Health Policy and Planning, 31, 2016, 1297–1309
doi: 10.1093/heapol/czw045
Advance Access Publication Date: 28 April 2016
Review

OXFORD

Opening the 'black box' of performance-based financing in low- and lower middle-income countries: a review of the literature

Dimitri Renmans[1,2,*], Nathalie Holvoet,[1] Christopher Garimoi Orach[3] and Bart Criel[2]

[1]Institute of Development Policy and Management, University of Antwerp, Belgium, [2]Department of Public Health, Institute of Tropical Medicine, Antwerp, Belgium and [3]School of Public Health, Makerere University, Uganda

*Corresponding author. Institute of Development Policy and Management, University of Antwerp, Lange Sint-Annastraat 7, 2000 Antwerpen, Belgium E-mail: dimitri.renmans@uantwerpen.be

Accepted on 21 March 2016

Abstract

Although performance-based financing (PBF) receives increasing attention in the literature, a lot remains unknown about the exact mechanisms triggered by PBF arrangements. This article aims to summarize current knowledge on how PBF works, set out what still needs to be investigated and formulate recommendations for researchers and policymakers from donor and recipient countries alike. Drawing on an extensive systematic literature review of peer-reviewed journals, we analysed 35 relevant articles. To guide us through this variety of studies, point out relevant issues and structure findings, we use a comprehensive analytical framework based on eight dimensions. The review inter alia indicates that PBF is generally welcomed by the main actors (patients, health workers and health managers), yet what PBF actually entails is less straightforward. More research is needed on the exact mechanisms through which not only incentives but also ancillary components operate. This knowledge is essential if we really want to appreciate the effectiveness, desirability and appropriate format of PBF as one of the possible answers to the challenges in the health sector of low-and lower middle-income countries. A clear definition of the research constructs is a primordial starting point for such research.

Key words: Performance-based financing, developing countries, health financing

Key Messages

- The findings are often contradictive, therefore we need more research on the influence of the context and the design of the performance-based financing (PBF) scheme.
- A common definition of the construct of PBF, supported by proponents and opponents, is needed to better structure research and the debate on PBF. Such a definition should make clear which projects should be seen as PBF and which elements compose it.
- More research is needed on the exact mechanisms through which not only incentives but also ancillary components operate and how these mechanisms interact with each other.
- Local norms and values matter and should be taken into account when deciding whether or not (or how) to implement PBF.

Introduction

Performance-based financing (PBF) is becoming increasingly popular in the health sector in developing countries. The Health Results Innovation Trust Fund, sponsored by the British and Norwegian governments, and managed and supported by the World Bank, committed US$420 million to PBF projects all over the world between its inception in 2007 and June 2014. The International Development Association added another US$2.4 billion to these committed funds (World Bank 2014). With many more countries, non-governmental organizations (NGOs) and development agencies implementing PBF schemes in the health sector, this amount is only the tip of the iceberg. Research; however, has only recently been catching up and a lot remains unknown about the exact mechanisms that are initiated by PBF schemes (Eldridge and Palmer 2009; Ireland et al. 2011; Witter et al. 2012). Therefore a huge need exists for more knowledge on how (not) to implement a PBF scheme.

In order to take stock of what research has (or has not) taught us about PBF schemes and to inform policymakers intending to implement one, we will provide a review of the literature that unlike earlier reviews does not focus on results per se (see e.g. Eldridge and Palmer 2009; Witter et al. 2012) but rather on trying to answer the following research questions: What changes (in behaviours) are being initiated through the implementation of a PBF scheme and how does it resonate with stakeholders, the context and other programmes? What lessons can be drawn for research and policymaking? This is somewhat in line with the review of Miller and Babiarz (2013), but the latter focused on any kind of financial incentive in the health sector, while ours is restricted to PBF schemes. Moreover, our scope goes beyond the study of possible perverse effects and the 'what', 'how' and 'whom to reward'-questions. More specifically, we intend to open up the 'black box' of PBF and look at what research has learned us about the effects of PBF schemes on different aspects of health service delivery. For that purpose, we use a comprehensive analytical framework to guide the review which is built around the behaviour and interactions between the main actors in a PBF scheme (Renmans et al. 2016). It is informed by the assumptions and predictions of Principal–Agent (P–A) theory, enriched by those of behavioural economics. The P–A theory addresses the same problem as the one at the basis of PBF, namely bringing the interests of the agents, i.e. the health workers, in line with the interests of the principal, i.e. the Ministry of Health (MoH). It proposes to do this in a similar way by linking rewards to predefined targets, but also by trying to create an institutional environment conducive for reaching the principals' objectives. Given the similarities between P–A theory and PBF, the former provides a particularly useful structure to analyse the different steps in the elaboration of a PBF scheme, and highlights the explicit and implicit underlying assumptions and possible (negative) side-effects of such an intervention. Additionally, it also allows us to unveil the often overlooked aspects in research on PBF and to organize these aspects in an orderly manner. Behavioural economics enriches this with a more complex view on human behaviour.

In the following section, we first outline our search strategy and give a concise overview of the analytical framework. The findings of our review are presented and discussed in the next section, and the article concludes with selected recommendations for policymakers and researchers.

Methodology

An essential starting point for a review on such a heavily debated topic is a definition of the study subject. In line with Musgrove (2011), we distinguish between 'results-based financing' (RBF), which is 'any programme that rewards the delivery of one or more outputs or outcomes by one or more incentives, financial or otherwise, upon verification that the agreed-upon result has actually been delivered' (Musgrove 2011:3) and 'PBF', which is differentiated from the former on three accounts: 'incentives are directed only to providers, not to beneficiaries [patients]; awards are purely financial; and payment depends explicitly on the degree to which services are of approved quality' (Musgrove 2011:3). Moreover, following Fritsche et al. (2014), we restrict our definition to contracting-in approaches that involve existing health facilities (public or private) that operate under the aegis of the government, and we do rule out contracting-out projects that give contracts to (international) NGOs or private organizations to perform health services outside the hierarchical control of the public health sector.

The above constructed definition of PBF may be seen too narrow a definition of PBF; within the PBF 'community of practice'[1] it is today widely acknowledged that PBF entails a reform package that is broader than merely the introduction of a novel way of purchasing. It may also have the ambition to address aspects like the management autonomy of the health facility (e.g. to hire and fire staff or have free use of resources), a separation of functions (purchaser, provider, verification officers[2]), improved management and planning of health care by health facilities, participation of local stakeholders in the decision-making process at the health facility level and/or the evaluation of services, improved M&E systems in health care etc. Nonetheless, many studies still stick to a more narrow definition of PBF and limit the description of the scheme to the financial incentive arrangements. Within the framework of this literature review, we opt for a middle ground and consider PBF as an incentive scheme directed to health providers (facilities and/or health workers), but accompanied by a new level of autonomy of the health facility (e.g. to decide on the use of resources), increased monitoring and a separation of functions between the purchaser, provider and/or the newly created verification officer of health services (see Figure 1). As we will see, many PBF schemes are composed of additional measures, although they should be seen as part of that particular PBF scheme, they are in this case not essential in order to be referred to as a PBF.

Search strategy

We searched the online databases Wiley Online, PubMed and ScienceDirect using the search words and filters presented in Table 1.

1 The 'PBF community of practice' brings together hundreds of researchers (pro and con), practitioners and policy makers that work on PBF. (http://groups.google.com/group/performance-based-financing/)
2 A 'verification officer' is an organization (private non-profit, private-for-profit or even a government agency) contracted by the PBF scheme administration to check whether the services reported by the facility and its staff have indeed been delivered and whether the requested quality was achieved. This function has been specifically created in the frame of PBF policies. In our article, we use the term verification officer in order to make a clear conceptual distinction between, on the one hand, the people whose task it is to verify the reports of health facilities engaged in a PBF scheme, and, on the other, the much more conventional formative supervisor (generally a cadre of the district management team) whose task it is to support, train and coach health staff at the lower levels of the health system.

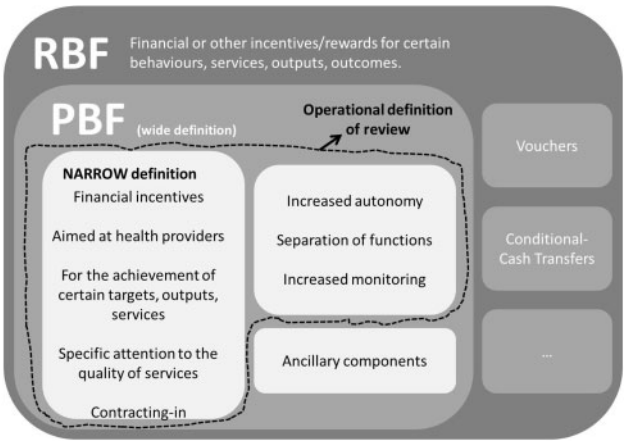

Figure 1. Used operational definition of PBF.

Our inclusion criteria were the following: original research article; PBF being the main subject and observations clearly related to PBF; PBF corresponding to the aforementioned definition; focus on low-or lower middle-income countries according to the classification used by the World Bank (World Bank 2015); and published in a peer-reviewed journal. Our exclusion criteria were the following: a review article; an editorial; an abstract meeting; the research population being unsalaried health workers; and the link between PBF and the observations is not clear. After an initial screening based on the title and abstract of the articles we scrutinized the remaining articles to check against our inclusion criteria.[3] The search was finalized in January 2016.

Analytical framework

Table 2 summarizes the eight dimensions, and their constitutive elements, of the analytical framework (Renmans et al. 2016). The first four are the actors in the centre of our framework: the MoH or the donors who are labelled 'the main principals' in P–A terminology (i); the health workers, who are called 'the agents' (ii); 'the verification officers' who verify the reports (see note 2) (iii); and the patients who are labelled 'the benefitting principals' (iv). Four other dimensions complement our framework: context and other stakeholders (v); the PBF contract (vi); the effects of the PBF scheme (vii); and the costs and benefits of the scheme (viii).[4]

The first dimension is constituted by the 'main principals' who delegate tasks to the agent and supply funds. They may be a MoH, an overarching medical board, a purchasing agency, a donor… Attention should be paid to the underlying objectives and values of the principals, and the more general political economy environment, as they may influence the set up and the success of the PBF scheme (Bertone and Meessen 2013). Understanding the political economy can also be useful for other countries that are planning, implementing or scaling-up a PBF scheme in their handling of the politics behind PBF (addressing questions such as which stakeholders need to be on board, what are the caveats at the political level etc.).

The health providers (health workers and facilities) or 'agents' constitute the second dimension of our framework. Their perception and acceptance of the PBF scheme is essential as it may reinforce certain forms of productive or detrimental behaviour such as 'gaming', i.e. actions that facilitate the attainment of the targets without contributing to a real or intended improvement in health outcomes (Baker 1992). Other well-known phenomena include 'cherry picking', i.e. choosing those patients who make it easier to reach the targets; blatantly manipulating information (Ireland et al. 2011); 'task trade-off', i.e. mainly focusing on actions that receive rewards to the detriment of other tasks and 'free riding', referring to health workers who do not contribute to the attainment of targets while still receiving the rewards (Laffont and Martimort 2002). Motivation obviously plays an essential role in such behaviour, hence the importance of investigating and distinguishing among different sources and kinds of motivation, from intrinsic and extrinsic to materialistic (financial) and non-materialistic motivation (social, moral, intrinsic) (see Paul and Robinson 2007). Related to this is the crowding out of intrinsic motivation by extrinsic motivators (Frey and Jegen 2001). Finally, as health workers' miscomprehension with respect to the PBF might also contribute to its failure (Ssengooba et al. 2012); it is also important to study the knowledge of the health workers regarding the PBF scheme.

Thirdly, the 'verification officer' has a decisive role. As he/she is in a P–A relationship with the funder, the same problems and opportunities as discussed in the previous paragraph have to be investigated. The independence of the verification officer is another issue to be scrutinized. There can be collusion between the verification officer and the health workers when they collectively manipulate their individual reports (Faure-Grimaud et al. 2003) or a conflict of interest might arise when the payment of the verification officer is correlated with the performance of the health facility she/he is verifying. The effect of this new actor/role on the existing relationship between the health workers and other more formative supervisors deserves attention as the verification role of this actor may create confusion concerning the role of the other supervisors. Finally, her/his capacity (knowledge, infrastructure, access) to correctly verify the records and power to act upon anomalies detected need to be studied.

Fourthly, the role of the patients ('benefitting principals') is essential. Elements that deserve to be studied more specifically include the degree to which the needs and wishes of the patients are taken into account by the health workers and the PBF scheme and the ways in which the patients participate in the setting up, implementation and evaluation of the PBF scheme.

Although often neglected, the context is an essential part of every (PBF) programme and the fifth dimension of our framework. The social, cultural, institutional (norms, laws, other policies, other sectors etc.), economic and epidemiologic context influences what is (and is not) acceptable, the interests of the principals and the agents, the degree of opportunism of the health workers, the perceptions of the PBF scheme, the preferred leadership styles etc. (Lubatkin et al. 2007; Wendt et al. 2009; Cuevas-Rodriguez et al. 2012). Moreover, other stakeholders (like health facilities, NGOs, religious organizations etc.) may also have a determining influence on the outcomes.

The sixth dimension is the contract that spells out the PBF scheme. We identify six important constitutive elements in such contracts: (i) the governance arrangements that structure the relationships between the different actors (supervisors, funders, patients, health workers) and shape the level of participation of the community and autonomy of the health facility. (ii) The matrix of indicators and quality measures consists of targets, objectives and indicators. Ideally, the targets and indicators should be 'SMART': Specific, Measurable, Attainable, Realistic and Time bound (Doran 1981). Moreover, they should be consistent with other objectives and targets, challenging enough, accepted by the health workers,

3 Specific justifications for not withholding certain articles can be obtained from the authors.

4 We would like to thank the reviewers for improving the aforementioned terminology.

Table 1. Used search words and filters

Database	Search words	Filters
PubMed	'Reimbursement, Incentive'[Mesh] OR 'PBF' OR 'performance-based incentives' OR 'P4P' OR 'payment for performance' OR 'RBF' OR 'Pay for performance'	**Publication dates:** 1 January 2000 to 31 December 2015
Wiley Online	• 'PBF' OR 'performance-based incentive' OR 'RBF' OR 'P4P' OR 'pay for performance' OR 'payment for performance'' • 'Incentive' AND 'health''	• In All Fields • Date range: 2000–2015 • Publication type: Journals • In Abstract (incentive); In All Fields (health) • Date range: 2000–2015 • Publication type: Journals
ScienceDirect	'Incentive' OR 'PBF' OR 'performance-based incentives' OR 'P4P' OR 'payment for performance' OR 'RBF' OR 'Pay for performance'	• In Abstract, Title or Keyword • **Date range:** 2000–2015 • **Publication type:** Journals: article or review article • **Sciences:** in 'arts and humanities', 'medicine and dentistry', 'nursing and health professions', 'psychology', 'social sciences'

Table 2. Dimensions and elements of our analytical framework

Dimension	Elements	Dimension	Elements
1. Main principals (donors, MoH etc.)	• Harmonization • Hidden objectives and values • Political economy	**6. Contract**	• Implementation • Governance arrangements (participation and interaction between agents, ownership etc.) • Matrix of indicators and quality measures (participation, measurability, attributability, seven requirements) • M&E arrangements (different options and effect on quality of services and Health Management Information System) • Financial incentive arrangements (appropriateness, height and timing) • Dispute settlement mechanism (new mechanism?, local context) • Ancillary components (effect on non-materialistic motivation)
2. Agents (health providers)	• Perception and acceptance of PBF and its components • Rent seeking behaviour (gaming, manipulation of info, cherry picking, task trade off, free-riding) • Motivation of health workers • Crowding out/in • Adverse selection • Knowledge of PBF		
3. Verification officers	• Scrutinize P–A relationship between verifier and funder • Collusion and conflict of interest • Capacity • Decision-making power		
4. Benefitting principals (patients)	• Needs and wishes (utility function) • 'Misdirected accountability' • (Ways of) participation	**7. Effects**	• Health Impact • Quality of care and services • Financial sustainability • Health workers motivation and self-esteem in the long run • Inequality • Existing institutions
5. Context and stakeholders	• Social and cultural context • Institutional context (norms, laws, other policies, other sectors etc.) • Economic context • Epidemiologic context • Other stakeholders (religious organizations, pharmaceutical companies etc.)	**8. Costs and benefits**	• Costs related to monitoring, premiums, transaction costs, other economic costs • Wider benefits (reforms, trust) • Comparison with other programs

oriented towards teamwork, easily attributed to the health workers' behaviour, observable and validated to measure what they are meant to measure (Stiglitz 1987; Liu and Mills 2007). (iii) The monitoring and evaluation (M&E) arrangements concern the way indicators are selected, the way the data are collected, monitored and reviewed, and the way the health workers are monitored by the verification officers. (iv) The financial incentive arrangements are related to the amount and timing of the incentives, as well as their fit within the local context. (v) A clear contract does not exclude disagreement, hence the need to scrutinize dispute settlement mechanisms. (vi) The last element of a PBF contract are its ancillary components e.g. improved formative supervision, improved management and planning of the health facility, increased accountability from facilities towards the government and possibly towards the wider population and community etc. Their effects on the performance and the extrinsic and intrinsic motivation of the health workers deserve closer scrutiny. The first five contractual elements can be considered essential components of any PBF scheme (see also earlier), while the sixth is equally important yet may differ from PBF scheme to PBF scheme. It is, however, important to compare the

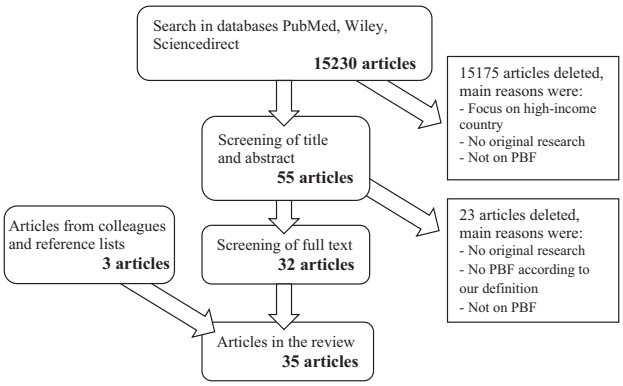

Figure 2. Flowchart of search results.

set-up on paper with the *de facto* implementation (Ssengooba et al. 2012).

The seventh dimension of our framework focuses on the positive and negative, expected and unexpected effects of a PBF scheme. These include effects at the level of health outcomes and the general health situation, the quantity and quality of health care services, geographical and interpersonal inequities—see the 'Inverse Care Law' (Hart 1971) —the Health Management Information System (HMIS), existing institutions such as norms and trust levels, and motivation and self-esteem of health workers and the financial sustainability of the intervention.

Finally, there is the need for a systematic analysis of the costs and benefits. The costs should include those related to monitoring, performance premiums, and transaction costs of the increased administrative burden (like the new role of verification officer) as well as other economic costs (Ireland et al. 2011). At the same time, the benefits may not be restricted to improved health outcomes, but may include, e.g. an increased feeling of appreciation or the facilitation of other necessary reforms. Moreover, PBF schemes should be compared with other programmes that envisage improving performance (e.g. non-monetary rewards, improved work conditions) (Mills 2014).

Results

The initial search of the databases generated 15 230 hits (see Figure 2) of which 55 articles remained for closer scrutiny after we screened the titles and abstracts and deleted duplicates. The main reasons for deletion were as follows: focus on high-and higher middle-income countries; reviews, editorials, or comments; or because they did not discuss PBF. After reading all articles, we concluded that 32 of them met our criteria, while the remaining studies were deleted because they were either not original research or because PBF was not in compliance with our definition. We subsequently added three articles found through the reference lists of articles already included or brought to our attention by colleagues. Table 3 gives an overview of the 35 articles that fulfilled our inclusion criteria and are used in the review.

In terms of geographic focus, there is a clear bias towards African countries, with a predominance of studies in Rwanda (14) and Burundi (7). We also observe that the subject is becoming increasingly popular; with the number of studies increasing from 13 articles until 2013, to 22 in 2014 and 2015.

One initial interesting observation is that PBF schemes differ on almost every single of the six 'contract' elements. This confirms the findings of Eldridge and Palmer (2009) who highlighted in their review that the 27 PBF schemes under study differed considerably with respect to the nature of the actors involved, the matrix of indicators and the incentive arrangements used. However, as evidenced in the remainder of this article, this diversity has not been reflected by research focusing on the specific impact of these different arrangements.

The discussion below is based upon an in-depth review of the 35 articles and structured according to the eight dimensions of our analytical framework. Some of the study limitations are highlighted at the end of the section.

Main principals

Notwithstanding the 2005 Paris Declaration on Aid Effectiveness[5], a lack of harmonization in terms of approaches and payment levels remains problematic (Fox et al. 2014; Paul et al. 2014). This has severe repercussions on the feasibility of national expansion of a PBF project and can contribute to a feeling of unfairness when different levels of incentives are used across different schemes in the same country (Paul et al. 2014).

Another important issue on the table since the Paris Declaration is ownership. Bertone and Meessen (2013) indicate that the 'underlying philosophy of a project [and in this case the donor] can influence the institutional rearrangement that its implementers are prepared to put in place' (p. 8). Similarly, Chimhutu et al. (2015) find that the government was not the driving force and even often ignored in the decisions on PBF in Tanzania. Kalk et al. (2010) and Paul et al. (2014) report complaints of health workers about indicators that mainly correspond to donor priorities and suffer from a lack of local embeddedness.

The political economy behind PBF only recently received the attention it deserves in the research literature. Van de Poel et al. (2015) clarify that the Cambodian government opted for a contracting-in approach (thus PBF) instead of a contracting-out approach, because the latter imposed a bigger burden on the budget and the MoH was unwilling to completely devolve health service delivery to international NGOs. In Tanzania, the pressure on the government to reach the Millennium Development Goals on child and maternal health and the eagerness of some donors (most importantly Norway) to implement a PBF scheme was more a matter of ideology and power battles between the donors than the result of a thorough study and consultation process (Chimhutu et al. 2015).

Agents

Studies reveal that the overall perception of health workers on PBF is mixed. On the one hand, it appears that many health workers have a positive stance towards PBF, mainly because of an increase in salary, but also because of the following: perceived positive effects on motivation, quality and volume of services; better access to information, more training and a feeling of recognition (Kalk et al. 2010; Bertone and Meessen 2013; Njoumemi and Fadimatou 2013; Chimhutu et al. 2014; Manongi et al. 2014; Paul et al. 2014). An increase in the number of nurses in Burundi and a decrease in turnover of health workers in Cambodia also point to a positive evaluation of PBF (Matsuoka et al. 2014; Falisse et al. 2015). On the other hand, health workers do have a number of criticisms on the way certain PBF schemes are implemented: dissatisfaction with the allocation method of allowances, the occurrence of nepotism, the

5 An agreement among donors and recipient countries to focus on five principles: ownership, alignment, harmonisation, managing for results and mutual accountability (see OECD 2005).

Table 3. Included articles

	Country	Article		Country	Article
1	*Benin*	(Paul et al. 2014)	19		(Kalk et al. 2010)
2	*Burundi*	(Falisse et al. 2012)	20		(Basinga et al. 2011)
3		(Bertone and Meessen 2013)	21		(Skiles et al. 2013)
4		(Peerenbom et al. 2014)	22		(Binagwaho et al. 2014)
5		(Bonfrer et al. 2014a)	23		(Zeng et al. 2014)
6		(Bonfrer et al. 2014b)	24		(de Walque et al. 2015)
7		(Falisse et al. 2015)[TQ3]	25		(Janssen et al. 2015)
8		(Rudasingwa et al. 2015)	26		(Lannes et al. 2015)
9	*Cambodia*	(Khim and Annear 2013)	27		(Lannes 2015)
10		(Matsuoka et al. 2014)	28		(Skiles et al. 2015)
11		(Van de Poel et al. 2015)	29	*Tanzania*	(Chimhutu et al. 2014)
12	*Cameroon*	(Njoumemi and Fadimatou 2013)	30		(Manongi et al. 2014)
13	*DR Congo*	(Soeters et al. 2011)	31		(Olafsdottir et al. 2014)
14		(Fox et al. 2014)	32		(Binyaruka et al. 2015)
15	*Rwanda*	(Meessen et al. 2006)	33		(Borghi et al. 2015)
16		(Soeters et al. 2006)	34		(Chimhutu et al. 2015)
17		(Meessen et al. 2007)	35	*Uganda*	(Ssengooba et al. 2012)
18		(Rusa et al. 2009)			

level of the payments, a lack of clarity and a feeling of unfairness (Khim and Annear 2013; Fox et al. 2014; Paul et al. 2014). There are also more general points of criticism expressed by health workers such as frustration about the conflict between targeted and untargeted tasks, a negative perception of PBF as a controlling mechanism, scepticism about the use of arbitrarily selected indicators, and the dependence upon donor funding (Kalk et al. 2010; Paul et al. 2014).

The acceptance of PBF is closely related to the motivation of the health workers. P–A theory, closely related to PBF, assumes that they act along the lines of the 'homo economicus' model, which posits that health workers would be mainly motivated by financial and self-regarded interests and act rationally to obtain them. However, several studies report expressions of intrinsic motivation by health workers (Kalk et al. 2010; Olafsdottir et al. 2014). It also seems that recognition for their work is essential to boost this intrinsic motivation (Kalk et al. 2010).

Although systematic research on the crowding out/in of intrinsic motivation within the framework of PBF is either ongoing or in the process of being published, some published studies do mention the issue. Bertone and Meessen (2013) found health workers in Bubanza (Burundi) claiming that clear tasks and objectives, an increase in resources and closer support 'generated pride and professionalism' (p. 7). However, in Ngozi, another Burundian province, they observe that the bonus was gradually perceived as a right and a fixed extra and that this may have led to less intrinsic motivation (Bertone and Meessen 2013).

It is clear that the jury is still out on 'the crowding motivation theory'; however, the mere observation that extrinsic motivation is an important facet of health workers' motivation warrants a closer look at 'rent seeking behaviour' or shirking. The first form is called gaming and has been observed in several PBF schemes. In Rwanda, the PBF programme aimed to reduce stock depletion. However, health workers refused to distribute the last boxes of medicine creating a *de facto* stock depletion yet still reaching the target (Kalk et al. 2010). In Tanzania, health facilities deployed harmful strategies to attract women to deliver at the facility by telling them that they would otherwise receive a fine or be denied vaccinations (Chimhutu et al. 2014).

Another side-effect is related to task trade-offs. Kalk et al. (2010) notice that practitioners became frustrated because of the limited time available, which made them choose between necessary activities in the intensive care unit and those needed for rewards (e.g. filling out forms). Binyaruka et al. (2015) even found a 'significant reduction' in non-targeted outpatient visits, possibly due to the data generation and verification activities (see also Janssen et al. 2015). Other studies found that the greatest increase of utilization was observed for the services that generated the highest incentive, which can also be an indication of the occurrence of task trade-off (Basinga et al. 2011; Chimhutu et al. 2014).

Cherry-picking is closely related to task trade-off but it concerns the choice 'between patients' instead of 'between tasks'; yet it receives much less attention in the literature. Lannes et al. (2015) found, using impact evaluation data, that 'easier to reach' patients (the less poor) were mainly focused on by the health workers in Rwanda, while Skiles et al. (2013) using less disaggregated data from the national Demographic and Health Survey did not find evidence that it would favour urban communities or wealthier quintiles.

A rather difficult issue to monitor is free-riding, which explains the lack of studies and observations on the issue. Khim and Annear (2013) report the importance of peer pressure to enhance performance and discourage free-riding.

The last form of 'rent-seeking behaviour' is the intended manipulation of reports. Kalk et al. (2010) report regular manipulations such as the arbitrary and retrospective filling of forms. In contrast, Khim and Annear (2013) state that the misreporting had decreased in Cambodia 'thanks to regular monitoring, random verification and the availability of web-based reporting' (p. 245).

Finally, it is essential that agents have a good understanding of the contract, a concern raised by health practitioners themselves (Paul et al. 2014). Ssengooba et al. (2012), Fox et al. (2014) and Janssen et al. (2015) found in their studies that lack of knowledge and understanding was one of the contributing factors of underperformance.

Verification officer

The study of Fox et al. (2014) underscores the importance of the introduction of a strong verification officer to detect fraud and rent seeking behaviour. First, according to Fox et al. (2014), a 'strong

verification officer' means an agency or organization, created for or assigned to the purpose of verifying the reported services, with enough authority to check the reports sent to them for deliberate errors and faults. This appeared not to be the case in a project in Cambodia where the director was concerned about over-reporting, yet was not able to verify this in the field (Matsuoka et al. 2014). Second, it also means that the verification officer has enough capabilities and knowledge to perform these checks. This seemed to be one of the problems in a PBF project in Uganda where the verification officers did not have enough clinical experience to make sense of the shorthand and recording practices in the primary registers (Ssengooba et al. 2012). Moreover, the workload for the verification officers became too heavy, which again decreased the reliability of the data obtained (Ssengooba et al. 2012). At the same time, a constructive relationship with the health workers is important as a lack of 'focus on the learning process' was seen as problematic in Rwanda (Janssen et al. 2015).

The verification officer's trustworthiness is another important aspect to investigate. Problems of collusion and conflicts of interest are observed in several studies (Falisse et al. 2012; Bertone and Meessen 2013; Chimhutu et al. 2014). A blatant form of conflict of interest and collusion was found in Ngozi, Burundi where the District Health Bureau was initially entrusted with this verification role while at the same time being evaluated itself on the basis of the performance of the health facilities. In the province of Bubanza, this was avoided by assigning this verification role to the purchasing agency (Bertone and Meessen 2013). However, Khim and Annear (2013) warn against this merger of roles (verification and purchasing) as it might lead to a conflict of interest to the disadvantage of the health workers. A less blatant form of conflict of interest is when there is discordance between the population and their representatives in the verifying Community-Based Organizations (CBOs)[6]—of higher socio-economic status - as found in Burundi by Falisse et al. (2012), which may be to the detriment of the poorest members of the community.

Finally, the relationship between the 'verification officer' and the funder of the programme can obviously be considered as another P–A relationship that may lead to side-effects similar to those discussed in the section on 'agents' in the analytical framework. In Burundi, Falisse et al. (2012) found that the verifying CBOs were mainly driven by financial incentives and that only one in four thought their work was important for the health system, signalling the absence of intrinsic motivation or a lack of focus on the common good. It was however not proven whether this was due to crowding out induced by the incentives.

Benefitting principals

There are several ways to include the patients' view in a PBF scheme. The Burundian project chose CBOs to monitor the performance of the health workers and collect the views of the patients via surveys. However, as has already been pointed out, these CBO's had an above average socio-economic status, making their representativeness questionable (Falisse et al. 2012). The real influence of the representatives is another important aspect. Again Falisse et al. (2012) found that in Burundi the information provided by the health committees was poorly used and medical staff had no obligation to take action after receiving patients' comments or recommendations.

Turning to the views of the benefitting principals (patients) on PBF, it appears that surprisingly little research has been done on this issue, despite it being central. Njoumemi and Fadimatou (2013) found that 60% of the population perceives PBF to be more effective than classic input payment, while only 30% thinks otherwise. This positive evaluation is confirmed by Lannes (2015) who found a positive influence on the satisfaction of patients with the quality of the services. However, Bonfrer et al. (2014a) find that the increase in quality was not acknowledged by the patients.

Context and other stakeholders

Taking the context into account before implementing a PBF scheme is essential, as the study of Olafsdottir et al. (2014) suggests. However, while several (case) studies briefly touch upon specific contextual dimensions, none of the research in our database provides an in-depth study of the contextual influence on the implementation and results of PBF schemes.

A first important issue is the institutional context in which a plethora of poorly coordinated, and often mutually conflicting incentive arrangements are functioning. These pre-existing financial incentives launched in the frame of *other* policies and programmes/projects, often addressing competing priorities (Ssengooba et al. 2012; Fox et al. 2014; Paul et al. 2014), can lead to unintended effects on the PBF scheme coming with its own distinct incentive structure, and vice-versa (typical of a situation of complexity). They not only affect the outcome of the PBF, but may also hamper its national scale-up. Conversely, policies breaking down demand-side barriers can improve the attainment of certain targets and the effectiveness of the PBF as proposed by *inter alia* Skiles et al. (2013), Matsuoka et al. (2014), Falisse et al. (2015) and Lannes et al. (2015) (e.g. through conditional cash transfers, health insurance or fee exemptions).

The very structure of the health sector is also important. Newly launched PBF schemes do not operate in a vacuum; they constitute an intervention in a health system characterized by complexity. Studies reveal that existing weaknesses in the health system, such as the poor quality of health care on offer, the suboptimal functioning of the routine Health Management and Information System, and problematic financial management, greatly influence the effectiveness of the PBF scheme (Meessen et al. 2006; Matsuoka et al. 2014; Chimhutu et al. 2014). However, not only weaknesses, but also specific arrangements of a health sector can interfere. For example, in Ngozi, Burundi, the facilities already received enough funds for training, drugs, infrastructure etc., from input-based payment, which reduced the incentive to achieve the bonuses which were partly intended for the purchase of the same items (Bertone and Meessen 2013). In the same project, Bertone and Meessen (2013) affirm that the alleged 'authoritarian role' (p. 853) of the Provincial Medical Director interfered with the co-management approach of the project and left no room for consensus building. In complex systems, power relations influence the implementation of new interventions, in this case PBF schemes.

The capacities of health facilities (human, financial, infrastructural resources) are critical in every programme and PBF is no exception. Meessen et al. (2007) and Olafsdottir et al. (2014) claim that a nationwide scale-up is probably not possible unless capacity is increased and Fox et al. (2014) state that the fragile context in the DR Congo presents 'considerable challenges in terms of appropriate design and implementation' (p. 2). Nevertheless, findings by Soeters

6 This is a 'public or private non-profit (including a church or religious entity) that is representative of a community or a significant segment of a community, and is engaged in meeting human, . . . or public safety community needs' (NNLM 2016).

et al. (2011) in DR Congo suggest that PBF is possible in a low-capacity context.

What seems to be particularly important is the 'right' type and amount (not too little or too much) of capacity. Overstaffing may lead to a dilution of the power of incentives (Fox et al. 2014), while sharing incentives with unqualified, temporary lay workers may indirectly contribute and help to sustain such practices (Matsuoka et al. 2014).

Although the focus of payment under a PBF scheme moves from inputs to outputs, appropriate and sufficient resources (drugs, materials) remain vital for good health outcomes (Khim and Annear 2013; Matsuoka et al. 2014; Olafsdottir et al. 2014). Skiles et al. (2013) explain why the Rwandan PBF was not 'an effective pro-poor strategy' (p. 830)[7] by referring to the lack of resources of health facilities in poorer communities and consequently the lower responsiveness to the needs of poor people and the inability of the Rwandan PBF set-up to respond to this.

This lack of capacity is also related to the general economic context. Uncertain and untimely payments due to insufficient funds can harm the credibility and impact of the PBF scheme (Ssengooba et al. 2012). However, a bad economic context at the level of the communities in which the health workers reside may have a positive effect on the acceptance of a PBF scheme as it is likely to increase the emphasis on extrinsic motivation and the desire to top up their salaries with incentives (Khim and Annear 2013).

Social and cultural attitudes are at least as important. Chimhutu et al. (2014) point to 'the type of fairness principle that prevails in a particular culture' (p. 9) and its influence on the acceptance of meritocratic payment schemes. The social context can also have a more direct influence on PBF and the attainment of targets when behaviours are not socially or culturally embedded (e.g. giving birth in health facilities) (Olafsdottir et al. 2014).

Lastly, other stakeholders than those explicitly stated in the PBF scheme are also part of the context. Other health facilities or organizations may enter in competition for badly needed competent health workers (Khim and Annear 2013), while higher bonuses in other facilities may lead to discontent (Paul et al. 2014). These other facilities are in turn influenced by PBF and have to be prepared to take up different tasks and increased pressure on the health system (e.g. increased referrals from PBF facilities) (Meessen et al. 2006) or, conversely, handle a decline of deliveries due to a shift of the deliveries toward PBF facilities (Van de Poel et al. 2015).

Contract

Like every policy, all constituting aspects of a PBF scheme need to be well elaborated in what the analytical framework refers to as the contract. The first important element is the institutional set-up or the governance arrangements. Research highlights that a difficult equilibrium must be found between strong and transparent structures to avoid fraud and corruption (Ssengooba et al. 2012; Khim and Annear 2013) and enough space for participation and autonomy (Soeters et al. 2006, 2011; Ssengooba et al. 2012; Paul et al. 2014; Van de Poel et al. 2015). Separating the different roles of fund holder, strategic purchaser and verification officer is said to promote neutrality and prevent any single officer from becoming too powerful (Soeters et al. 2006; Bertone and Meessen 2013; Khim and Annear 2013; Peerenboom et al. 2014).

A second element is the matrix of indicators and the quality measures. A participatory selection process is generally considered to design a set of indicators which are understood by the different stakeholders and/or adapted to the local context (Kalk et al. 2010). On the other hand, stringent and too optimistic timeframes may undermine such a process and lead to the selection of a rigid matrix poorly adapted to a changing context and thus jeopardizing the outcome (Ssengooba et al. 2012). A number of observations presented in the different studies may inform decisions regarding the matrix. First, indicators that elicit the largest behavioural change concern services 'over which the provider has greater control (e.g. prenatal care quality) and are less dependent on patients' health-seeking behaviour (e.g. timely prenatal care visits)' (Basinga et al. 2011:1425) (see also Skiles et al. 2013, 2015; Binyaruka et al. 2015). Second, a limited number of indicators and an assessment process that is not too complex ensures that verification is tangible and helps to ensure a clear link between the desired actions and incentives (Soeters et al. 2006; Khim and Annear 2013; Fox et al. 2014; Janssen et al. 2015). However, quality of care is a complex issue and is not easily captured by a limited set of indicators. In Rwanda, the overall quality score dropped to the baseline every time the quality indicators that composed it were modified (Janssen et al. 2015). This casts doubts on whether the indicators were actually indicating the underlying construct of quality. Striking a balance between comprehensiveness and feasibility of the matrix is, thus, challenging but needed. The absence of sufficient and complete data can also prove problematic and the identification of correct denominators may be especially difficult (Khim and Annear 2013). All this can lead to oversimplified indicators that ignore complexity, and only succeed in capturing the most visible and most easy to measure parts of performance. Finally, some PBF programmes may opt for targets instead of a 'fee for service'-logic, in such cases it is important to take into account that targets that are not challenging enough or not sufficiently discriminatory may reduce the incentive to perform as well as possible (Bertone and Meessen 2013; Peerenboom et al. 2014). However, finding the right target is not easy as setting different targets for responding to a difference at baseline may penalize already well performing facilities as they may find it harder to improve even more; while working with a fixed target for everyone, as in Benin, may lead to already higher performing facilities earning the most and increasing the existing inequality (Paul et al. 2014).

In Cambodia, the M&E arrangements helped to limit rent seeking behaviour and reduce absenteeism (Khim and Annear 2013). By definition, and in order to be effective, M&E needs to take place on a regular basis. Sound M&E systems thus entail a substantial increase in workload, funding, equipment, knowledge and human resources (Ssengooba et al. 2012; Khim and Annear 2013; Matsuoka et al. 2014; Janssen et al. 2015).

Once performance is assessed through the M&E system, rewards can be assigned. Health workers and researchers emphasize that rewards should take into account the extra costs and increased workload for health staff, but also the possible opportunity costs of ceased revenues due to possible new regulations as a consequence of the PBF programme (e.g. cessation of private practices) (Soeters et al. 2011; Khim and Annear 2013; Fox et al. 2014; Janssen et al. 2015). When the incentives do not fully cover these costs, PBF is likely to face opposition (Paul et al. 2014). As different groups of health workers (doctors, nurses etc.) bear different costs, they are motivated by different levels of incentives (Khim and Annear 2013). Interestingly, the need for a clear link between performance and incentives is disputed by Chimhutu et al. (2014) who claim that the perception of this link may suffice to lead to the desired behavioural changes. However, we can question the sustainability of such a practice. Whether

7 However, it was not 'pro-rich' either (Skiles et al. 2013).

performance payments should be directed to the individuals, the facilities or via the facility to the individuals remains unclear and needs further comparative research.

When disputes arise concerning the eligibility of a reward, a dispute settlement mechanism may be useful. However, the only study that explicitly refers to such a mechanism is the study of Soeters et al. (2006); they point out that in Rwanda a committee composed of all stakeholders is responsible for arbitration. However, as of yet, there is no detailed documentation on the precise functioning of such mechanisms, nor is there evidence regarding their impact.

The last ingredients of PBF schemes are the ancillary components. These are often undervalued in a description of what PBF is, yet they appear to be key when it comes to success (Khim and Annear 2013; Janssen et al. 2015). The most important agreement among health managers, health workers and researchers is that accompanying the financial incentives with qualitative feedback, training, coaching and formative supervision is essential in order to reach good results and motivate health workers (Rusa et al. 2009; Kalk et al. 2010; Soeters et al. 2011; Bertone and Meessen 2013; Matsuoka et al. 2014; Manongi et al. 2014; Paul et al. 2014; Janssen et al. 2015) The increased focus by the management boards of the facilities on planning and better management is indeed seen as enhancing the performance of health facilities (Soeters et al. 2011; Bertone and Meessen 2013). Better planning and more precise PBF targets contribute to more clarity on the responsibilities and tasks of the health workers, which in turn are perceived to generate a positive effect on their performance (Bertone and Meessen 2013; Khim and Annear 2013; Manongi et al. 2014). According to research in Benin, it was probably through these 'other elements of its package' (Paul et al. 2014: 212) that the PBF scheme motivated health workers.

Just as important as a well-designed scheme is its implementation. Keeping to the timing turns out to be essential. Delaying a PBF project in Uganda led to health workers forgetting the targets or the project as a whole (Ssengooba et al. 2012), while an untimely or incorrect payment of the incentives can lead to distrust and uncertainty which undermines the credibility of the project (Bertone and Meessen 2013; Chimhutu et al. 2014; Fox et al. 2014). Khim and Annear (2013) conclude from their study that clear rules and procedures for implementation are essential and help to overcome these problems. However, flawed implementation is sometimes also due to a lack of knowledge of the necessary human, financial and technical resources and of the possible responses of stakeholders and institutions (Ssengooba et al. 2012; Bertone and Meessen 2013).

Effects

The effects of PBF schemes on the targeted health indicators are mostly evaluated as mixed (e.g. Kalk et al. 2010; Binagwaho et al. 2014; Manongi et al. 2014; Zeng et al. 2014; Binyaruka et al. 2015; de Walque et al. 2015; Falisse et al. 2015). However, in Burundi, Rwanda and Tanzania, PBF led to more outreach activities and to new initiatives to increase the performance on the indicators (Bonfrer et al. 2014a; Chimhutu et al. 2014; Janssen et al. 2015). Paul et al. (2014) report an increase of professionalism and respect of national norms, a reduction of absenteeism and more cooperation. These improved outputs, however, are only intermediate results while the actual objective of PBF is to improve health outcomes. Kalk et al. (2010) report a positive impact of the health sector reform (of which PBF is one component) on infant and under-five mortality rates in Rwanda. The link between PBF and better health outcomes is, however, difficult to prove in such complex reforms. Van de Poel et al. (2015) did not find any effect on neonatal mortality in Cambodia despite the rise in institutional deliveries; nor did Skiles et al. (2015) on morbidity from diarrhoea, fever or symptoms of ARI in Rwanda. The shift of institutional deliveries from non-PBF facilities to PBF facilities in Cambodia and the persistence of demand-side barriers in Rwanda were depicted as the main reasons for this lack of effect (Skiles et al. 2015; Van de Poel et al. 2015). Other studies (Soeters et al. 2011; Bonfrer et al. 2014b; Rudasingwa et al. 2015) emphasize the importance of the quality of the increased output to reach better outcomes and show that PBF is able to improve this. However, the issue of quality is mostly studied in a narrow and superficial way, confining it to certain clinical subcomponents of care or even to the quality of the physical infrastructure. As already mentioned, in Rwanda, the indicators that composed the quality score were adapted a few times, after which the quality scores dropped to the baseline level before they started climbing again (Janssen et al. 2015), which may hint at the indicators not 'indicating' overall quality. In short, because researching and monitoring quality is difficult and at times even simply ignored (see Chimhutu et al. 2014), the effectiveness of PBF schemes as a quality improvement tool is unclear.

This also holds for the effect of PBF on equity in health services utilization for which evidence is either lacking or inconclusive. In Rwanda, Skiles et al. (2013, 2015) found that the PBF programme did not increase inequities; but it did not reduce them either. However, Binyaruka et al. (2015) found a 'potential pro-poor effect' in Tanzania on institutional deliveries. This contradicts the findings of Lannes et al. (2015) and Bonfrer et al. (2014b) who found a pro-rich effect. The reverse was found for vaccinations (Bonfrer et al. 2014b).

When scrutinizing the impact of PBF on health workers, it is important to distinguish between the short and the long term, as research suggests that the increased workload may eventually lead health workers to 'feel constantly tired' (Kalk et al. 2010). Other research indicates that health workers are starting 'to feel blasé about PBF' and taking the premiums for granted (Paul et al. 2014), hence decreasing their motivation. On the other hand, it has also been reported that PBF and the accompanied ancillary components increase the responsibility, the feeling of appreciation and the satisfaction of the health workers (Chimhutu et al. 2014; Manongi et al. 2014; Paul et al. 2014).

Another important issue is the impact of PBF on existing social institutions and norms. The levels of trust are especially fragile, and research shows that it is possible that a sense of unfairness and perception of nepotism or favouritism may take over in the minds of the health workers, depleting trust levels between the health workers and their superiors (Khim and Annear, 2013; Paul et al. 2014). Yet, these same studies and Kalk et al. (2010) also showcased that incentives at the level of service units may foster team spirit between the health workers. Although Paul et al. (2014) remark that the PBF scheme did not seem to improve collaboration and teamwork between the different levels of the health system in Benin. In Rwanda, however, Janssen et al. (2015) do report increased dynamism between actors at different levels. The interaction with the patients also seems to have improved (Kalk et al. 2010; Khim and Annear 2013; Paul et al. 2014) but Kalk et al. (2010) also hint at the fact that the relationship with the patient turned into a 'client' relationship; whether this is a positive outcome or not was questioned by some of the interviewees in the study.

The influence on existing institutional arrangements, such as decision-making rights, was exemplified by the study of Bertone

and Meessen (2013). Other research emphasizes the creation of new institutions that strengthen the management and planning capacity of the health facilities (Manongi et al. 2014; Peerenboom et al. 2014). Very few studies have until now focused on these institutional effects.

The HMIS can also benefit from PBF as it may improve the completeness of the data (Peerenboom et al. 2014) and increase its use for management purposes (Paul et al. 2014). However, it is not clear from these studies whether this concerns all the data or only those pertaining to the PBF scheme. This is conditional on close monitoring and verification of reports, since difficulties can occur due to manipulation of the data (Meessen et al. 2006).

Finally, the issue of (financial) sustainability of the schemes is questioned as the important and uncertain financial role played by donors impedes long-term budgeting (Peerenboom et al. 2014). Whereas the national scale-up in Burundi points to a certain long-term coherence, the budget for it remains largely aid-dependent and thus fragile (Falisse et al. 2015).

Costs and benefits

The only thorough and systematic cost-benefit analysis of PBF in a low-income country has been performed in Tanzania (Borghi et al. 2015). It found that half of the start-up economic costs were spent on training. Although the bulk of the recurrent economic costs (US$2.3 million for 13 months) were caused by the working hours used by the health workers to generate data (37%), the management of the project (28%) and the verification (13%); the pay outs only accounted for a small part (22%). The transaction costs (costs other than the pay outs) are remarkably higher than those observed in other studies (Meessen et al. 2006; Soeters et al. 2006, 2011) which never exceed the 50% as is the case in the Tanzanian study. This difference is related to the much broader approach to costs (including time spent by health workers and managers on PBF activities, the market value of all resources donated and used, the verification system etc.) adopted in Borghi et al. (2015). This administrative burden is also reported in Cambodia (Khim and Annear 2013) and Benin (Paul et al. 2014). These administrative costs can be tempered when large enough target populations are served (Soeters et al. 2006, 2011;) or the PBF is fully integrated in the health system instead of implemented through technical support (Borghi et al. 2015). All in all, surprisingly little is known about the transaction costs and thus the efficiency of PBF compared with other performance improvement programmes. The many negative (e.g. the possible crowding out effect) and positive (e.g. the improved data) externalities, are difficult to account for which makes it even more difficult to decide on PBF's efficiency.

Limitations of this review

Similar to other reviews, our findings depend on the quality of findings in the studies under review which were selected based on the inclusion criteria discussed in the methodology section. Given the fact that we base our work on studies from different countries and that the context has an important influence, the results cannot be easily extrapolated beyond the case study settings highlighted in Table 4. A similar remark can be made about the unit of analysis of the several studies in this review. Although we clearly defined PBF at the outset of the review[8], important differences remain between the PBF schemes under study: different institutional arrangements, levels of autonomy and/or more or different ancillary components. Thus the reason for some of the contradictory findings may be the difference in the unit of analysis and further research is needed on this issue. Finally, the occurrence of a publication bias towards more positive results is not unlikely as more critical studies may need more time to develop theoretically and methodologically.

8 This has led to the exclusion of some interesting studies (e.g. Witter et al. 2011) on initiatives which some experts (but not all) would classify as PBF

Discussion

In what follows we highlight the main lessons learned and recommendations for policymakers and researchers. Some may seem self-evident but the analysis of the articles in the 'Results' section nevertheless points to the relevance of reiterating them.

Politics matter

The first important lesson for policymakers (from donor and developing countries alike) is that politics and ideology matter. Discourse on development cooperation often tends to be technocratic, disregarding local political and democratic realities. However, there is no such thing as a neutral policy. PBF is no different and starts from some implicit assumptions about what is just and according to which philosophical vision(s) society should be structured (see Meessen 2013). This is even more relevant as Chimhutu et al. (2014) emphasize the difference in acceptance of PBF between egalitarian and more economically liberal societies. Clearly, values and ideas should remain in the centre of policy decisions in order to find the solutions that fit specific societal contexts.

Real inclusive democratic ownership is an important stepping stone to effective health policies. Moreover, under the assumption that PBF is a policy choice of the government and the local stakeholders, involving the different actors (health managers, health workers, patients etc.) can result in a more contextualized PBF arrangement which is more readily accepted and more faithfully implemented. Also, when governments are in charge, they can more easily harmonize different approaches, policies and incentives in order to limit unfair differences between facilities and ensure a coherent incentive strategy. By doing so, they facilitate possible national scale-up and avoid conflicting instructions and priority setting.

Finally, inclusive democratic ownership can help in keeping the focus of accountability towards the population instead of the donors. Another way to avert such 'misdirected or upward accountability' towards the donors is to incorporate representatives of the patients/population into the governance structure of the PBF scheme. Importantly, the population should feel adequately represented by their peers and the latter should have enough power to represent them (see McCoy et al. 2012).

Context matters

As for any other programme in the health sector the context is an important factor in the effectiveness and appropriateness of a PBF scheme. Not only other policies and social and cultural configurations, but also the available capacities, the quality and structure of the health system, the economic situation etc. have an important influence on a PBF scheme. Hence, good knowledge of the local context and of the various stakeholders' expectations is essential for a sound PBF design (see Ssengooba et al. 2012). Although policymakers are not always in the position to influence contextual elements, they should always take them into account when deciding on the PBF implementation (modalities). In order to guide such decisions within a constantly changing context, it may be advisable to use an action-research approach. Such a research set-up, where

researchers and implementers cooperate closely, requires M&E systems that are open to unplanned effects—be they positive or negative—and place equal emphasis on the E(valuation) as on the M(onitoring) dimensions. This is particularly relevant since it has been found that RBF mechanisms tend to favour monitoring to the detriment of evaluation (Liverani and Lundgren 2007). Sufficient knowledge, authority and capacity to perform the 'M' and the 'E' are thus essential.

Every component matters

The different case studies indicate that a PBF scheme is not a one-dimensional programme that can be copy-pasted into different settings. We distinguished six elements of a PBF contract (see also Renmans et al. 2016), which all comprise a number of choices to be made and questions to address: how much autonomy; how to monitor; who will monitor; how (much) to pay; which targets; how to handle disputes; how much training etc. Enough time should be invested in preparing a PBF scheme, taking into account lessons from other programmes/projects, yet at the same time being aware of the differences in context and objectives.

In preparing and studying a PBF scheme, the ancillary components should not be neglected. Indeed, an important finding of our review is that they can play an important role in motivating and improving the performance of the health workers. Paul et al. (2014) even claim that they may account for a very substantial part of the PBF success. These ancillary components should therefore not be at the margin of PBF research but rather be treated with as much attention as the incentives themselves.

A special case is the role of verification whereby policymakers should be wary of an approach which smacks too much of policing. The latter may conflict with and jeopardize the acceptability of formative supervision seen as an instrument for continuous training of health workers, certainly if both practices are conducted by the same person or the same body. Formative supervision is in essence a relational and qualitative approach, built on trust and mutual respect, and which entails dialogue, constructive feedback, problem-solving, training etc.

Clearly defining and delineating the different components of the PBF scheme and consequently monitoring and reviewing their effects consistently may help to keep the implementation of PBF schemes flexible. Some studies (Kalk et al. 2010; Chimhutu et al. 2014) found health workers engaged in practices that were detrimental for patients' health or that misguided the results; hence the need for adaptations to the scheme to counter these tendencies. With context being in a constant flux (e.g. changing health priorities), a PBF scheme should be sufficiently responsive to change and new knowledge.

Research matters

As follows from the abovementioned recommendations, research plays an important role in improving the implementation of PBF schemes. However, as already stated, research must not be a substitute for democracy. It is questionable whether everything that has proved its value should also be implemented. Nevertheless, research does have an essential role to play in informing policymakers. Despite the rise in studies on PBF some important research questions remain unanswered or even unexplored.

This is particularly valid as regards the impact of PBF on health outcomes. Although it is the ultimate objective of every health policy, robust data on this issue is lacking. Positive effects on the health outputs are being observed but the larger impact on health remains under scrutinized. Another important under researched issue is the influence of context on the effectiveness and acceptability of a PBF scheme; in particular the influence of other stakeholders, like other health facilities, pharmaceutical companies or international interest groups, is barely touched upon in the literature. Similarly, the focus of research is, understandably, mostly on health workers while the effect of PBF on the experiences of the patients is too often left out. The long-term effects on health workers' self-esteem, their intrinsic and extrinsic motivation, their capabilities, the trust between health workers and managers etc. also remain largely a blind spot. This lack of longitudinal research severely limits our knowledge of the durability of PBF results and of the more structural effects. Another major gap is the lack of insight into the interactions between the incentives, the verification process, ancillary components and the wider context: e.g. the influence of verification on the more formative supervision (see Bosch-Capblanch and Garner 2008) or the effect of the incentives on the implementation of other reforms (is the first a catalyst for the latter, see Meessen et al. 2011). Last but not least, the results reached by PBF schemes should be compared with other performance improving programmes on the basis of a thorough cost-benefit analysis, while keeping in mind that not everything is quantifiable and politics should remain central (Mills 2014). From this vantage point, the study of Borghi et al. (2015) is a particularly useful example on how to move forward.

If we wish to provide answers to this (non-exhaustive) list of research gaps, then research should focus on opening the black box of PBF. Unlocking and unpacking the programme theory of PBF and mapping out the pitfalls, side effects, opportunities and consequently the desirability of this financing mechanism should be a responsibility of PBF researchers. The use of theory-based evaluation (White 2009) can be illuminating and this review could be a starting point for the development of a PBF programme theory.

Construct validity matters

In order to perform sound research it is important to clearly identify the constructs under study. Many studies label projects wrongly as 'PBF' when, e.g. a contracting-out approach is being used (Zeng et al. 2013). This is a problem of construct validity, which in this case means that the 'failure to adequately explicate a construct may lead to incorrect inferences about the relationship between operation and construct' (Shadish et al. 2002:73). Moreover, in our study we highlighted that PBF schemes may dramatically differ on the six constitutive elements of the contract. Considering PBF as an undifferentiated theoretical concept is therefore fictional. As already highlighted in the limitations section, this makes it difficult to compare different PBF schemes from different contexts or even decide on whether a certain project is PBF or not. This reality deserves more attention within research. An important way forward therefore is to clearly discern the different elements of a PBF scheme, and unravel which elements initiate which mechanisms and how they interact. This should start at the very beginning of every research paper with a comprehensive description of the PBF scheme at hand; a practice that has not always been observed in the studies under review which, for example, made it sometimes difficult to decide whether or not a particular study satisfied the inclusion criteria. This also holds for the individual PBF components: for instance, a construct like 'community involvement' might veil substantial differences among different levels of participation (from filling out satisfaction surveys to being involved in decision making at the health facility) with potentially diverging implications in terms of motivation and performance. Again, theory-based evaluation might be a wholesome research approach to distinguish between the different elements and mechanisms.

Last but not least, a similar remark can be made with regard to the construct of performance. Here the paper of Musgrove (2011) is more elaborate as it points out that performance can refer to concepts as different as outputs, outcomes and impacts. Calling a PBF (un)successful without looking at the different aspects of performance is at the very least imprudent and we therefore advocate in favour of research that looks 'beyond' the narrow output indicators of the PBF scheme and that adopts a wider approach to performance, consistent with the need to clearly handle a systems perspective in the study of PBF and to fully acknowledge the complex character of health systems.

Notes

1. The 'PBF community of practice' brings together hundreds of researchers (pro and con), practitioners and policy makers that work on PBF. (http://groups.google.com/group/performance-based-financing/)
2. A 'verification officer' is an organization (private non-profit, private-for-profit or even a government agency) contracted by the PBF scheme administration to check whether the services reported by the facility and its staff have indeed been delivered and whether the requested quality was achieved. This function has been specifically created in the frame of PBF policies. In our article, we use the term verification officer in order to make a clear conceptual distinction between, on the one hand, the people whose task it is to verify the reports of health facilities engaged in a PBF scheme, and, on the other, the much more conventional formative supervisor (generally a cadre of the district management team) whose task it is to support, train and coach health staff at the lower levels of the health system.
3. Specific justifications for not withholding certain articles can be obtained from the authors.
4. We would like to thank the reviewers for improving the aforementioned terminology.
5. An agreement among donors and recipient countries to focus on five principles: ownership, alignment, harmonisation, managing for results and mutual accountability (see OECD 2005).
6. This is a 'public or private non-profit (including a church or religious entity) that is representative of a community or a significant segment of a community, and is engaged in meeting human, … or public safety community needs' (NNLM 2016).
7. However, it was not 'pro-rich' either (Skiles et al. 2013).
8. This has led to the exclusion of some interesting studies (e.g. Witter et al. 2011) on initiatives which some experts (but not all) would classify as PBF

Funding

This work was supported by a PhD grant of the Flemish Interuniversity Council (VLIR) (VLADOC2014PR008).

Conflict of interest statement. None declared.

References

Baker GP. 1992. Incentive contracts and performance measurement. *Journal of Political Economy* **100**: 598–614.

Basinga P, Gertler PJ, Binagwaho A *et al*. 2011. Effect on maternal and child health services in Rwanda of payment to primary health-care providers for performance: an impact evaluation. *The Lancet* **377**: 1421–8.

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

Binagwaho A, Condo J, Wagner C *et al*. 2014. Impact of implementing performance-based financing on childhood malnutrition in Rwanda. *BMC Public Health* **14**: 1132.

Binyaruka P, Patouillard E, Powell-Jackson T *et al*. 2015. Effect of paying for performance on utilisation, quality, and user costs of health services in tanzania: a controlled before and after study. *PLoS One* **10**: e0135013.

Bonfrer I, Soeters R, Van de Poel E *et al*. 2014a. Introduction of performance-based financing in burundi was associated with improvements in care and quality. *Health Affairs (Millwood)* **33**: 2179–87.

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

Borghi J, Little R, Binyaruka P, Patouillard E, Kuwawenaruwa A. 2015. In Tanzania, the many costs of pay-for-performance leave open to debate whether the strategy is cost-effective. *Health Affairs* **34**: 406–14.

Bosch-Capblanch X, Garner P. 2008. Primary health care supervision in developing countries. *Tropical Medicine and International Health* **13**: 369–83.

Chimhutu V, Lindkvist I, Lange S. 2014. When incentives work too well: locally implemented pay for performance (P4P) and adverse sanctions towards home birth in Tanzania - a qualitative study. *BMC Health Services Research* **14**: 1–12.

Chimhutu V, Tjomsland M, Songstad NG, Mrisho M, Moland KM. 2015. Introducing payment for performance in the health sector of Tanzania-the policy process. *Global Health* **11**: 1.

Cuevas-Rodríguez G, Gomez-Mejia LR, Wiseman RM. 2012. Has agency theory run its course? Making the theory more flexible to inform the management of reward systems. *Corporate Governance: An International Review* **20**: 526–46.

de Walque D, Gertler PJ, Bautista-Arredondo S *et al*. 2015. Using provider performance incentives to increase HIV testing and counseling services in Rwanda. *Journal of Health Economics* **40**: 1–9.

Doran GT. 1981. There'sa SMART way to write management's goals and objectives. *Management Review* **70**: 35–6.

Eldridge C, Palmer N. 2009. Performance-based payment: some reflections on the discourse, evidence and unanswered questions. *Health Policy and Planning* **24**: 160–6.

Falisse J-B, Meessen B, Ndayishimiye J, Bossuyt M. 2012. Community participation and voice mechanisms under performance-based financing schemes in Burundi. *Tropical Medicine and International Health* **17**: 674–82.

Falisse J-B, Ndayishimiye J, Kamenyero V, Bossuyt M. 2015. Performance-based financing in the context of selective free health-care: an evaluation of its effects on the use of primary health-care services in Burundi using routine data. *Health Policy and Planning* **30**: 1251–60.

Faure-Grimaud A, Laffont J-J, Martimort D. 2003. Collusion, delegation and supervision with soft information. *Review of Economic Studies* **70**: 253–79.

Fox S, Witter S, Wylde E, Mafuta E, Lievens T. 2014. Paying health workers for performance in a fragmented, fragile state: reflections from Katanga Province, Democratic Republic of Congo. *Health Policy and Planning* **29**: 96–105.

Frey BS, Jegen R. 2001. Motivation crowding theory. *Journal of Economic Surveys* **15**: 589–611.

Fritsche GB, Soeters R, Meessen B. 2014. *Performance-Based Financing Toolkit*. Washington, DC: The World Bank.

Ireland M, Paul E, Dujardin B. 2011. Can performance-based financing be used to reform health systems in developing countries? *Bulletin of the World Health Organization* **89**: 695–8.

Janssen W, Ngirabega Jde D, Matungwa M, Van Bastelaere S. 2015. Improving quality through performance-based financing in district hospitals in Rwanda between 2006 and 2010: a 5-year experience. *Tropical Doctor* **45**: 27–35.

Kalk A, Paul FA, Grabosch E. 2010. 'Paying for performance' in Rwanda: does it pay off? *Tropical Medicine and International Health* **15**: 182–90.

Khim K, Annear PL. 2013. Strengthening district health service management and delivery through internal contracting: lessons from pilot projects in Cambodia. *Social Science and Medicine* **96**: 241–9.

Laffont J-J, Martimort D. 2002. *The Theory of Incentives: The Principal-Agent Model*. Princeton, NJ: Princeton University Press.

Lannes L. 2015. Improving health worker performance: The patient-perspective from a PBF program in Rwanda. *Social Science and Medicine* **138**: 1–11.

Lannes L, Meessen B, Soucat A, Basinga P. 2015. Can performance-based financing help reaching the poor with maternal and child health services? The experience of rural Rwanda. *The International Journal of Health Planning and Management*. doi:10.1002/hpm.2297.

Liu X, Mills A. 2007. Motivation and performance -related pay. In: Preker A.S., Liu X., Velenyi E. V., Baris E. (eds.), *Public Ends, Private Means: Strategic Purchasing of Health Services*. Washington, DC: The World Bank, 237–58.

Liverani A, Lundgren HE. 2007. Evaluation systems in development aid agencies: an analysis of DAC Peer Reviews 1996—2004. *Evaluation* **13**: 241–56.

Lubatkin M, Lane PJ, Collin S, Very P. 2007. An embeddedness framing of governance and opportunism: towards a cross-nationally accommodating theory of agency. *Journal of Organizational Behavior* **28**: 43–58.

Manongi R, Mushi D, Kessy J, Salome S, Njau B. 2014. Does training on performance based financing make a difference in performance and quality of health care delivery? Health care provider's perspective in Rungwe Tanzania. *BMC Health Services Research* **14**: 154.

Matsuoka S, Obara H, Nagai M, Murakami H, Chan Lon R. 2014. Performance-based financing with GAVI health system strengthening funding in rural Cambodia: a brief assessment of the impact. *Health Policy and Planning* **29**: 456–65.

McCoy DC, Hall JA, Ridge M. 2012. A systematic review of the literature for evidence on health facility committees in low- and middle-income countries. *Health Policy and Planning* **27**: 449–66.

Meessen B. (2013). Financement basé sur la performance: structurons mieux le débat. http://www.santemondiale.org/ihpfr/?p=4465.

Meessen B, Kashala J-PI, Musango L. 2007. Output-based payment to boost staff productivity in public health centres: contracting in Kabutare district, Rwanda. *Bulletin of the World Health Organization* **85**: 108–15.

Meessen B, Musango L, Kashala JP, Lemlin J. 2006. Reviewing institutions of rural health centres: the Performance Initiative in Butare, Rwanda. *Tropical Medicine and International Health* **11**: 1303–17.

Meessen B, Soucat ALB, Sekabaraga C. 2011. Performance-based financing: just a donor fad or a catalyst towards comprehensive health-care reform? *Bulletin of the World Health Organization* **89**: 153–6.

Miller G, Babiarz KS. 2013. Pay-for-performance incentives in low- and middle-income country health programs. National Bureau of Economic Research Working Paper Series, **No. 18932**.

Mills A. 2014. Health Care Systems in Low- and Middle-Income Countries. *New England Journal of Medicine* **370**: 552–7.

Musgrove P. (2011). *Rewards for Good Performance or Results: A Short Glossary*. Washington, DC: The World Bank. http://www.rbfhealth.org/sites/rbf/files/documents/Rewards%20for%20Good%20Performance%20or%20Results%20-%20Short%20Glossary.pdf.

National Network of Libraries of Medicine (NNLM). (2016). Community Based Organization Defined. https://nnlm.gov/sea/funding/cbodef.html, accessed 11 February 2016.

Njoumemi Z, Fadimatou A. 2013. Financement basé sur la performance pour le suivi-évaluation du système de santé au Cameroun. *African Evaluation Journal* **1**: 11.

OECD. (2005). *The Paris Declaration on Aid Effectiveness*. Paris: Organisation for Economic Cooperation and Development.

Olafsdottir AE, Mayumana I, Mashasi I *et al*. 2014. Pay for performance: an analysis of the context of implementation in a pilot project in Tanzania. *BMC Health Services Research* **14**: 392.

Paul E, Robinson M. 2007. Performance Budgeting, Motivation and Incentives. In: Robinson M. (ed). *Performance Budgeting: Linking Funding and Results*. Basingstoke: Palgrave Macmillan, 330–75.

Paul E, Sossouhounto N, Eclou DS. 2014. Local stakeholders' perceptions about the introduction of performance-based financing in Benin: a case study in two health districts. *International Journal of Health Policy and Management* **3**: 207–14.

Peerenboom PB, Basenya O, Bossuyt M *et al*. 2014. La bonne gouvernance dans la réforme du financement du système de santé au Burundi. *Santé Publique* **26**: 229–40.

Renmans D, Paul E, Dujardin B. (2016). Analysing PBF through the lenses of the principal-agent theory. (Unpublished).

Rudasingwa M, Soeters R, Bossuyt M. 2015. The effect of performance-based financial incentives on improving health care provision in burundi: a controlled cohort study. *Global Journal of Health Science* **7**: 39854.

Rusa L, Ngirabega Jde D, Janssen W *et al*. 2009. Performance-based financing for better quality of services in Rwandan health centres: 3-year experience. *Tropical Medicine and International Health* **14**: 830–7.

Shadish WR, Cook T D, Campbell DT 2002. *Experimental and Quasi-experimental Designs for Generalized Causal Inference*. Boston, MA: Houghton Mifflin.

Skiles MP, Curtis SL, Basinga P, Angeles G. 2013. An equity analysis of performance-based financing in Rwanda: are services reaching the poorest women? *Health Policy and Planning* **28**: 825–37.

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

Soeters R, Habineza C, Peerenboom PB. 2006. Performance-based financing and changing the district health system: experience from Rwanda. *Bulletin of the World Health Organization* **84**: 884–9.

Soeters R, Peerenboom PB, Mushagalusa P, Kimanuka C. 2011. Performance-based financing experiment improved health care in the Democratic Republic of Congo. *Health Affairs* **30**: 1518–27.

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

Stiglitz JE. 1987. principal and agent (ii). In: Eatwell, J, Milgate, M, Newman P. (eds). *The New Palgrave: A Dictionary of Economics*. Basingstoke: Palgrave Macmillan.

Hart J. 1971. The inverse care law. *The Lancet* **297**: 405–12.

Van de Poel E, Flores G, Ir P, O'Donnell O. 2015. Impact of performance-based financing in a low-resource setting: a decade of experience in Cambodia. *Health Economics* doi:10.1002/hec.3219.

Wendt H, Euwema MC, van Emmerik IJH. 2009. Leadership and team cohesiveness across cultures. *Leadership Quarterly* **20**: 358–70.

White H. 2009. Theory-based impact evaluation: principles and practice. *Journal of Development Effectiveness* **1**: 271–84.

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* **2**: Cd007899.

Witter S, Zulfiqur T, Javeed S, Khan A, Bari A. 2011. Paying health workers for performance in Battagram district, Pakistan. *Human Resources for Health* **9**: 23.

World Bank. (2014). *HRITF Annual Report 2014: RBF a Smarter Approach to Delivering More and Better Reproductive, Maternal, Newborn, and Childhealth Services*. Washington DC: The World Bank. http://www.hritfreport.org/wp-content/uploads/2014/09/HRITF-2014-Annual-Report.pdf.

World Bank. (2015). Country and lending groups. http://data.worldbank.org/about/country-and-lending-groups, accessed 1 May 2015.

Zeng W, Cros M, Wright KD, Shepard DS. 2013. Impact of performance-based financing on primary health care services in Haiti. *Health Policy and Planning* **28**: 596–605.

Zeng W, Rwiyereka AK, Amico PR, Avila-Figueroa C, Shepard DS. 2014. Efficiency of HIV/AIDS health centers and effect of community-based health insurance and performance-based financing on HIV/AIDS service delivery in Rwanda. *American Journal of Tropical Medicine and Hygiene* **90**: 740–6.