that the predominant research design employed – with-

[20] Although the first GAVI evaluation concluded from its quantitative analysis that "receiving rewards has little effect on performance", and that apparent gains resulted from faulty and misleading indicators.

and-without counterfactual designs, are not generally applicable to programmes of a very different and more complex and generally universal application.[21] Also, interventions involving national governments often represent complex, multi-faceted policy approaches, rather than discrete, localised interventions such as those characterised by most RBF interventions. Pearson and Vähämäki et al. locate such types of initiatives in the larger spectrum of results-based approaches, and suggest that these types of initiatives may be closer in nature to initiatives such as General Budget Support and different in kind from PBR projects involving discrete service providers. Also, as Section 4.3 touches upon, there are other potential methodologies that could be used for the evaluation of such types of programmes.

Other domains beyond the health sector

4.24 As Section 2.1 indicates, almost all the research and evaluation to date has concerned PBR initiatives in the health area, with just one evaluation identified concerning infrastructure (provision of water and sanitation connectivity by local utilities) and a couple starting to look at education. One cannot assume that findings with respect to initiatives in the health sector, given its own peculiarities, necessarily would be transferable to other domains.

4.25 There now does seem to be more interest in results-oriented approaches across all development domains. This may represent an opportunity to build in appropriate evaluation into the design of such initiatives that can be informed by, but not necessarily copy, what has been tried in the health area.

DFID Pilots

DFID is currently planning comprehensive evaluations of three DFID-sponsored PBR pilots, in Ethiopia, Rwanda, and Uganda. The first two of these will be evaluating RBA initiatives in the education section, which helps to address the gaps discussed in the text.

Given the preliminary status of the evaluation plans, it is premature to form concrete judgements. Nevertheless, to a significant extent, collectively they do appear to be addressing many of the issues identified in this study, and addressing some of the key questions discussed. This is particularly the case for the Ethiopia pilot evaluation, which is proposing complementary methods to address some key questions, such as understanding why the approach works (or not) in combination with quantitative data on value added, value for money, unintended consequences, etc. The Rwandan evaluation will be using a theory-of-change model developed in conjunction with the government. Both evaluations appear to have a significant process evaluation component.

But some questions about the draft evaluation approaches still remain. For example, all will use the status quo as the counterfactual, which is unlikely to remain stable over the three-year time period for the evaluation, rather than other alternative approaches. The proposed impact designs all have weaknesses that, unless complemented by other data sources (e.g. other forms of quantitative as well as qualitative data, particularly those that can help identify potential mechanisms influencing changes along the results chain), may severely limit the ability to draw valid conclusions. Further, all the proposed results indicators are subject to distortion, with just one of these evaluations planning on examining unintended effects explicitly.

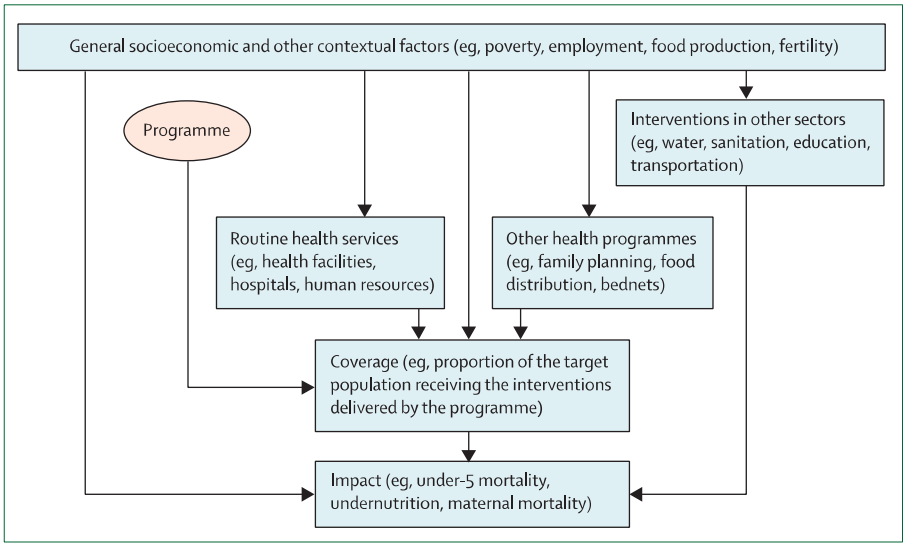
4.3 Methodological implications and alternatives

4.26 The following are some ideas and general suggestions with respect to evaluation approaches that may be appropriate for evaluating future PBR initiatives. It is beyond the scope of this report to discuss these methodological considerations in any detail, as these are well documented in the evaluation literature.

[21] If programmes can be phased, certain forms of quasi-experimental designs may be possible under certain circumstances, although contamination of comparison or control groups represents a significant danger, and this approach can be controversial from programmatic and ethical perspectives.

Articulate the theory of change

4.27 The theory of change (or intervention, or programme logic) is not an evaluation method per se, but an essential planning tool useful for virtually all forms of evaluations. The DAC EQS emphasise the need for articulation of the theory of change early in the evaluation process. It can assist with evaluability and in identifying what types of evaluation questions might be appropriate to address at given points of time in the lifecycle of a programme or project. It can spell out the assumptions linking inputs and activities, outputs, and outcomes and impacts at all levels, which can aid both in evaluation design and in programme planning. There are theory-of-change models that can take into account complexity, including the interaction of many different factors internal and external to the programme design. As discussed above, there are few PBR approaches that are "pure", and identification of other facilitating and intervening factors would seem particularly relevant. Ideally, a participatory approach should be used in the articulation of the theory of change.



4.28 Adjacent is an example of a theory-of-change model developed by Victora et al. (2010) to illustrate how the PBR intervention ("the programme") would need to interact with other factors to result in impact (in this case, improvements to under-five mortality, under-nutrition, maternal mortality).[22]

A more balanced mixed methods approach

4.29 Every possible evaluation method has strengths and limitations. It is generally recognised that a combination of methods can help address at least some of the inherent limitations of any method used on its own. As discussed previously, almost all evaluations to date of PBR approaches have been research studies using experimental or statistical impact evaluation designs. There is clearly a need for continued research of this nature, although hopefully informed by the theory of change and other forms of evaluation findings, so that impact evaluation can be as focused as possible. For example, it only makes sense to apply RCTs when randomisation in practice will be feasible and when meaningful counterfactuals are possible. This also rarely would be the design of choice in complex situations with many factors interacting and where there are uncertainties at play. There are other possible approaches that can be considered when this is not the case.

[22] This is a copy of Figure 1: Outline of factors affecting maternal and child health and nutrition appearing in Victora et al.

4.30 Given the almost exclusive focus on impact evaluations using experimental and statistical approaches, it would seem appropriate to complement these and to provide greater balance through utilisation of a variety of other forms of evaluation that can address a wider variety of evaluation questions than is possible with a "rigorous" impact evaluation.

4.31 Most of the systemic reviews call for process evaluation, as discussed below, but at the same time also for more "rigorous" research, along the same lines as the studies that have already been carried out. Perhaps this is to be expected, given that those making such suggestions are largely health sector researchers who generally are unaware of the evaluation field and the wide range of other potential methodologies in use by evaluators, including other models of causality and means of establishing contribution.

Process evaluation

4.32 Even though the authors of most of the research studies, as well as of the systemic reviews, largely have a background in experimental research, a very common observation and recommendation, for reasons discussed throughout the report, is the need for process evaluation. For example, as Witter et al., 2012 have indicated: "Robust effectiveness evaluations should be complemented by in-depth process evaluations to uncover the mechanisms by which the intervention may or may not work, and to probe the motivational effects which are intended to be at the core of the intervention." Process evaluation has been identified as essential information to help interpret impact evaluation findings.

4.33 Process evaluation of some form (which can be both quantitative and qualitative in nature, and potentially explore a wide range of questions such as those identified above) is a basic prerequisite in the evaluation process. In many cases, it may be premature to undertake impact evaluation until it can first be established what is actually taking place. While there is an increasing recognition of the need for such complementary forms of data and evaluation about the processes and mechanisms at place, there sometimes can be a tendency to treat these as minor sideshows to the main event.[23]

Formative evaluation

4.34 Formative evaluation is undertaken while an initiative is still under way[24], to provide feedback and guidance about how the undertaking can be enhanced. It can include both process and impact components. For example, this can help explain what is taking place, reasons for changes, identify if things are moving in the right direction or not, and the likelihood that desired outcomes can be achieved.

[23] For example, the HRITF Guide: *Impact Evaluation in Practice* indicates that qualitative data, monitoring data, and process evaluations are complementary to impact evaluation and need to be able to inform and interpret the results from impact evaluations. But this guide that is quite comprehensive in other respects hardly comes back to this need or suggests how it should be done and how these data can be used.

[24] In comparison with summative evaluation, which makes a judgement about an initiative's merit and worth after completion, or at least after it has been in existence for some time.

4.35 Formative evaluation can take a variety of forms. For example, ongoing monitoring and evaluation (M&E), largely undertaken by the programme itself, can be considered one form of formative evaluation (or representing one form of data that can aid in making formative assessments). Frequently, particularly for large-scale initiatives or when there may be questions about how the initiative is proceeding, independent evaluation may be needed to provide more information about the nature of the intervention and how it is working.

> Targeted evaluations that can provide useful information need not be very expensive, or take a long time to implement. For example, Kalk et al. (2010) made use of a number of semi-structured key informant interviews with high-ranking government staff, hospital management and administrative staff, doctors and nurses and patients, along with a document review, to identify the actual effects of incentives used in Rwanda, including motivational aspects, and were able to document various confounding factors and negative effects such as gaming.

4.36 Traditional research approaches, such as most of those identified in this study, can take years to complete, and many more years before there is a sufficient body of reliable evidence. There is a need for evaluation that can help inform both policy and practice in the short and medium term, on a much tighter timeframe.

Theory-based approaches to evaluation

4.37 Theory-based approaches to evaluation use the theory of change as a basis for identifying plausible, testable hypotheses, including causal assumptions, which can then use a variety of methods, qualitative and quantitative, for testing these. Theory-based approaches are in increasingly common use in evaluation, in development as well as in other domains. Theory-based approaches typically will explore and confirm the existence of causal processes or 'chains'. Complementary approaches going under names such as General Elimination Theory, and the modus operandi method are consistent with a theory-based approach. The former involves identifying potential rival plausible explanations for the findings other than the intervention itself, and then obtaining evidence needed to rule them out. The latter involves tracing the 'signature', where one can trace an observable chain of events and identify the mechanisms that link to the impact, which can result in generative causality.

4.38 One example of such an approach was used by Ssengooba et al. (2012) to explore factors responsible for the failure (identified through a World Bank sponsored RCT) of a performance-based contracting approach in Uganda. This evaluation used what the authors describe as an "open box" approach, using a theory-based and prospective case study design, grounded in systems thinking and complexity theories. The aim of this approach was to be able to "open the black-box and build plausible explanations derived from empirical observations". This method was used to explore factors associated with the failure of the programme, such as inadequate design, which in turn led to hasty adaptations and something implemented in practice quite different from as intended and not attuned to the needs of the low-income target populations.

4.39 Victora et al. (2010) describe their use of what they refer to as a "platform approach" in Mozambique, drawing upon a variety of different sources of data. This approach "uses the district as the unit of design and analysis; is based on continuous monitoring of different levels of indicators; gathers additional data before, during, and after the period to be assessed by multiple methods; uses several analytical techniques to deal

with various data gaps and biases; and includes interim and summative evaluation analyses." Victora et al. propose this platform approach as an alternative to traditional designs, such as those comparing a few programme districts (or populations) with a handful of others that do not have that particular programme when programmes are being scaled up across many different settings. They also propose this approach when the comparators are likely to have, to some degree, initiatives similar to those under scrutiny; when an important question is to understand why certain programmes are implemented in some areas rather than others; and to be able to explore which of the various programmes or delivery approaches implemented by different partners works best in a given context.

Alternative methods and designs

4.40 In a recently released report commissioned by DFID, Stern et al. (2012) identify a range of alternative methodologies that can be used for impact evaluation. In keeping with the scope of their study, they specifically identify a range of theory-based, case-based, and participatory approaches. While the focus of this work is on evaluation of the full spectrum of development aid and not on PBR specifically, they do identify a range of methodological approaches that might be particularly worthy of more attention in this area. As they observe, some of the most potentially useful approaches to causal inference, including approaches that can take into account multiple causality and configurations such as are typical with respect to PBR, are not generally known or applied in the evaluation of development aid.[25]

4.41 Also, as they observe, most development interventions can be best seen as contributing factors rather than as "causes". They "work" as part of a causal package in combination with other 'helping factors". This seems particular relevant, given the research findings to date, with respect to PBR. Incentives almost never "work" in isolation of other factors. Contribution analysis in particular might be particularly appropriate for evaluation of many PBR schemes.

Realist evaluation

4.42 Realist evaluation (see for example Pawson, 2006, and Pawson et al., 2012) is an approach that identifies what works for whom in what circumstances. This matches very directly with what has emerged from this study as the major need: to identify what forms of PBR in development might be most appropriate. This acknowledges that it is the wrong question to ask if PBR or, more broadly, incentives "work" or not. A more appropriate and useful question is to ask under what circumstances they can work, and under what circumstances alternative approaches might be more appropriate.

4.43 Realist evaluation puts a premium on explanation, which can aid in programme design even without more extensive information. Realist evaluation identifies generalizable conclusions based upon various configurations of context, mechanisms, and outcomes that can be applied to other settings. Realist evaluation searches for and refines

[25] Picciotto (2012) has a useful discussion of the limitations of experimental designs for impact evaluation and the existence of various alternative approaches. Bamberger, Rugh and Mabry (RealWorld Evaluation) also provide many useful and practical ways of undertaking evaluation in challenging circumstances.

explanations of programme effectiveness in various context-mechanism configurations, in a way that can assist in learning from pilot operations, in informing policy, and in applying findings to other settings.

Concluding observations about methodology

4.44 Given that *all* methods have strengths and limitations, evaluation of PBR should use a mix of methods, applying the most appropriate (and simplest, and most cost-effective) method to address questions of most interest. Having said this, a realist evaluation approach seems to be very closely suited to the range of questions and challenges concerning PBR that have been identified in this study and thus may warrant being given priority in the choice of potential evaluation designs.

5 Overall conclusions and recommendations

5.1 The importance of an outcome (or results) orientation, focusing on the actual benefits arising rather than on inputs and services provided, for all public services including development aid, is well recognised. By paying for actual, verifiable results achieved, PBR has been promoted as a means of ensuring the achievement of results.

5.2 Nevertheless, there is very limited evidence about the effectiveness of PBR approaches. Perhaps the most optimistic conclusion that can be drawn from available evidence is that contracting out may increase access and use of health services in the short term rather than broader health outcomes. Implementation of PBR in practice has encountered severe challenges and difficulties, and as the recent Vähämäki et al. review indicated: "is being questioned by practitioners and researchers".

5.3 PBR represents one possible (or a collection of) means of supporting a results-oriented approach to development aid. It will not be suitable in all contexts, and even strong proponents of PBR acknowledge that PBR is not a panacea. While there is very limited evidence about the circumstances under which PBR may work, it is clear that it can only be effective as part of a package of approaches rather than just on its own. It should only be attempted when improved performance is within the control of management or staff and not impeded by other barriers, where financial incentives are most likely to motivate the necessary actions to produce the desired results, and where incentives could help focus attention on neglected areas that otherwise would not receive attention.

5.4 The overall quality of research and evaluation of PBR approaches has been identified as weak. Actual implementation of sophisticated impact evaluation designs in practice has proved very challenging, to the extent that systemic reviews have identified biases and methodological limitations in almost all the studies conducted to date, severely limiting the ability to draw valid conclusions. There has been limited attention to many of the generally accepted evaluation criteria and standards. More profoundly, many important questions about PBR that could help inform policy and practice have not been addressed. There is, however, a range of potential evaluation methodologies that may be well suited to addressing many of the questions about PBR.

Implications and recommendations for policy and programming

- There is a need for some healthy scepticism, with recognition that the value of PBR is, at least as of yet, unproved, with the likelihood of unintended negative effects.
- One should embark upon a PBR approach only after considering its potential impact and cost effectiveness in comparison to other possible strategies. These considerations should recognise that there are many situations where PBR would not be appropriate, such as when barriers to results are beyond the control of those expected to bring about changes, when other innovations or forms of support are required, or lack of motivation is not the barrier to improved effectiveness.
- Potential unintended effects should be anticipated and articulated at the design stage, and monitored on an ongoing basis.

- Each application of PBR should be tailored to the particular situation, recognising that one model is unlikely to be appropriate equally across the board.
- Consideration should be given to the meaning of a 'hands off' approach. At a minimum, it is appropriate to insist upon adherence to ethical guidelines and standards, including those of the country or community where the project is being implemented and to principles of aid effectiveness that DFID has agreed to. A hands-off approach should not be seen as a barrier to independent evaluation (although the nature of the evaluation, so that it does not necessarily become intrusive or dictate the form of programme approach, should be open to negotiation).
- In particular given the challenges identified to implementation of PBR, programmes should include an internal M&E capability.

Implications and recommendations for evaluation

- Evaluation should start by identifying priority questions that can best inform policy and programming decisions, and only then consider potential methodologies that can best address these questions in a timely and cost-effective manner.
- The most important question for evaluation to address should not be "does it work?", but to identify the mechanisms and sets of circumstances under which PBR approaches can most likely result in behavioural change leading to changes in outcomes, recognising that this is very much a question of impact.
- Evaluation should explore unintended consequences of incentive approaches, identifying when these are most likely to occur, when these may offset benefits and how these can be minimized.
- Other potential evaluation questions that should be considered include (but are not limited to): cost effectiveness and comparison with other potential approaches and strategies, appropriate size and nature of incentives, sustainability, and equity.
- There is a particular need for evaluation to explore and to describe the *process* by which PBR initiatives are implemented in practice, and the reasons why changes from the original conception may be needed. Methods can range from routine M&E to more comprehensive ad hoc evaluation studies.
- A mixed method approach should be taken, involving both quantitative and qualitative methods. In all cases, the theory of change should be articulated, using this as a basis to identify what types of questions can be evaluated at given points in time and to aid the choice of methods.
- Priority should be given to methods that can provide explanation. In this regard, theory based and realist evaluation approaches should be given special consideration.
- Given the complex context in which PBR schemes work, always in combination with other factors, it may be more appropriate to use a contribution analysis approach rather than aim, perhaps unrealistically, to identify linear cause-and-effect.

Annex 1 – Key informants and databases

Key informant contacts

Agency
World Bank
University of California Berkeley*
Jameel Poverty Action Lab (J-PAL)
The GAVI Alliance
African Development Bank (AfDB)
OECD
NORAD
University of California San Francisco
Norwegian Knowledge Centre for the Health Services
University of Leeds
Centre for Global Development (UK)
Makerere University, Uganda
University of Rome
International Initiative for Impact Evaluation (3IE)*
OECD/DAC Evaluation Network contact list (c. 85 persons)

*Contacts with no response.

Databases searched

- International Initiative for Impact Evaluation (3IE)
- Results-Based Financing (RBF) for Health (rbfhealth)
- The Gavi Alliance
- Jameel Poverty Action Lab (J-PAL)
- Center for Global Development
- World Bank
- The Lancet
- Google Scholar

Annex 2 – Evaluations, reviews, and other documents considered

Reviews and syntheses

Citation	*Domain*	*Type of intervention*
Brenzel, Logan. (2009). Taking Stock: World Bank Experience with Results-Based Financing (RBF) for Health. World Bank (HDNHE) internal unpublished study.	health	CCTs (mainly)
Canavan Ann, Toonen, Jurriën, and Elovainio, Riku. (2008). KIT Development Policy & Practice.	health	PBF
Eldridge, Cynthia and Natasha Palmer. (2009). Performance-based payment: some reflections on the discourse, evidence and unanswered questions. *Health Policy and Planning*.24:160–166	health	donor to government, within the public sector, government/donor to non-state provider, non-state provider and health workers
IEG (Independent Evaluation Group). 2011. *Evidence and Lessons Learned from Impact Evaluations on Social Safety Nets*. Washington, DC: World Bank.	Non-contributory programmes targeting the poor such as health and education, land redistribution, and microfinance	non-contributory programs that target the poor and vulnerable, e.g. CCTs, other supports including workfare, school feeding, educational fee waivers, thus not really PBR
Lagarde M, Palmer N. (2009). The impact of contracting out on health outcomes and use of health services in low and middle-income countries. Cochrane Database of Systematic Reviews, Issue 4. Art. No.: CD008133. DOI: 10.1002/14651858.CD008133.	health	contracting out
Lagarde, M. and Palmer, N. The impact of health financing strategies on access to health services in low and middle income countries. Cochrane Database of Systematic Reviews 2006, Issue 3. Art. No.: CD006092. DOI: 10.1002/14651858.CD006092.	health	contracting out

Liu, Xingzhu, David R Hotchkiss, and Sujata Bose. (2008). The effectiveness of contracting-out primary health care services in developing countries: A review of the evidence. *Health Policy and Planning*. 23:1–13.	health	contracting
Ministry of Justice. Summary of common PbR analytical issues. (2012). Payment by Results Analytical Sub-Group.	cross government (UK)	all PBR
Oxman AD, Fretheim A. (2008). An overview of research on the effects of results-based financing. Report Nr 16-2008. Oslo: Nasjonalt kunnskapssenter for helsetjenesten. Study carried out for NORAD.	health	all
Oxman, Andrew D and Atle Fretheim. (2009a). Can paying for results help to achieve the Millennium Development Goals? Overview of the effectiveness of results-based financing. *Journal of Evidence-Based Medicine*. **2.** 70–83.	health	all (recipients/CCTs; health professionals, organisations
Oxman, Andrew D Oxman and Atle Fretheim. (2009b). Can paying for results help to achieve the Millennium Development Goals? A critical review of selected evaluations of results-based financing. *Journal of Evidence-Based Medicine*. **2.** 184–195.	health	all
Pearson, Mark, Martin Johnson & Robin Ellison (2010). *Review of Major Results Based Aid (RBA) and Results Based Financing (RBF) Schemes: Final Report*. Commissioned by Aid Effectiveness and Accountability Department, DFID.	all	all
Pearson, Mark. (2011). *Results based aid and results based financing: What are they? Have they delivered results? London: HLSP Institute.*	all	all
Vähämäki, Janet, Martin Schmidt, and Joakim Molander. (2011). *Review: Results Based Management in Development Cooperation*. Sweden: Riksbankens Jubileumsfond.	all development	any form of RBM, interpreted very broadly

Witter S, Fretheim A, Kessy FL, Lindahl AK. (2012). Paying for performance to improve the delivery of health interventions in low- and middle-income countries . *Cochrane Database of Systematic Reviews*. Issue 2. Art. No.: CD007899. DOI: 10.1002/14651858.CD007899.pub2.	health (Providers of healthcare services)	mainly CCTs, RBF

Individual Evaluations

Citation	*Domain*	*Type of intervention*	*Type of evaluation*	*Geographic area*
Allsop, Terry, Robin Ellison, Larry Orr, Mark Pearson and Jawaad Vohra. (2012). Evaluation of the Pilot Project of Results-Based Aid in the Education Sector in Ethiopia: Inception Report. (DFID pilot)	education	RBA	Interrupted time series quasi-experimental design, plus qualitative data collection	Ethiopia
Averill, Kate, Kara Scally-Irvine, Deddi Nordiawan, Marcus Howard, and Jonathan Gouy. (2011). Independent Evaluation of the Water and Sanitation Hibah Program, Indonesia: Final Evaluation Report. AUSAid.	water/ sanitation/ infrastructure	PBR	Rapid Evaluation Appraisal Method	Indonesia
Basinga, Paulin, Paul J Gertler, Agnes Binagwaho, Agnes L B Soucat, Jennifer Sturdy, Christel M J Vermeersch. (2011). *Lancet* 377: 1421–28	health	P4P payments to facilities	RCT (districts)	Africa (Rwanda)

Basinga, Paulin, Paul J. Gertler, Agnes Binagwaho, Agnes L.B. Soucat, Jennifer R. Sturdy, and Christel M.J. Vermeersch. (2010). Paying Primary Health Care Centers for Performance in Rwanda. Policy Research Working Paper 5190. The World Bank.	health	P4P payments to facilities	RCT (districts)	Africa (Rwanda)
Center for Global Development. (2012). Terms of Reference Proposed Methodology for a Process Evaluation of Results Based Aid (DFID Ethiopia pilot).	education	RBA	DFID pilot process evaluation (to be undertaken directly by CGD)	Ethiopia
CEPA LLP/ (2010). *GAVI second evaluation report. The GAVI Alliance*.	health	RBA	mixed	International
Chee, Grace, Natasha Hsi, Kenneth Carlson, Slavea Chankova, Patricia Taylor. September 2007. Evaluation of the First Five Years' of GAVI Immunization Services Support Funding. GAVI Alliance.	health	RBA	mixed qualitative + quantitative	International
DFID. (2012). Terms of Reference (TOR): Data Verification and Evaluation of Project of Results-Based Aid (RBA) in the Education Sector – Rwanda. (DFID Rwanda pilot)	education	RBA	prospective mathematical model + process evaluation	Rwanda
GAVI Immunisation Services Support (ISS) update. (2009). http://www.gavialliance.org/library/news/statements/2009/update-on-immunisation-services-support-%28iss%29/	health	RBA	mixed	
Kalk, Andreas Friederike, Amani Paul, amd Eva Grabosch. (2010). 'Paying for performance' in Rwanda: does it pay off? *Tropical Medicine and International Health*. 15(2), 182–190.	health	P4P	interviews + doc review	Rwanda

Macro International Inc. (2009). The five-year evaluation of the Global Fund to Fight AIDS, to Fight AIDS, Tuberculosis, and Malaria.	health	RBA	mixed	International
Martinez, Javier, Mark Pearson, Birte Holm Sørensen, Barbara James, and Claudia Sambo. (2012). *Evaluation of the Health Results Innovation Trust Fund*. Norad.	health	RBF, CCT (not COD or RBA)	RCTs	International
Mills, Anne. To contract or not to contract? Issues for low and middle income countries. (1998). *Health Policy and Planning. 13*(10), 32-40.	health	contracting of services to the private sector	review	Asia, Africa, Pacific (5 countries/ case studies implemented not as part of the study)
NU Health Programme. (2012). Inception Report (Final) [DFID Uganda pilot).	health	RBF		Uganda
Olken, Benjamin A., Onishi, Junko and Wong, Susan. Should Aid Reward Performance? Evidence from a Field Experiment on Health and Education in Indonesia. (2012). NBER Working Paper No. 17892. National Bureau of Economic Research. (plus two other versions of this same paper)	health + education	CCT (to individuals and block grants to villages)	RCT	Indonesia
Ssengooba, Freddie, Barbara McPake, Natasha Palmer. (2012). Why performance-based contracting failed in Uganda – An "open-box" evaluation of a complex health system intervention, *Social Science & Medicine*, Volume 75, Issue 2, Pages 377-383	health	contracting	theory-based case, presented as an alternative to black box RCTs	
Toonen, Jurien, Ann Canavan, Petra Vergee,r and Riku Elovainio (2009). Learning lessons on implementing performance based financing, from a multi-country evaluation kit. Royal Tropical Institute).	health	PBF	synthesis report drawing lessons from 4 country study reports	Africa

Victora, Cesar G, Robert E Black, J Ties Boerma, Jennifer Bryce. (2010). Measuring impact in the Millennium Development Goal era and beyond: a new approach to large-scale effectiveness. *The Lancet*. Published Online July 9, 2010 DOI:10.1016/S0140-6736(10)60810-0.	health		platform approach drawing upon multiple sources of data	Mozambique
World Bank. Project appraisal document. Second Punjab education sector project. (2012). World Bank internal document.	education	RBF	monitoring only	Pakistan

Other documents

Bamberger, Rugh and Mabry/ (2012). *RealWorld Evaluation*. 2nd Ed. Sage.

Basinga, Paulin, Serge Mayaka, and Jeanine Condo. (2011). Performance-based financing: the need for more research. Bulletin of the World Health Organization. 89:698-699.

Battye, Fraser and Paul Mason. (2012). Thinking about...evaluation and payment. Paper based upon presentation to the 2012 United Kingdom Evaluation Society annual conference.

Bemelmans-Videc, Marie-Louse, Jeremy Lonsdale, and Burt Perrin.(2007). *Making Accountability Work: Dilemmas for Evaluation and for Audit*. Transaction Publishing,

Birdsall, Nancy and William D. Savedoff. (2011). *Cash on Delivery: A new approach to foreign aid*. Rev. ed. Center for Global Development;

Birdsall, Nancy, Ayah Mahgoub, and William D. Savedoff (2010). *Cash on Delivery: A New Approach to Foreign Aid*. CGD Brief.

Brown, Jessica. (2008). Cash on Delivery Aid: Incentive Issues in a Multi-Model Aid System. Unpublished paper.

Building Evidence on RBF for Health 2011: Third Annual Impact Evaluation Workshop. Bangkok, Thailand. World Bank. http://web.worldbank.org/WBSITE/EXTERNAL/TOPICS/EXTHEALTHNUTRITIONANDPOPULATION/EXTHSD/0,,contentMDK:23151124~menuPK:2643950~pagePK:64020865~piPK:51164185~theSitePK:376793,00.html

Building Evidence on RBF for Health: Third Annual Impact Evaluation Workshop. (2011). Bangkok. http://web.worldbank.org/WBSITE/EXTERNAL/TOPICS/EXTHEALTHNUTRITIONANDPOPULATION/0,,contentMDK:23151124~menuPK:2643981~pagePK:64020865~piPK:51164185~theSitePK:282511,00.html

Chi-Man Yip, Winnie, William Hsiao, Qingyue Meng, Wen Chen, Xiaoming Sun. (2010). Realignment of incentives for health-care providers in China. *Lancet*, 375: 1120–30.

Economic Policy Research Institute (EPRI). (2011). Designing and implementing social transfer programmes. Chapter 15: Monitoring and evaluation. Available at http://epri.org.za/resources/book

England, Roger. (2000). Contracting and Performance Management in the Health Sector, Some Pointers on How to Do It. DFID Health Systems Resource Centre.

England, Roger. (2004). Experiences of contracting with the private sector A selective review. DFID Health Systems Resource Centre.

Evaluation of the Health Results Innovation Trust Fund. (2012). NORAD.

Evidence Review Team 1: Supply Side Financial Incentives. (2011). Evidence Synthesis Packet.

Evidence Review Team 2: Conditional Cash Transfers. (2011). Evidence Synthesis Packet.

Evidence Review Team 3: Demand Side Incentives. (2011). Evidence Synthesis Packet.

Fryatt, Robert, Anne Mills, Anders Nordstrom. (2010). Financing of health systems to achieve the health Millennium Development Goals in low-income countries. *The Lancet*, Volume 375, Issue 9712, Pages 419 - 426.

Gertler, Paul J., Sebastian Martinez, Patrick Premand, Laura B. Rawlings, Christel M. J. Vermeersch. (2011). *Impact Evaluation in Practice.* World Bank.

IDD and Associates. (2006). *Evaluation of General Budget Support: Synthesis Report*. DFID and seven other partner governments.

Ireland, Megan, Elisabeth Paul, and Bruno Dujardin. (2011). Can performance-based financing be used to reform health systems in developing countries? *Bull World Health Organ* 2011;89:695–698.

Lob-Levyt, Julian. (2009). Vaccine coverage and the GAVI Alliance Immunization Services Support initiative. *The Lancet*, 373*9659), Page 209

Loevinsohn, Benjamin. (2008). Performance-based contracting for health services in developing countries : a toolkit. The International Bank for Reconstruction and Development / The World Bank

Loevinsohn, Benjamin. (2008). Performance-Based Contracting for Health Services in Developing Countries: A Toolkit. The World Bank.

Macq, Jean and Jean-Christophe Chiem. (2011). Looking at the effects of performance-based financing through a complex adaptive systems lens. *Bull World Health Organ*. 9. 699–700.

Mcpake, Barbara and Elias E Ngalande Banda. (1994). Contracting out of health services in developing countries. *Health Policy and Planning*; 9(1): 25-30

Meessen, Bruno; Agnès Soucat, and Claude Sekabaraga. (2011). Performance-based financing: just a donor fad or a catalyst towards comprehensive health-care reform?. *Bull World Health Organ* [online]. Vol.89, n.2, pp. 153-156.

Montagu, Dominic, and Gavin Yamey. (2011). Pay-for-performance and the Millennium Development Goals. *The Lancet*. Vol. 377 April 23, 2011, 1383-1385.

Morgan, Lindsay, Alix Beith, and Rena Eichler. (2011). Performance-Based Incentives for Maternal Health: Taking Stock of Current Programs and Future Potentials USAIDHealth Systems.

Organisation for Economic Co-operation and Development, Development Assistance Committee (OECD/DAC). (2002). *Glossary of Key Terms in Evaluation and Results Based Management*. Paris, OECD.

OECD/DAC Network on Development Evaluation. (2011). *Evaluating Development Co-Operation: Summary of Key Norms and Standards*. 2nd Ed. Paris, OECD.

Olsen, Ingvar Theo. (2011). Result-based Financing in the Health Sector: Experiences from Norway and from low-income Countries. Internal Norad document.

Palmer, Strong, Natasha Wali, Lesley Abdul, and Sondorp, Egbert. (2006). Contracting out health services in fragile states. *British Medical Journal*, 332, 718-721.

Pawson, Ray and Ana Manzano-Santaella. (2012). A realist diagnostic workshop. *Evaluation*, 18(2) 176 –191.

Pawson, Ray. (2006). *Evidence-based policy: A realist perspective*. Sage.

Ravallion, Martin. (2012). Fighting Poverty One Experiment at a Time: A Review of Abhijit Banerjee and Esther Duflo's Poor Economics: A Radical Rethinking of the Way to Fight Global *Poverty Journal of Economic Literature*, 50:1, 103–114

Stern, Elliot, Nicoletta Stame, John Mayne, Kim Forss, Rick Davies and Barbara Befani (2012). *Broadening the Range of Designs and Methods for Impact Evaluations*, DFID Working Paper, 38. London: DFID.

Annex 3 – Examples of outcome measures

Measures used in the studies included in the Witter et al. (2012) Review[26]

Settings: Vietnam, China, Uganda, Rwanda, Tanzania, Democratic Republic of Congo, Burundi, Philippines

Outcomes	Impacts	Number of studies
Provider performance (quality of care)	The impact of performance-based financing on service delivery is highly uncertain. Four studies measured coverage of tetanus vaccinations among pregnant women, with mixed findings. Results from one study showed a small or no impact on tuberculosis case detection	5
Utilisation of services: antenatal care	The impact of performance-based financing on attendance rates for antenatal care is highly uncertain. The study results point in both negative and positive directions	2
Utilisation of services: institutional deliveries	Whether performance-based financing leads to an increase in institutional deliveries is unclear. A wide range of effect estimates are reported in the studies, including substantially larger increases in areas *without* PBF, to almost a 2-fold increase in areas *with* PBF.	4
Utilisation of services: preventive care for children, including vaccination	Performance-based financing may or may not lead to increased utilisation of preventive care services for children. One study found that attendance rates for children's preventive services doubled, but the impact on immunisation rates ranged from negative to positive across the 4 studies.	4
Utilisation of services: number of outpatients	Utilisation of services may increase as a consequence of PBF, but this has not been rigorously evaluated and the studies where this has been assessed have not yielded consistent results	4

[26] Copied from Witter et al., pp. 3-4.

Outcomes	Impacts	Number of studies
Patient outcomes	The impact of performance-based financing on patient outcomes was evaluated in only 1 study. The results were inconsistent across the 4 measures that were used in the study: performance-based financing seemed to have an impact on rates of wasting and General Self Reported Health in this study, but not on CRP levels or on anaemia rates	1
Unintended effects	Only 2 studies reported on unintended effects - in both studies the authors voiced concerns 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	2
Resource use	PBF payments tend to increase facility revenues and to increase staff pay, but their impact on wider resource use indicators, such as other funding sources, patient payments and efficiency of service provision are not yet established	8

DEPARTMENT FOR INTERNATIONAL DEVELOPMENT

DFID, the Department for International Development: leading the UK government's fight against world poverty.

Since its creation, DFID has helped more than 250 million people lift themselves from poverty and helped 40 million more children to go to primary school. But there is still much to do.

1.4 billion people still live on less than $1.25 a day. Problems faced by poor countries affect all of us. Britain's fastest growing export markets are in poor countries. Weak government and social exclusion can cause conflict, threatening peace and security around the world. All countries of the world face dangerous climate change together.

DFID works with national and international partners to eliminate global poverty and its causes, as part of the UN 'Millennium Development Goals'. DFID also responds to overseas emergencies.

DFID works from two UK headquarters in London and East Kilbride, and through its network of offices throughout the world.

From 2013 the UK will dedicate 0.7 per cent of our national income to development assistance.

Find us at:
DFID,
1 Palace Street
London SW1E 5HE

And at:
DFID
Abercrombie House
Eaglesham Road
East Kilbride
Glasgow G75 8EA

Tel: +44 (0) 20 7023 0000
Fax: +44 (0) 20 7023 0016
Website: www.dfid.gov.uk
E-mail: enquiry@dfid.gov.uk
Public Enquiry Point: 0845 300 4100
If calling from abroad: +44 1355 84 3132