

University of Groningen

## Do incentives matter when working for god? The impact of performance-based financing on faith-based healthcare in Uganda

Duchoslav, Jan; Cecchi, Francesco

*Published in:*  
World Development

*DOI:*  
[10.1016/j.worlddev.2018.09.011](https://doi.org/10.1016/j.worlddev.2018.09.011)

**IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.**

*Document Version*  
Publisher's PDF, also known as Version of record

*Publication date:*  
2019

[Link to publication in University of Groningen/UMCG research database](#)

*Citation for published version (APA):*

Duchoslav, J., & Cecchi, F. (2019). Do incentives matter when working for god? The impact of performance-based financing on faith-based healthcare in Uganda. *World Development*, 113, 309-319. <https://doi.org/10.1016/j.worlddev.2018.09.011>

### Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

### Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

*Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.*



# Do incentives matter when working for god? The impact of performance-based financing on faith-based healthcare in Uganda

Jan Duchoslav<sup>a,c,\*</sup>, Francesco Cecchi<sup>b,c,1</sup>

<sup>a</sup> Development Strategy and Governance Division, International Food Policy Research Institute, Malawi

<sup>b</sup> Faculty of Economics and Business, University of Groningen, Netherlands

<sup>c</sup> Development Economics Group, Wageningen University, Netherlands



## ARTICLE INFO

### Article history:

Accepted 5 September 2018

Available online 26 September 2018

### Keywords:

Performance-based financing  
Extrinsic incentives  
Efficiency of healthcare delivery  
Religious nonprofit organizations  
Uganda

## ABSTRACT

Can extrinsic incentives motivate faith-based healthcare providers? This paper challenges the finding that religious providers are intrinsically motivated to serve (poor) patients, and that extrinsic incentives may crowd-out such motivation. We use a unique panel of output and expenditure data from small faith-based nonprofit healthcare facilities in Uganda to estimate the effect of introducing performance-based financing. The output of the observed facilities is less than 50% of their potential. Performance-based financing increases output and efficiency robustly by at least 27%, with no apparent reduction in the perceived quality of services. Religious nonprofit healthcare providers may well be intrinsically motivated, but respond positively to extrinsic incentives. Whether working for god or not, incentives matter.

© 2018 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Public healthcare systems in many developing countries suffer from severe dysfunctionalities and endemic absenteeism: among public health workers in Uganda, unauthorized absence from duty may be as high as 50% (Björkman & Svensson, 2009). One way to improve the delivery of health services is to allow competition among different providers, regardless of their ownership status, guided by a principle of non-discrimination in the allocation of resources. This calls for a shift of responsibilities to the private and nonprofit sector. In fact, private healthcare represents a large share of health provision around the world, and this share is even greater among the poorest and most vulnerable. According to the World Health Organization Global Health Expenditure database, private healthcare providers account for 60% of health spending in low-income countries (Walton & Matthees, 2017). While for profit enterprises are growing rapidly, especially in urban areas, a large share of them is still faith-based: they are so called religious nonprofit organizations (RNPOs). In Uganda, 82% of all private nonprofit health facilities are coordinated by one of three faith-based organizations: the Uganda Protestant Medical Bureau (UPMB), the Uganda Catholic Medical Bureau (UCMB), and the Uganda Muslim Supreme Council (UMSC)—with a far greater share among

smaller dispensaries in rural areas (Reinikka & Svensson, 2010). Since 2000, the Ugandan government initiated a program in which every nonprofit primary health unit received an untied grant to help them offer their services. In a seminal paper, Reinikka and Svensson (2010) show that RNPOs responded to this unconditional surge in resources by increasing output. They interpret this to be consistent with the view that religious nonprofit providers are “working for God”, and thus intrinsically motivated and non-opportunistic.

Given growing budget pressures in many countries, and growing frustration with the lack of progress engendered by standard funding practices, a different approach to increasing healthcare output is recently becoming more popular: setting incentives that make the amount of funding a healthcare provider receives conditional on performance. There is increasing body of evidence that Performance-Based Financing (PBF), as this approach is often referred to, can increase both output and efficiency of healthcare in developing countries if the incentives are clear and well designed (Brenzel, 2009; Bhatnagar & George, 2016; Eldridge & Palmer, 2009; Hecht, Batson, & Brenzel, 2004; Honda, 2013; Novignon & Nonvignon, 2017). So far nonetheless, most rigorous studies on PBF have focused on public and private healthcare facilities, and did not investigate the heterogeneity of outcomes across sectors (Banerjee, Glennerster, & Duflo, 2008; Basinga et al., 2011; Bonfrer, Van de Poel, & Van Doorslaer, 2014; Bhatnagar & George, 2016; Meessen, Kashala, & Musango, 2007; Morgan, 2010; Soeters, Peerenboom, Mushagalusa, & Kimanuka,

\* Corresponding author at: IFPRI Malawi, P.O. Box 31666, Lilongwe 3, Malawi.

E-mail addresses: [j.duchoslav@cgiar.org](mailto:j.duchoslav@cgiar.org) (J. Duchoslav), [f.cecchi@rug.nl](mailto:f.cecchi@rug.nl) (F. Cecchi).

<sup>1</sup> Address: Nettlebosje 2, 9747 AE, Groningen, Netherlands.

2011; Sekabaraga, Diop, & Soucat, 2011).<sup>2</sup> Do intrinsically motivated RNPOs respond to extrinsic incentives in the same way as public sector? Or do extrinsic incentives erode the intrinsic motivation inherent to religious nonprofit healthcare outfits, potentially going as far as reducing their efficiency and quality of service?

We use a panel dataset from Uganda spanning a period of thirteen years and up to 246 small- to mid-sized health units belonging to the UCMB to estimate the effects of introducing PBF on healthcare output and – by extension – on the efficiency of their healthcare service delivery. We first analyze the data using data envelopment analysis (DEA) and stochastic frontier analysis (SFA)—standard approaches in healthcare production studies, both estimating the degree of inefficiency compared to some optimal benchmark frontier. Frontier efficiency measurements are common in studies focused on healthcare provision in developed countries, but – to the best of our knowledge – we are the first ones to apply them in a PBF evaluation in the context of a developing country. Next we estimate the parameters of a production function by means of a regression analysis and a more general parametric approach, using a dynamic version of the generalized method of moments (system-GMM). In this case, PBF can be seen as a new technology which shifts the whole production frontier. We find that the output of the observed facilities is less than 50% of their potential. Also, performance-based financing increases output and efficiency robustly by at least 27%. By conducting an independent client satisfaction survey we also show that this comes at no expense of the perceived quality of services provided. Jointly these results point towards a RNPO sector that is both responsive to extrinsic incentives and in dire need for increased efficiency. Whether working for God or not, incentives seem to matter—and can help deliver more healthcare services at a lower cost.

The remainder of the paper is structured as follows: in Section 2 we provide the background for the study, focusing on the latest literature on PBF and a description of the healthcare system in Uganda. Section 3 describes the data, while the methodological approach is outlined in Section 4. We present the main results on efficiency in Section 5 and additional results on perceived quality of service in Section 6. In Section 7, we discuss the results and conclude.

## 2. Background

### 2.1. PBF: one policy, many faces

The growing evidence that the health of its population is an important determinant of a country's economic growth (Bloom & Canning, 2000; Weil, 2007) has provided an additional argument – besides the ethical ones – for the need for functional and accessible healthcare provision. In contrast to the traditional perception that healthcare provision is essentially a function of structural inputs (including people, infrastructure, knowledge, drugs, material, equipment, and technology), the PBF paradigm focuses on the processes transforming these inputs into outputs (Eichler, 2006). Though usually thought of as complementary, the right processes can – to an extent – make up for the lack of inputs (Peabody, Tozija, Munoz, Nordyke, & Luck, 2004). By improving the transforming processes, more output can be produced using the same limited inputs.

In recent years, PBF has become one of the favorite ways to stimulate such improvements (Brenzel, 2009; Eldridge & Palmer, 2009; Hecht et al., 2004; Honda, 2013). However, while the

principal-agent problem has been successfully reduced by conditioning payment on performance in many other professional contexts (Miller, 2008; Zhao, 2005), it is rather difficult in processes with such multi-dimensional output as healthcare. While some level of agreement on best practices (increasingly grounded in economic theory and based on achieving specific, measurable, attainable, relevant, and time-bound – or SMART – indicators) has emerged over the past years (Fritsche, Soeters, & Meessen, 2014), the way in which various PBF components affect the multiple dimensions of healthcare delivery is still not fully understood (Renmans, Holvoet, Orach, & Criel, 2016; Renmans, Holvoet, Criel, & Meessen, 2017b).

Even if performance is understood in its most limited sense as output, thus excluding quality considerations, the many different types of output produced by a healthcare provider have to be taken into account when assessing its performance—either individually or according to some conversion logic. Expanding the notion of performance to include the quality and relevance of produced output, which are hard to quantify in a single metric, complicates the matter even further, and there is an ongoing debate on how best to measure these aspects of healthcare production. Despite these challenges, recent PBF schemes have now incorporated quality indicators in their design, typically based on checklists of observable structural and process measures (Josephson et al., 2017).<sup>3</sup> Such designs usually pay for output conditional on quality in a setting where the principal contracts the health facilities and the management of these providers then contracts the staff.

While this general framework is becoming commonplace, PBF programs operate in their specific settings, and many try to experiment with innovations, making each design unique. In our case, incentive payments are determined at the facility level, and make up only a fraction of total facility income with capped incremental bonuses. The allocation of the bonus payments is at the discretion of the in-charge of the facility, and typically redistributed to employees.<sup>4</sup>

Several studies document positive effects of PBF, at least on public healthcare delivery (Basinga et al., 2011; Bonfrer et al., 2014; Meessen, Musango, Kashala, & Lemlin, 2006; Meessen et al., 2007; Soeters et al., 2011; Sekabaraga et al., 2011). Others find no lasting effects (Banerjee et al., 2008; Morgan, 2010; Turcotte-Tremblay, Spagnolo, De Allegri, & Ridde, 2016); characteristically, when incentives do not trickle down to individuals in one way or another, or if other PBF design feasibility criteria are not met.

Existing literature, however, also identifies several potential pitfalls of PBF. Oxman and Fretheim (2008) warn against the danger of widening the already existing gap between poorly- and well-performing facilities, which may lead to an increasing gap in access to quality healthcare. Other concerns include the risk of increased gaming, i.e. systematic reporting bias (Kalk, Paul, & Grabosch, 2010; Kalk, 2011; Lu, 1999), target-led distortions resulting in the production of services with negative marginal value (Wynia, 2009), and cherry-picking of patients who are most suited to achieve targets (Ireland, Paul, & Dujardin, 2011).<sup>5</sup> Finally, direct

<sup>3</sup> While indices of structural and process measures are now probably the most common way in which PBF schemes promote quality of healthcare production, other approaches have also been employed. The PBF program described by Soeters et al. (2011) for example tries to ensure quality maintenance through comprehensive agreements with providers and regulators, and measure it through patient-perceived quality surveys that do not directly influence bonuses, and quality reviews done by peripheral health authorities at primary level or through peer group reviews at hospital level.

<sup>4</sup> This has since been standardized in most PBF designs by the indices management tool, which uses a group evaluation system rather than one where only the in-charge decides on the individual bonus payments of staff.

<sup>5</sup> These unintended consequences may be a sign that the PBF design offers too strong incentives. To help diminish this concern it is important to regularly review targets and incentives, as is the case in our setting.

<sup>2</sup> Even though some of these studies do include a share of faith-based nonprofit facilities, they do not investigate the differences that may arise from variations in intrinsic motivation.

financial incentives, as well as the bureaucratization of healthcare delivery, which is – to an extent – necessary to implement a PBF program, may end up crowding out intrinsic motivation (Frey & Jegen, 2001; Ireland et al., 2011; Kalk, 2011), inducing a decline in physician professionalism and morale (Wharam et al., 2009). The importance of intrinsic motivation in healthcare provision is clear (Brock, Lange, & Leonard, 2016; Leonard & Masatu, 2010; Reinikka & Svensson, 2010), and the theory behind extrinsic incentives crowding it out is well established (Deci & Ryan, 1985). However, its applicability to PBF is still subject to discussion (Lohmann, Houfort, & De Allegri, 2016), and empirical evidence of the effects of PBF on intrinsic motivation (which could be specific to each PBF scheme) remains inconclusive (Bertone, Lagarde, & Witter, 2016; Bhatnagar & George, 2016; Chimhutu, Lindkvist, & Lange, 2014; Lohmann et al., 2018; Shen et al., 2017).

Many of these pitfalls, and especially the last one, acquire particular salience when embedded in faith-based, mission-driven, nonprofit settings. If PBF crowds-out intrinsic motivation it may even have the opposite effect to the desired one.

## 2.2. Setting and intervention

The health sector in Uganda is characterized by a high degree of fragmentation with a mixture of public, private nonprofit and private for-profit healthcare providers (Björkman & Svensson, 2009), with government health spending representing only about a quarter of the total in most years since 2000 (World Health Organization, 2018). Although the Ugandan Ministry of Health takes non-governmental providers into account in its planning, providing partial funding to some of them, private health facilities and practices account for half of Uganda's reported healthcare output and operate independently of public ones (Government of Uganda – Ministry of Health, 2010). The policies governing healthcare services in the country are consequently as diverse as the service providers. In this complex situation, the Ugandan Ministry of Health piloted a large-scale PBF program for healthcare providers, which mostly turned out to be a failure due to a number of design problems (Morgan, 2010; Ssengooba, McPake, & Palmer, 2012). Other PBF programs have since been implemented in the country, including one which involves RNPO-run facilities (Renmans, Holvoet, & Criel, 2017a), but these have not yet – to the best of our knowledge – been comprehensively evaluated.

In this paper, we focus on one of the largest nonprofit private healthcare providers in the country – the Uganda Catholic Medical Bureau (UCMB). The UCMB runs an extensive country-wide network of hospitals and health centers accounting for over a third of private healthcare facilities in Uganda (Government of Uganda – Ministry of Health, 2010). The structure of the administration of the UCMB healthcare facilities mirrors that of the Catholic church itself: each of the 15 dioceses has its own health office, which is responsible for the operation of the health units within its territory. The central UCMB office coordinates the diocesan health offices, sets policy on the national level, monitors and evaluates the diocesan offices and the individual health units, and represents them nationally and internationally.

In 2008, the UCMB selected the diocesan health office in Jinja – one of the smaller dioceses – to pilot a PBF scheme in its six health centers to test the practical feasibility of this new approach before possibly extending the scheme to all its health units.<sup>6</sup> The Diocese

of Jinja was selected for its manageable size, for its long cooperation with Cordaid, an international NGO which helped set the scheme up, and for its historically good performance relative to other dioceses.

The scheme was introduced in the 2008/2009 fiscal year. It was supported by Cordaid with financial, material, and technical assistance,<sup>7</sup> but implemented by the diocesan health office in Jinja. For each health unit, the scheme set two long-term output targets based on its performance in the previous 3 years, in consultation with the unit's management. The targets are defined in terms of the Standard Unit of Output (SUO) – a weighted average of the most commonly performed procedures and services, taking into account their relative input requirements in terms of cost and time:

$$SUO = \text{outpatients} + 5 \times \text{inpatients} + 2 \times \text{deliveries} + 0.3 \times \text{ANC} + 0.2 \times \text{immunizations} \quad (1)$$

where *outpatients*, *inpatients*, *deliveries*, *ANC*, and *immunizations* are the numbers of outpatient visits, inpatient admissions, deliveries, ante-natal care visits (including family planning), and immunizations respectively.

Starting in fiscal year 2009/2010, each of the participating facilities received a financial bonus conditional on reaching the lower pre-specified target, and incrementally increasing up to its maximum if the upper target was reached.<sup>8,9</sup> The realized bonuses increased the income of participating facilities – about a third of which comes from user fees and the remainder from government grants – by about 5% on average. Although targets and bonuses were set at the unit level, the heads of the units typically used at least some of the extra income to top up the salaries of their employees proportionally to their hours worked, thus bringing individual incentives in line with those at the facility level.

## 3. Panel data

To gauge the effect of PBF on the efficiency of healthcare delivery, we use a range of output and input measurements collected by the UCMB from all its health units, which amount to a panel spanning up to 246 mid-sized health units over a period of thirteen years (fiscal year 2001/2002 – fiscal year 2013/2014). The last six fiscal years follow the introduction of PBF in the treated centers.

The data are compiled by individual health facilities, which report them to their superior diocesan health office. The diocesan health offices verify the data from the ledgers of randomly selected facilities on a monthly bases, and send it on to the UCMB, which uses the same random selection system to verify the data from the health offices. This data is collected for internal monitoring purposes and to guide decision making and as an advocacy device. UCMB engages in monthly data verification activities and omits any data that it has reason to consider unreliable.

Output is measured using the Standard Unit of Output (SUO, see Section 2.2 for details). Total *expenditures* are measured in millions of 2014 Ugandan Shillings (US\$). Panel descriptive statistics for

<sup>7</sup> Cordaid covered the financial costs of the scheme, offered technical support and training, and – where missing – provided IT equipment.

<sup>8</sup> Technically, output was only one of 5 aspects of performance incentivized by the PBF scheme. The others were equity (measured in user fees per SUO), efficiency (measured in cost per SUO), productivity (measured in SUO per number of qualified staff), and quality of service delivery (measured in terms of qualified staff per total staff, combined with results of drug prescription and patient satisfaction surveys). However, the targets set for these indicators were generally low enough to be met in nearly all cases. This leaves the output targets as the only ones practically affecting the size of the PBF bonuses paid out.

<sup>9</sup> This PBF design is rather uncommon in two ways: As recommended by Fritsche et al. (2014), PBF programs typically weigh outputs in monetary terms. More importantly, most do not involve output targets, but rather reward each unit of output – the notable exception being a PBF initiative in Malawi focused on maternal and newborn health (Lohmann et al., 2018).

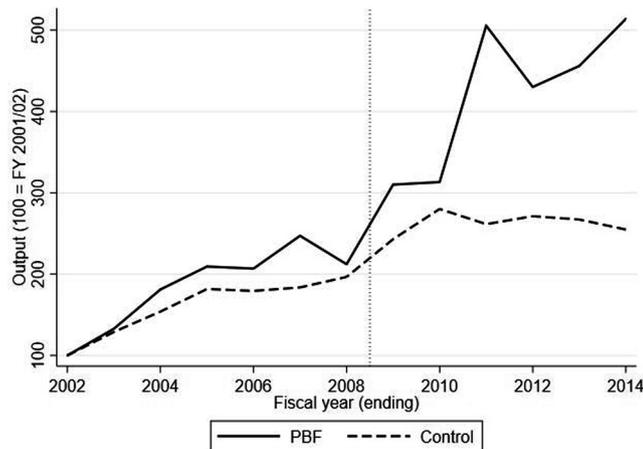
<sup>6</sup> One of the health centers dropped out of the scheme a few years after its introduction due to unspecified issues. Preferring to err on the cautious side, we treat it as under PBF for the remaining years as well. The PBF scheme has since been extended to government-run health centers in the same area to test its feasibility in a public setting. Since then, it has not – to the best of our knowledge – been expanded further.

**Table 1**  
Panel descriptive statistics.

Year	PBF in place	Observations <sup>a</sup>		SUO	Inputs		
		Control	PBF		expenditures	staff	beds
2001/02	No	75	4	10,471 (8753)	78.1 (95.3)	11.7 (7.4)	19.8 (16.7)
2002/03	No	58	2	10,892 (10,309)	79.9 (56.9)	12.0 (8.0)	22.6 (16.8)
2003/04	No	116	4	11,982 (8760)	86.7 (88.1)	11.7 (7.8)	19.4 (15.6)
2004/05	No	119	3	13,587 (9543)	93.9 (91.2)	12.1 (8.3)	20.7 (16.0)
2005/06	No	141	4	12,798 (9917)	86.4 (89.3)	10.8 (7.4)	19.8 (15.7)
2006/07	No	147	4	13,180 (10,012)	79.1 (75.6)	11.0 (7.7)	20.3 (15.7)
2007/08	No	166	4	12,956 (10,729)	72.3 (77.2)	10.4 (7.4)	19.5 (16.9)
2008/09	Yes	164	3	15,533 (12,370)	75.9 (79.9)	9.8 (6.6)	18.3 (15.2)
2009/10	Yes	180	4	18,361 (14,918)	97.7 (127.0)	10.7 (6.7)	21.3 (16.4)
2010/11	Yes	180	4	15,635 (12,963)	91.1 (105.2)	11.0 (7.6)	20.8 (16.0)
2011/12	Yes	169	4	14,660 (11,803)	91.3 (135.4)	11.2 (8.3)	22.0 (16.0)
2012/13	Yes	179	5	15,063 (11,799)	85.0 (111.6)	11.5 (6.9)	21.9 (16.1)
2013/14	Yes	159	5	15,042 (10,295)	92.4 (116.1)	12.1 (7.7)	22.3 (15.8)

Notes: Means (with standard deviations in parentheses) are reported for SUO, expenditures, staff, and beds.

<sup>a</sup> The panel has considerable gaps, leading to differences in how many of the six PBF facilities data is available for from year to year. This is also the case for control facilities. However, most of the unbalancedness of the panel stems from an expansion of the dataset through time, and to a much lesser extent through attrition and time-series gaps. Our analysis addresses this issue.



**Fig. 1.** Output trends.

SUO – as well as for the other two main factors of production, i.e. the total number of staff (*staff*) and capital proxied by bed capacity (*beds*) – are presented in Table 1. Fig. 1 shows the trends in output of PBF and control facilities.<sup>10</sup> It plots the average output for PBF and control facilities after standardizing each facility's output at 100 at the beginning of the observed period. The dotted vertical line marks the introduction of the PBF scheme.

#### 4. Empirical strategy

Most analyses of healthcare production typically involve data envelopment analysis (DEA) or stochastic frontier analysis (SFA)

<sup>10</sup> See Table 1 for the exact numbers of observations for both PBF and control facilities available for each year.

approaches (Hollingsworth, 2008). Each approach has its advantages as well as drawbacks. DEA does not require any distributional assumptions nor to impose a functional form on the production technology, but it is very sensitive to outliers and measurement error due to its non-stochastic nature. This is a problem especially in the context of developing countries where the reliability of data may be in question. SFA methods are considerably less sensitive to outliers and measurement error than DEA, but require imposing assumptions on the distribution of efficiency or the error term and about the form of the production function. In what follows, we use both approaches to model the production of healthcare as a function of three groups of inputs: expenditures, labor (number of staff) and capital (proxied by the number of beds). Complementarily, we present the results of a parametric analysis, starting from a pooled ordinary least squared (OLS) model, and gradually relaxing simplistic assumptions, correcting the initial estimate to include time and facility fixed effects, first-order autocorrelation, and ultimately modeling healthcare production as a dynamic process (Scott & Coote, 2010).

##### 4.1. Data envelopment analysis

The DEA approach, where efficiency scores are obtained non-parametrically through linear programming methods and regressed on facility characteristics, is popular because its non-stochastic first stage does not require one to make any assumptions about the functional form of production technology. Instead, the efficiency factor  $\theta$  is defined as the output of each facility – or decision-making unit (DMU) in the parlance of DEA – relative to the output of a virtual facility with the same levels of input, which in turn is a linear combination of the most efficient facilities in the dataset. As such,  $\theta$  is a facility's current output expressed as a fraction of its maximum potential output given the current levels of inputs and maximum efficiency. Formally, we solve the following linear program for each facility in each fiscal year:

$$\begin{aligned} & \min_{\theta, \lambda} \theta, \\ & \text{s.t. } -\frac{1}{\theta} \mathbf{SUO}_i + \mathbf{SUO}' \lambda \geq 0, \\ & \quad \mathbf{x}_i - \mathbf{X}' \lambda \geq 0, \\ & \quad \mathbf{N}' \lambda = 1, \\ & \quad \lambda \geq 0 \end{aligned} \tag{2}$$

where  $0 \leq \theta \leq 1$ ,  $SUO_i$  is the output of the  $i$ -th facility,  $\mathbf{x}_i$  is a vector of its inputs (*expenditures*, *beds*, *staff*),  $\mathbf{SUO}$  is a vector of outputs of all the  $N$  facilities in a given year,  $\lambda$  is an  $N \times 1$  vector of constant weights, and  $\mathbf{N}$  is a  $N \times 1$  vector of ones.

In the second stage, we estimate the effects of the introduction of PBF on  $\theta$ , using the double-bootstrapped maximum likelihood truncated regression procedure proposed by Simar and Wilson (2007) to correct for biases arising from within-facility correlation of efficiency:

$$\theta_{it} = \alpha + \beta PBF_{it} + \delta_i + \tau_t + \varepsilon_{it} \tag{3}$$

where  $PBF_{it}$  equals 1 if the  $i$ -th facility had a PBF program in place in fiscal year  $t$  and 0 otherwise,  $\delta_i$  are spatial fixed effects on the diocese level, and  $\tau_t$  are fiscal year fixed effects. We use code adapted from Wolszczak-Derlacz and Parteka (2011) to estimate the second stage.

#### 4.2. Stochastic frontier analysis

The deterministic nature of DEA gives the first stage of this approach the advantage of not necessitating any functional or distributional assumptions, but it also renders it extremely sensitive to outliers. The issue is further aggravated in developing country contexts like ours, where data collection methods are often suboptimal, making the data exceptionally noisy. SFA – the most common alternative to DEA – is considerably less sensitive to outliers and noise in the data, but does require an explicitly defined production function. To allow for sufficient functional flexibility, we use a translog production function with expenditures, number of staff and capital (proxied by bed capacity) as factors of production:

$$\ln SUO_{it} = \alpha + \gamma' \mathbf{factors}_{it} + (v_{it} - u_{it}) \tag{4}$$

where  $\mathbf{factors}_{it}$  is a vector representing a translog production function with *expenditures* <sub>$it$</sub>  (total expenditures in millions of US\$), *staff* <sub>$it$</sub>  (total number of staff), *beds* <sub>$it$</sub>  (bed capacity proxying for capital) as production factors,  $u_{it}$  is measure of inefficiency (a Euclidean distance from the estimated production frontier) with a truncated normal distribution, and  $v_{it}$  a normally distributed stochastic error term. Following Battese and Coelli (1995), we use a maximum likelihood random effects model to estimate the impact of PBF on  $u_{it}$ :

$$u_{it} = \zeta + \beta PBF_{it} + \delta_i + \tau_t + w_{it} \tag{5}$$

where  $w_{it}$  is a normally distributed stochastic error term and the rest of the notation is the same as above, so that:

$$\ln SUO_{it} = \alpha + \gamma' \mathbf{factors}_{it} + (v_{it} - (\zeta + \beta PBF_{it} + \delta_i + \tau_t + w_{it})) \tag{6}$$

#### 4.3. Panel regression analysis and dynamic system GMM

To ease potential concerns about the assumptions and drawbacks of both DEA and SFA, we also present the results of a parametric analysis, regressing output on PBF and the equivalent of the same translog production function as in (4). We then correct a bias due to the presence of facility-level fixed effects by estimating a fixed-effects (FE) model, and we include fiscal year fixed effects to produce the following fully specified model:

$$\ln SUO_{it} = \alpha + \beta PBF_{it} + \gamma' \mathbf{factors}_{it} + \eta_i + \tau_t + \varepsilon_{it} \tag{7}$$

where  $PBF_{it}$  is a dummy equal to one if PBF is in place in the given facility and fiscal year.  $\mathbf{factors}_{it}$  is a vector representing a translog production function with *expenditures* <sub>$it$</sub>  (total expenditures in millions of US\$), *staff* <sub>$it$</sub>  (total number of staff), *beds* <sub>$it$</sub>  (bed capacity as a proxy for capital) as production factors.  $\eta_i$  are facility-level fixed effects,  $\tau_t$  are fiscal year fixed effects, and  $\varepsilon_{it}$  is a stochastic error term. Although we technically estimate the effect of PBF on output rather than on allocative efficiency, it has no consequence for the magnitude and statistical significance of the estimated effect  $\beta$  since all factors of production enter on the right-hand side of the regression equation. Substituting allocative efficiency for output on the left-hand side would only affect the values of  $\gamma$ , leaving  $\beta$  unaffected.

It can be expected that performance in one period is heavily influenced by performance in previous time periods (Scott & Coote, 2010). We test for serial autocorrelation following Wooldridge (2002), and verify the assumption of stationarity using a series of augmented Dickey-Fuller (ADF) tests such as the Fisher-type test suggested by Choi (2001) and that of Im, Pesaran, and Shin (2003). We then reestimate the model using a three-step feasible generalized least squares estimator (FGLS) to correct for autocorrelation.

Following Roodman (2009), our final estimator is a panel-robust two-step Blundel-Bond system GMM with forward orthogonal-deviations, both first differences and levels of the independent variables as standard instruments, and a GMM-style instrument that collapses all available lags of the lagged dependent variable for each time period into one moment. This allows for correcting both autocorrelation and imbalances in the panel. Dynamic panel data analysis helps reduce – if not resolve – key econometric problems often arising from empirical studies that use conventional cross-sectional or time-series datasets. The large number of data points increases the efficiency of econometric estimates and, by utilizing information on both the inter-temporal dynamics and the individuality of the entities being investigated, it better controls for the effects of missing or unobserved variables (Hsiao, 2003). Also, by following facilities over a 13-year time span, we can construct a proper recursive structure to study the before-after effect, addressing concerns over the short-lived nature of PBF-induced increases in performance (Maynard, 2012; Oxman & Fretheim, 2008).

### 5. Results

We estimate the effect of PBF on productivity using the traditional two-stage DEA and SFA approaches. The linear program in the first DEA stage produces largely varied efficiency scores  $\theta$  across facilities and years with mean value  $\mu_\theta = 0.439$  and standard deviation  $\sigma_\theta = 0.240$ . In words, the health facilities have on average been producing only 44% of the output that they could potentially produce, given their inputs and the efficiency of the best-performing units (similarly estimated efficiency scores of healthcare facilities in developed countries, as reported by Hollingsworth (2008) and O'Neill, Rauner, Heidenberger, & Kraus (2008), typically fall between 80% and 90%). In the second double-bootstrapped maximum-likelihood stage, we estimate that the introduction of PBF increases efficiency  $\theta$  by 20.1% points (i.e. by 45.8%) – or almost one standard deviation (Table 2, column 1).

Though estimated simultaneously, the SFA approach can effectively be thought of as a two-stage process which first estimates a production frontier corrected for inefficiency  $u$ , and then regresses  $u$  on PBF. We estimate that the introduction of PBF reduces the initial  $u$  of 0.817 by 0.476 (Table 2, column 2). Since the frontier is modeled as a translog function, these estimates require a simple

**Table 2**  
PBF effects on output and efficiency – DEA and SFA

Dep. var.	(1)	(2)	
	DEA	SFA	
	2nd stage	1st stage	2nd stage
	$\theta$	$\ln SUO$	$u$
PBF	0.201*** (0.038)		-0.476** (0.241)
Translog inputs	Yes	Yes	Yes
Time FE	Yes	No	Yes
Diocese FE	Yes	No	Yes
Mean dep. var.	0.439	9.321	0.817
Std. dev. dep. var.	(0.240)	(0.699)	(0.228)
Obs.	1903	1903	
Wald $\chi^2$ (df)	1045.71 (31)	2877.92 (9)	

Notes: (Cluster) robust (bootstrapped) standard errors in parentheses. \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

arithmetic manipulation to be directly comparable to our linearly estimated DEA results: The inefficiency  $u$  of 0.817 is equivalent to an efficiency  $\theta$  of 44.2%, and the PBF-induced reduction in  $u$  by 0.476 is equivalent to an increase in  $\theta$  of 26.9% points (i.e. 60.9%).

**Result 1.** *The efficiency of healthcare facilities in our sample is on average 44%, or about half that of health facilities in developed countries.*

The SFA procedure is more robust to noise in the data than DEA, but both techniques are likely to bias the estimated gains in efficiency (46–61%) upwards due to their static nature. We thus continue investigating the effect of PBF parametrically. We start by presenting a naïve pooled OLS model (Table 3, column 1). A Wald test confirms the presence of facility-level fixed effects ( $F(215, 1677) = 5.07$  is significant at the 1% level), and a cluster-consistent Hausman-type test following Arellano (1993) reveals that a random effects model would be inconsistent ( $\chi^2(10) = 28.251$  is significant at the 1% level). We additionally control for a possible time trend by including fiscal year fixed effects (Table 3, column 2). A Wald test confirms the joint significance of fiscal year fixed effects ( $F(12, 215) = 12.65$  is significant at the 1% level). Following Wooldridge (2002), we reject the null hypothesis of no first order autocorrelation in the panel ( $F(1, 192) = 61.532$  is significant at the 1% level). We verify the assumption of stationarity through a series of ADF tests, all of which reject the null hypothesis of non-stationarity at the 1% significance level. To correct for first-order autocorrelation, we therefore estimate the model using an FGLS estimator (Table 3, column 3).

These parametric models produce large, significant effects of PBF on output which, apart from the naïve OLS result in column 1, are very much in line with the estimates of our previous DEA and SFA (49–51%). However, so far our estimates are static, and do not account for the dynamic nature of the data—output last year is likely to be a fundamental determinant of output in the current year. Our final and favored estimator is therefore a system GMM (Table 3, column 4). The magnitude of the estimated effect of the introduction of PBF on healthcare output decreases drastically to 27%. Its statistical significance also decreases in the process, but always remains below the 5% level. This is consistent with Scott and Coote (2010), who assert that healthcare provision is an inherently dynamic system, with current output and efficiency predicting future ones. Importantly, the increase is systematic and similar across individual indicators. Columns 1 to 3 of Table 4 show that output in terms of outpatients, inpatients, and immunizations

increases significantly in a similar fashion to the overall effect on the SUO. Column 4 reports the results for deliveries. Also in this case the coefficient is very consistent with the other estimations. However, as can be seen from the table, this estimation fails the Arellano-Bond test for AR(1) in first differences, as well as the Hansen test for instrument exogeneity. This implies that the moment conditions used in column 4 are not valid, and that the results should therefore be interpreted with care.<sup>11</sup> Taken together, the findings of Table 4 indicate that the expansion of total output did not come as a trade-off between different outputs, but rather as a shift in the total amount of medical services performed.<sup>12</sup>

**Result 2.** *The introduction of PBF robustly increases output through improved efficiency of healthcare provision by at least 27%, or over a third of a standard deviation.*

## 6. Perceived quality

A 27% increase in output is far from modest, especially in light of the fact that it is due to a financial incentive scheme worth only about 5% the total income of the participating facilities, and that we control for the increase in budget. Therefore, it is quite plausible that such a sudden rise in output could come at the expense of quality. For instance, to maximize revenue, health staff could perform the services in a hasty manner. Such response to the incentives would suggest that there is an opportunity cost in terms of quality of service. If the staff are aware of their implicit decision, this could indicate an erosion of professionalism and intrinsic motivation. To see whether this might be at stake, we conducted a two-wave patient satisfaction survey in the PBF facilities and in a sub-sample of those without PBF.

While patient satisfaction is subjective, several studies endorse its validity as an instrument for measuring quality of healthcare (Andaleeb, 2001; Davies & Ware, 1988; Johansson, Oléni, & Fridlund, 2002) and it has recently been placed at the core of policy recommendations regarding PBF in the United States (Ryan, 2009; Wharam et al., 2009). Leonard (2008) shows that satisfaction is jointly produced with quality during the course of a consultation and that patients respond to increased quality by being more likely to be satisfied. Moreover, patient satisfaction reflects both process quality and clinical quality (Marley, Collier, & Meyer Goldstein, 2004), making it a good measure of the overall quality of healthcare delivery. This said, increased efforts by health facilities to improve the experience of patients and persuade them to come back (an intermediary variable in the PBF theory of change) may weaken the correlation between patient satisfaction, perceived quality and actual quality of service in our setting. In other words this analysis should be seen as an effort to assuage the concern that patient care was neglected in favor of output growth, rather than as evidence of changes in intrinsic motivation.

We use the output and expenditure data along with several other characteristics to match the six PBF facilities with six similar ones receiving fixed funding, identified among other facilities in linguistically and culturally affine areas. The matches are based on a propensity score calculated from a set of indicators not constituting a building block of the SUO (the facilities' income and expenditures, catchment population, number of staff, bed capacity, number of carried out diagnostic procedures and minor surgical

<sup>11</sup> An alternative dynamic estimate that allows for first-order moving-average MA (1) errors solves this problem and reveals quantitatively similar results.

<sup>12</sup> We do not report the effect of PBF on antenatal care visits as our model presents strong evidence against the null hypothesis that the overidentifying restrictions are valid (although the Sargan test has a tendency to underreject). Rejecting this null hypothesis requires to reconsider the model or instruments, and we could not find suitable substitutes.

**Table 3**  
PBF effects on output and efficiency–panel regressions.

	(1) OLS	(2) FE	(3) FGLS	(4) sysGMM
Dep. var.	ln <i>SUO</i>	ln <i>SUO</i>	ln <i>SUO</i>	ln <i>SUO</i>
PBF	0.734*** (0.129)	0.492*** (0.131)	0.509*** (0.095)	0.266** (0.130)
ln <i>SUO</i> <sub>it-1</sub>				0.377*** (0.053)
Translog inputs	Yes	Yes	Yes	Yes
Facility FE	No	Yes	Yes	No
Diocese FE	No	No	No	Yes
Time FE	no	Yes	Yes	Yes
SE clustered by facility	Yes	Yes	No	No
Mean dep. var.	9.321	9.321	9.321	9.372
Std. dev. dep. var.	(0.699)	(0.699)	(0.699)	(0.684)
Obs.	1903	1903	1896	1572
R <sup>2</sup>	0.594	0.384		
Wald $\chi^2(230)$			11063***	
Instruments				52
AR(2) z				0.76
Sargan $\chi^2(8)$				7.30
Hansen $\chi^2(8)$				7.64
Difference in Hansen $\chi^2(1)$				1.66

Notes: Robust SE in parentheses. \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . Inputs: *expenditures, staff, capacity*. To avoid proliferation of instruments in the system GMM model, we take the spatial fixed effects one level up within their nested structure, using diocese-instead of facility-level fixed effects.

**Table 4**  
PBF effects on individual output indicators.

	(1) sysGMM	(2) sysGMM	(3) sysGMM	(4) sysGMM
Dep. var.	ln <i>outpatients</i>	ln <i>inpatients</i>	ln <i>immunizations</i>	ln <i>deliveries</i>
PBF	0.213* (0.113)	0.245 (0.197)	0.320*** (0.120)	0.285*** (0.086)
Dep. var. <sub>it-1</sub>	0.438*** (0.054)	0.475*** (0.074)	0.531*** (0.060)	0.746*** (0.082)
Dep. var. <sub>it-2</sub>		0.103** (0.043)		0.033 (0.046)
Dep. var. <sub>it-3</sub>				0.124*** (0.034)
Translog inputs	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes
Diocese FE	Yes	Yes	Yes	Yes
Mean dep. var.	8.705	6.452	8.333	4.547
Std. dev. dep. var.	(0.689)	(1.122)	(0.869)	(1.148)
Obs.	1572	1380	1572	1303
Instruments	52	51	52	50
AR(2) z	0.36	0.19	-0.22	1.69*
Sargan $\chi^2(df)$	5.15 (8)	4.02 (6)	11.43 (8)	5.00 (4)
Hansen $\chi^2(df)$	4.01 (8)	3.55 (6)	9.55 (8)	5.63 (4)
Difference in Hansen $\chi^2(1)$	0.35	0.02	0.06	2.73*

Notes: SE in parentheses. \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . Inputs: *expenditures, staff, capacity*. An additional lag of the dependent variable is needed in model (3) to satisfy identification requirements. Similarly, two additional lags of the dependent variable are needed in model (4).

operations, average length of stay of admitted patients, number of fatalities, and the availability of mental counseling), as measured in the last year prior to the introduction of the scheme. By matching on values collected before the implementation of PBF, we ensure that the propensity scores are not influenced by any potential confounding effects of the intervention.

To gauge the perceived quality of the services in these facilities, we adapted the Hospital Consumer Assessment of Healthcare Providers and Systems (HCAHPS) survey (Centers for Medicare & Medicaid Services, 2011) to the Ugandan environment and administered it in two waves in 2012 and in 2014.<sup>13</sup> To address concerns

that the perceived quality might be highly dependent on the relative differences vis-à-vis the nearest available alternatives, each of the units was further matched with the nearest similarly-sized public facility as well as a village half-way between the catchment areas of each private-public pair, resulting in a final sample of 24 facilities and 12 neighboring villages.

In the villages, we randomly selected 12 adult respondents from a previously recorded household census. At the facilities, accidental sampling was used instead, interviewing the first 10 people exiting the facility on an unannounced day. In total, 384 interviews were carried out in each wave. Excluding respondents who had not visited the catholic facilities in the three years prior to the interview – resulted in 430 non-incidentally truncated interviews. From answers to a set of questions regarding various aspects of

<sup>13</sup> The survey instrument, data and do-files will gladly be shared by authors upon request.

**Table 5**  
Survey sample balance.

	Control		PBF		Diff.	t
	Mean	Std. dev.	Mean	Std. dev.		
2012						
<i>quality</i>	0.877	0.098	0.830	0.134	−0.047	2.7825***
<i>health</i>	0.560	0.278	0.499	0.286	−0.061	1.5326
<i>assets</i>	0.267	0.186	0.268	0.200	0.001	−0.0420
<i>primary</i>	0.383	0.489	0.455	0.500	0.072	−1.0294
<i>female</i>	0.713	0.455	0.718	0.452	0.005	−0.0851
<i>age</i>	37.438	19.008	35.857	14.685	−1.581	0.653
<i>govt</i>	0.407	0.005	0.429	0.007	0.022	2.5974**
2014						
<i>quality</i>	0.904	0.101	0.903	0.110	−0.001	0.1028
<i>health</i>	0.607	0.280	0.624	0.278	0.017	−0.4787
<i>assets</i>	0.274	0.176	0.227	0.164	−0.047	2.1539**
<i>primary</i>	0.431	0.498	0.489	0.502	0.057	−0.8851
<i>female</i>	0.734	0.444	0.832	0.375	0.098	−1.8556*
<i>age</i>	41.633	16.562	35.827	15.580	−6.106	2.9377***
<i>govt</i>	0.362	0.004	0.390	0.006	0.028	3.8334***

Notes: \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

perceived quality of service, we factor out an index of perceived quality (*quality*), scaled from 0 (most dissatisfied) to 1 (most satisfied). Other personal-level confounding characteristics measured through the survey include a physical health index (*health*) based on the SF-12 Health Survey as proposed by Ware, Kosinski, and Keller (1996), an asset index (*assets*) obtained by principal factor analysis (Sahn & Stifel, 2003) approximating wealth, education (*primary*) – a dummy equal to one if the respondent completed primary education, sex (*female*) – a dummy equal to one if the respondent is female, and age (*age*) in years. Finally, *govt* is the mean perceived quality of the matched state-owned facility measured in the same way as *quality*. By using the mean perceived quality of the main public facility rather than the quality perceived by each respondent, we de facto control for the relative reputation of the public competitor. As can be seen in Table 5, the sample was well balanced across all the confounding characteristics (except for *govt*) in 2012.

Lacking a baseline survey, we cannot estimate the effect of the introduction of PBF on the perceived quality of healthcare provision in a difference-in-differences (DID) setting. We can, however, use the DID approach to compare trends in PBF and non-PBF facilities after the new financing system was introduced and after the management and staff of the facilities had presumably gotten accustomed to it.

A meaningful measure of perceived quality of the health facilities can only be obtained from respondents who had recently received treatment there. This can potentially introduce a serious self-selection bias. Before proceeding to a full DID model specification for comparing the trends in perceived quality of healthcare delivery in PBF and non-PBF facilities, we therefore first estimate a bivariate sample-selection model as proposed by Heckman (1979) to check for such bias. We then estimate the PBF trend effects using a simple DID model with cluster-robust standard errors on the sub-sample of non-incidentally truncated interviews, gradually adding covariates until reaching the following fully specified model:

$$quality_{it} = \alpha + \beta_1 PBF_i + \beta_2 (PBF_i \times FU_t) + \beta_3 FU_t + \gamma' \mathbf{X}_i + \delta' \mathbf{Z}_{it} + \zeta govt_i + \eta_i + \varepsilon_{it} \quad (8)$$

where *FU* is a dummy equal to one for all observations from the 2014 follow-up,  $\mathbf{X}$  is a vector of production factors proxying the size of the private facility (*expenditures*, *staff* and *beds*),  $\mathbf{Z}$  is a vector of respondent characteristics comprising *health* (a physical health

index), *assets* (an asset index approximating wealth), *primary* (a dummy equal to one if the respondent completed primary education), *female* (a dummy equal to one if the respondent is female), and *age* (respondent's age in years). *govt* is the mean perceived quality of the nearby state-owned health facility,  $\eta$  are location type fixed effects indicating whether the interview took place at a private facility, a public facility, or a village in between the two, and  $\varepsilon$  is a stochastic error term.

In Table 6 we first estimate a Heckman bivariate sample-selection model to check for self-selection into visiting the private facilities. While some of the observed respondent characteristics affect the likelihood of visiting the private facility within 3 years prior to the interview, the coefficient on the inverse Mills ratio is statistically insignificant (Table 6, column 1). In other words, the self-selection bias in the sample does not affect the coefficient estimates in the second stage. It is therefore safe to estimate the PBF trend effects using a simple DID model with cluster-robust standard errors on the sub-sample of non-incidentally truncated interviews. Starting with a naïve specification which only includes the DID terms and size controls (Table 6, column 2), we gradually include respondent characteristics (Table 6, column 3), the mean level of perceived quality of the nearest similarly-sized public facility, and controls for the location and timing of the interview (Table 6, column 4).

The estimates of perceived quality are stable across the specifications. Perceived quality of healthcare increased between 2012 and 2014 in the PBF as well as non-PBF facilities (see coefficients on *FU* in Table 6). There is also some evidence that the increase was much more pronounced in the case of PBF facilities (see coefficients on *PBF* × *FU* in Table 6). While we do not have quantitative data on staff motivation, we substantiated this result by conducting in-depth interviews with managerial as well as rank-and-file staff in the PBF facilities. These interviews suggest the additional salary bonuses which became possible thanks to PBF made staff feel more appreciated for their efforts. We interpret these results as indicative evidence that PBF did not lead to a decline in quality, and if anything increased the pace of improvements in perceived quality. However as mentioned earlier, this does not allow us to exclude possible negative changes in the performance of the services, either unnoticed by the patients or not captured by our survey instrument.

**Result 3.** Efficiency improvements resulting from PBF do not reduce the quality of care as perceived by patients.

**Table 6**  
PBF effects on trend in perceived quality of healthcare – regressions.

Dep. var.	(1)		(2)	(3)	(4)
	private	quality	quality	quality	quality
<i>PBF</i> × <i>FU</i>		0.037 (0.022)	0.038** (0.016)	0.035** (0.016)	0.038** (0.015)
<i>PBF</i>		−0.045*** (0.017)	−0.047 (0.026)	−0.040 (0.026)	−0.035 (0.026)
<i>FU</i>		0.031* (0.016)	0.030** (0.011)	0.028* (0.013)	0.031* (0.014)
Facility size	No	Yes	Yes	Yes	Yes
Respondent characteristics	Yes	No	No	Yes	Yes
Competitor quality	Yes	No	No	No	Yes
Location type FE	No	No	No	No	Yes
SE clustered by facility pair	No	No	Yes	Yes	Yes
Mean dep. var.		0.880	0.880	0.880	0.880
Std. dev. dep. var.		(0.115)	(0.116)	(0.115)	(0.115)
N	743	430	440	430	430
Mills $\lambda$		−0.027 (0.050)			
$R^2$			0.069	0.082	0.085

Notes: Robust SE in parentheses. \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

## 7. Discussion and conclusions

Using a panel of output and expenditure data from small healthcare facilities in Uganda, we estimate the contribution of performance-based financing towards achieving greater efficiency in faith-based nonprofit healthcare delivery. This is the first study on PBF to focus explicitly on RNPOs. The main strength of our analysis comes from using a combination of parametric and non-parametric econometric techniques. We rely on a dynamic panel estimation to verify the robustness of more traditional two-stage DEA and SFA procedures—the first time frontier efficiency methods have been used to assess a PBF intervention to the best of our knowledge. We find that the technical efficiency of the health facilities in our sample is on average low at 44%. This number could potentially be underestimated considering the quality of our dataset and the susceptibility of DEA to biased outliers. It is about half of the efficiency typically found in health facilities in developed countries (Hollingsworth, 2008; O'Neill et al., 2008), and somewhat lower than the efficiency scores found elsewhere in Africa (Akazili, Adjuik, Jehu-Appiah, & Zere, 2008; Kirigia, Emrouznejad, Sambo, Munguti, & Liambila, 2004; Masiye, 2007; Renner et al., 2005; Tlotlego, Nonvignon, Sambo, Asbu, & Kirigia, 2010; Zere et al., 2006). Nevertheless, the large performance differences we find between individual facilities in our sample with a standard deviation of 24%, which are smaller than in developed countries (12%, Hollingsworth, 2008), but largely comparable to those elsewhere in Africa (between 23% and 33% [Kirigia et al., 2004]), suggest that there is significant space for improvement. We indeed find just that, observing that healthcare providers respond strongly to targets by increasing output through improved efficiency. In our case, output rose by at least 27%. The result is statistically and economically significant and in line with that of other studies of PBF. Moreover, the increased efficiency does not come at the expense of perceived quality. Efforts to improve the performance of health facilities seem to pay-off egregiously.

Jointly we interpret these results as evidence that faith-based healthcare is far from its productivity frontier, and responds to extrinsic incentives. This does not disprove that faith-based healthcare is intrinsically motivated. Quite on the opposite, we believe that intrinsic motivation plays a vital role in all healthcare provision since the times of the Hippocratic Oath, regardless of ownership status. Our results point precisely to the fact that faith-based

non-profit, public and private for-profit outfits are probably not that different after all. Intrinsic motivation may still be important, but we bring evidence that extrinsic incentives have an even greater role to play in increasing overall efficiency.<sup>14</sup> Qualitative interviews with staff in PBF facilities – in line with evidence from several recent evaluations of PBF programs (Olafsdottir et al., 2014; Bertone et al., 2016; Lohmann et al., 2018) – suggest quite the opposite effect of crowding-out: PBF made staff feel appreciated for their efforts, and played an important role in increasing output. If well designed, performance-based financing can help mission-driven healthcare services do more at a lower unit-cost, probably without crowding-out motivation.

Existing literature raises several other concerns about the potential pitfalls of introducing a PBF scheme, all very salient to our context and analysis: Lu (1999) points out that incentivized output targets may encourage facility administrators to over-report output. While this is in general possible, we are confident that such misreporting is unlikely in our case. UCMB datasets have long been used at the national level as an advocacy device, requiring high levels of reliability, and as an internal means of performance comparison across facilities, inducing all facilities to minimize unintended under-reporting regardless of the financing mechanism at play. Moreover, unlike other healthcare providers in Uganda, UCMB engages in monthly data verification and monitoring activities throughout the country. To spot any plausible fallacies in the reporting, it also activated additional post-episode-of-care verifications at the community level for PBF facilities.

Another potential issue arises from the way in which the targets are aggregated from partial output indicators. If the relative weight given to individual output indicators does not reflect their actual cost in terms of factors of production, an introduction of weighted output targets such as the present one defined in terms of the SUO could distort the balance of provided services, potentially resulting in the production of services with negative marginal value (Wynia,

<sup>14</sup> In fact, while the output increases observed by Reinikka and Svensson (2010) rely on a budget expansion, our results net out the effect of budgetary increases. It is important to stress that any effect we estimate is budget-neutral, in that it removes the effect on output that can be directly derived from the increase in budget. This is especially salient given that the budget available for PBF in our setting may be small relative to other similar interventions. See Soeters (2017) for an overview of PBF standards.

2009). However, the three output indicators which constitute the largest part of total SUO production in observed facilities – the number of *outpatients* (52% of total SUO production), *inpatients* (36%) and *immunizations* (8%) – all rose at roughly the same rate following the introduction of PBF.

Further research would benefit from an earlier engagement in data collection on quality of service and intrinsic motivation, to observe actual changes introduced by PBF. Also, a pipeline randomization of the roll-out of the interventions would assuage some of the concerns about non-random placement. Finally, while our study shows that PBF has the potential to increase efficiency of faith-based healthcare delivery substantially, it has little to say about which incentives structures are best suited to do so. Future interventions could vary incentives to test which are most effective in mission-driven healthcare.

### Conflict of interest

None.

### Acknowledgments

We would like to thank Dr. Sam Orach of the Uganda Catholic Medical Bureau (UCMB) for making the input and output data of UCMB health centers available to us. We are also grateful to our team of enumerators for their dedicated work, and to Aisha Nanyiti for supervision of the enumeration team. The fieldwork for this research would not have happened without the financial support of the Netherlands Organisation for Scientific Research (project number W 07.72.108), and the logistical support of Muna Siraje. Finally, we would like to thank the two anonymous referees for their comments which helped us to considerably improve this paper. All remaining errors are our own.

### References

- Akazili, J., Adjuik, M., Jehu-Appiah, C., & Zere, E. (2008). Using data envelopment analysis to measure the extent of technical efficiency of public health centres in Ghana. *BMC International Health and Human Rights*, 8(11), 1–12.
- Andaleeb, S. S. (2001). Service quality perceptions and patient satisfaction: A study of hospitals in a developing country. *Social Science & Medicine*, 52(9), 1359–1370.
- Arellano, M. (1993). On the testing of correlated effects with panel data. *Journal of Econometrics*, 59(1–2), 87–97.
- Banerjee, A. V., Glennerster, R., & Duflo, E. (2008). Putting a band-aid on a corpse: Incentives for nurses in the Indian public health care system. *Journal of the European Economic Association*, 6, 487–500.
- Basinga, P., Gertler, P. J., Binagwaho, A., Soucat, A. L. B., Sturdy, J., & Vermeersch, C. M. J. (2011). Effect on maternal and child health services in Rwanda of payment to primary health-care providers for performance: An impact evaluation. *Lancet*, 377(9775), 1421–1428.
- Battese, G. E., & Coelli, T. J. (1995). A model for technical inefficiency effects in a stochastic frontier production function for panel data. *Empirical Economics*, 20, 325–332.
- Bertone, M. P., Lagarde, M., & Witter, S. (2016). Performance-based financing in the context of the complex remuneration of health workers: Findings from a mixed-method study in Rural Sierra Leone. *BMC Health Services Research*, 16(1).
- Bhatnagar, A., & George, A. S. (2016). Motivating health workers up to a limit: Partial effects of performance-based financing on working environments in Nigeria. *Health Policy and Planning*, 31(7), 868–877.
- Björkman, M., & Svensson, J. (2009). Power to the people: Evidence from a randomized field experiment on community-based monitoring in Uganda. *Quarterly Journal of Economics*, 124(2), 735–769.
- Bloom, D. E., & Canning, D. (2000). The health and wealth of nations. *Science*, 287(5456), 1207–1209.
- Bonfrer, I., Van de Poel, E., & Van Doorslaer, E. (2014). The effects of performance incentives on the utilization and quality of maternal and child care in Burundi. *Social Science & Medicine*, 123, 96–104.
- Brenzel, L. (2009). Taking stock: World bank experience with results-based financing for health.
- Brock, J. M., Lange, A., & Leonard, K. L. (2016). Generosity and prosocial behavior in healthcare provision: Evidence from the laboratory and field. *Journal of Human Resources*, 51(1), 133–162.
- Centers for Medicare and Medicaid Services. (2011). CAHPS Hospital Survey (HCAPS): Quality Insurance Guidelines Version 6.0.
- 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(23), 1–12.
- Choi, I. (2001). Unit root tests for panel data. *Journal of International Money and Finance*, 20(2), 249–272.
- Davies, A. R., & Ware, J. E. (1988). Involving consumers in quality of care assessment. *Health Affairs*, 7(1), 33–48.
- Deci, E. L., & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior*. New York: Springer.
- Eichler, R. (2006). Can pay for performance increase utilization by the poor and improve the quality of health services.
- Eldridge, C., & Palmer, N. (2009). Performance-based payment: Some reflections on the discourse, evidence and unanswered questions. *Health Policy and Planning*, 24(3), 160–166.
- Frey, B. S., & Jegen, R. (2001). Motivation crowding theory. *Journal of Economic Surveys*, 15(5), 589–611.
- Fritsche, G. B., Soeters, R., & Meessen, B. (2014). *Performance-based financing toolkit*. Washington, D.C.: The World Bank.
- Government of Uganda – Ministry of Health. (2010). Health Sector Strategic Plan III: 2010/11–2014/15.
- Hecht, R., Batson, A., & Brenzel, L. (2004). Making health care accountable: Why performance-based funding in developing countries is getting more attention. *Finance & Development* (March), 16–19.
- Heckman, J. (1979). Sample selection bias as a specification error. *Econometrica*, 47(1), 153–161.
- Hollingsworth, B. (2008). The measurement of efficiency and productivity of health care delivery. *Health Economics*, 17(8), 1107–1128.
- Honda, A. (2013). 10 Best resources on ... pay for performance in low- and middle-income countries. *Health Policy and Planning*, 28(5), 454–457.
- Hsiao, C. (2003). *Analysis of panel data*. Cambridge: Cambridge University Press.
- Im, K. S., Pesaran, M. H., & Shin, Y. (2003). Testing for unit roots in heterogeneous panels. *Journal of Econometrics*, 115(1), 53–74.
- 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(9), 695–698.
- Johansson, P., Oléni, M., & Fridlund, B. (2002). Patient satisfaction with nursing care in the context of health care: A literature study. *Scandinavian Journal of Caring Sciences*, 16(4), 337–344.
- Josephson, E., Gergen, J., Coe, M., Ski, S., Madhavan, S., & Bauhoff, S. (2017). How do performance-based financing programmes measure quality of care? A descriptive analysis of 68 quality checklists from 28 low- and middle-income countries. *Health Policy and Planning*, 32(8), 1120–1126.
- Kalk, A. (2011). The costs of performance-based financing. *Bulletin of the World Health Organization*, 89, 319.
- Kalk, A., Paul, F. A., & Grabosch, E. (2010). 'Paying for Performance' in Rwanda: Does It Pay off? *Tropical Medicine and International Health*, 15(2), 182–190.
- Kirigia, J. M., Emrouznejad, A., Sambo, L. G., Munguti, N., & Liambila, W. (2004). Using data envelopment analysis to measure the technical efficiency of public health centers in Kenya. *Journal of Medical Systems*, 28(2), 155–166.
- Leonard, K. K. L., & Masatu, M. C. M. (2010). Professionalism and the know-do gap: Exploring intrinsic motivation among health workers in Tanzania. *Health Economics*, 19, 1461–1477.
- Leonard, K. L. (2008). Is patient satisfaction sensitive to changes in the quality of care? An exploitation of the Hawthorne effect. *Journal of Health Economics*, 27(2), 444–459.
- Lohmann, J., Houffort, N., & De Allegri, M. (2016). Crowding out or no crowding out? A self-determination theory approach to health worker motivation in performance-based financing. *Social Science & Medicine*, 169, 1–8.
- Lohmann, J., Wilhelm, D., Kambala, C., Brenner, S., Muula, A. S., & De Allegri, M. (2018). 'The money can be a motivator, to me a little, but mostly PBF just helps me to do better in my job'. An exploration of the motivational mechanisms of performance-based financing for health workers in Malawi. *Health Policy and Planning*, 33(2), 183–191.
- Lu, M. (1999). Separating the true effect from gaming in incentive-based contracts in health care. *Journal of Economics & Management Strategy*, 8(3), 383–431.
- Marley, K. A., Collier, D. A., & Meyer Goldstein, S. (2004). The role of clinical and process quality in achieving patient satisfaction in hospitals. *Decision Sciences*, 35(3), 349–369.
- Masiye, F. (2007). Investigating health system performance: An application of data envelopment analysis to Zambian hospitals. *BMC Health Services Research*, 7(58), 1–11.
- Maynard, A. (2012). The powers and pitfalls of payment for performance. *Health Economics*, 12, 3–12.
- Meessen, B., Kashala, J.-P. I., & Musango, L. (2007). Output-based payment to boost staff productivity in public health centres: Contracting in Kabutare District, Rwanda. *Bulletin of the World Health Organization*, 85(2), 108–115.
- Meessen, B., Musango, L., Kashala, J.-P. I., & Lemlin, J. (2006). Reviewing institutions of rural health centres: The performance initiative in Butare, Rwanda. *Tropical Medicine & International Health*, 11(8), 1303–1317.
- Miller, G. J. (2008). Solutions to principal-agent problems in firms. In C. Ménard & M. Shirley (Eds.), *Handbook of new institutional economics* (pp. 349–370). Berlin: Springer.
- Morgan, L. (2010). Some days are better than others: Lessons learned from Uganda's first results-based financing pilot.

- Novignon, J., & Nonvignon, J. (2017). Improving primary health care facility performance in Ghana: Efficiency analysis and fiscal space implications. *BMC Health Services Research*, 17(1), 1–8.
- Olafsdottir, A. E., Mayumana, I., Mashasi, I., Njau, I., Mamdani, M., Patouillard, E., ... Borghi, J. (2014). Pay for performance: An analysis of the context of implementation in a pilot project in Tanzania. *BMC Health Services Research*, 14(1), 1–9.
- O'Neill, L., Rauner, M., Heidenberger, K., & Kraus, M. (2008). A cross-national comparison and taxonomy of DEA-based hospital efficiency studies. *Socio-Economic Planning Sciences*, 42, 158–189.
- Oxman, A. D., & Fretheim, A. (2008). An overview of research on the effects of results-based financing.
- Peabody, J. W., Tozija, F., Munoz, J. A., Nordyke, R. J., & Luck, J. (2004). Using vignettes to compare the quality of clinical care variation in economically divergent countries. *Health Services Research*, 39(6), 1951–1970.
- Reinikka, R., & Svensson, J. (2010). Working for god? Evidence from a change in financing of nonprofit health care providers in Uganda. *Journal of the European Economic Association*, 8(6), 1159–1178.
- Renmans, D., Holvoet, N., & Criel, B. (2017a). Combining theory-driven evaluation and causal loop diagramming for opening the 'black box' of an intervention in the health sector: A case of performance-based financing in Western Uganda. *International Journal of Environmental Research and Public Health*, 14(9), 1–20.
- Renmans, D., Holvoet, N., Criel, B., & Meessen, B. (2017b). Performance-based financing: The same is different. *Health Policy and Planning*, 32(6), 860–868.
- Renmans, D., Holvoet, N., Orach, C. G., & Criel, B. (2016). Opening the 'black box' of performance-based financing in low- and lower middle-income countries: A review of the literature. *Health Policy and Planning*, 31(9), 1297–1309.
- Renner, A., Kirigia, J. M., Zere, E. A., Barry, S. P., Kirigia, D. G., Kamara, C., & Muthuri, L. H. (2005). Technical efficiency of peripheral health units in Pujehun District of Sierra Leone: A DEA application. *BMC Health Services Research*, 5(77), 1–11.
- Roodman, D. (2009). How to Do xtabond2: An introduction to difference and system GMM in Stata. *Stata Journal*, 9(1), 86–136.
- Ryan, A. (2009). Hospital-based pay-for-performance in the United States. *Health Economics*, 18, 1109–1113.
- Sahn, D., & Stifel, D. (2003). Exploring alternative measures of welfare in the absence of expenditure data. *Review of Income and Wealth*, 49(4), 463–489.
- Scott, A., & Coote, W. (2010). Do regional primary-care organisations influence primary-care performance? A dynamic panel estimation. *Health Economics*, 19, 716–729.
- Sekabaraga, C., Diop, F., & Soucat, A. (2011). Can innovative health financing policies increase access to MDG-related services? Evidence from Rwanda. *Health Policy and Planning*, 26(Suppl 2), ii52–ii62.
- Shen, G. C., Nguyen, H. T. H., Das, A., Sachingongu, N., Chansa, C., Qamruddin, J., & Friedman, J. (2017). Incentives to change: Effects of performance-based financing on health workers in Zambia. *Human Resources for Health*, 15(1), 1–15.
- Simar, L., & Wilson, P. W. (2007). Estimation and inference in two-stage, semi-parametric models of production processes. *Journal of Econometrics*, 136(1), 31–64.
- Soeters, R. (Ed.). (2017). *Performance-based financing in action: Theory and instruments*. The Hague: SINA Health.
- Soeters, R., Peerenboom, P. B., Mushagalusa, P., & Kimanuka, C. (2011). Performance-based financing experiment improved health care in the democratic Republic of Congo. *Health Affairs*, 30(8), 1518–1527.
- Ssengooba, F., McPake, B., & Palmer, N. (2012). Why performance-based contracting failed in Uganda – An open-box evaluation of a complex health system intervention. *Social Science & Medicine*, 75(2), 377–383.
- Tlotlego, N., Nonvignon, J., Sambo, L. G., Asbu, E. Z., & Kirigia, J. M. (2010). Assessment of productivity of hospitals in Botswana: A DEA application. *International Archives of Medicine*, 3(1), 1–14.
- Turcotte-Tremblay, A.-M., Spagnolo, J., De Allegri, M., & Ridde, V. (2016). Does performance-based financing increase value for money in low- and middle-income countries? A systematic review. *Health Economics Review*, 6(1), 30.
- Walton, T., & Matthees, K. (2017). Improving emerging markets healthcare through private provision.
- Ware, J. E. J., Kosinski, M., & Keller, S. D. (1996). A 12-item short-form health survey of scales and preliminary construction tests of reliability and validity. *Medical Care*, 34(3), 220–233.
- Weil, D. N. (2007). Accounting for the effect of health on economic growth. *Quarterly Journal of Economics*, 122(3), 1265–1306.
- Wharam, J. F., Paasche-Orlow, M. K., Farber, N. J., Sinsky, C., Rucker, L., Rask, K. J., ... Sulmasy, D. P. (2009). High quality care and ethical pay-for-performance: A society of general internal medicine policy analysis. *Journal of General Internal Medicine*, 24(7), 854–859.
- Wolszczak-Derlacz, J., & Parteka, A. (2011). Efficiency of European public higher education institutions: A two-stage multicountry approach. *Scientometrics*, 89(3), 887–917.
- Wooldridge, J. M. (2002). *Econometric analysis of cross section and panel data*. Cambridge, MA: The MIT Press.
- World Health Organization (2018). Global health expenditure database.
- Wynia, M. K. (2009). The risks of rewards in health care: How pay-for-performance could threaten, or bolster, medical professionalism. *Journal of General Internal Medicine*, 24(7), 884–887.
- Zere, E., Mbeeli, T., Shangula, K., Mandlhate, C., Mutirua, K., Tjivambi, B., & Kapenambili, W. (2006). Technical efficiency of district hospitals: Evidence from Namibia using data envelopment analysis. *Cost Effectiveness and Resource Allocation*, 4(5), 1–9.
- Zhao, H. (2005). Incentive-based compensation to advertising agencies: A principal-agent approach. *International Journal of Research in Marketing*, 22(3), 255–275.