Evaluating the Impact of Performance-Based Financing in Healthcare Provision: A Case Study in Uganda

Jan Duchoslav* Francesco Cecchi*†

Working paper
9 September 2016
Click here for the latest version

Abstract

We estimate the effect of introducing performance-based financing on the allocative efficiency and quality of healthcare provision. We use a unique panel of output and expenditure data from small private not-for-profit healthcare facilities in Uganda in combination with a patient satisfaction survey focused on the perceived quality of healthcare. In the observed facilities, performance-based financing increases output and allocative efficiency by more than 25% with no apparent reduction in quality. The result is robust to more traditional estimation techniques.

1 Introduction

Improvements in healthcare provision form one of the cornerstones of economic development: 3 of the 8 Millennium Development Goals were directly related to healthcare, and the new Sustainable Development Goals call for increased health financing in developing countries. There is certainly room for improvement in this respect. According to the World Bank (2016), low-income countries spent 7 times less money per capita on health than middle-income ones, and 26 times less than non-OECD high-income countries in 2014. However, when expressed as a percentage of GDP, health expenditures were roughly the same in low-income countries (5.7%) as in middle- and high-income non-OECD countries (5.8% and 5.5% respectively).[1] If low-income countries are to increase spending on health, they must either divert it from elsewhere, or wait for economic growth to increase the total available funds. Alternatively, they can try to do more with less—to increase allocative efficiency.

Performance-based financing (PBF) is becoming a popular measure to do just that (Hecht et al., 2004; Brenzel, 2009; Eldridge and Palmer, 2009; Honda, 2013). As the term suggests, PBF makes the amount of funding a healthcare provider receives conditional on performance, most commonly defined in terms of output. Healthcare providers should respond positively to such incentive schemes as long as the marginal revenue offsets the marginal costs associated with increased output (Rusa et al., 2009). While this mechanism is straightforward in theory, the evidence that it works well in practice is inconclusive, and mostly focuses on partial outcomes (Eijkenaar et al., 2013). Moreover, Scheffler (2010) points out

*Development Economics Group, Wageningen University
†Faculty of Economics and Business, University of Groningen

[1] High-income non-OECD countries include for example Argentina, Russia or Saudi Arabia. OECD countries, whose healthcare spending figures are somewhat inflated by the United States, spent 12.6% of their GDP – or 149 times more per capita – than their low-income counterparts.

that most evaluations lack proper control groups, or confound the effect of performance incentives with additional elements of the program that might impact practices. Even the most rigorously designed randomized control trials (RCTs) often prove to be unworkable, or prone to last-minute political and organizational interferences that weaken the validity of the findings, especially in the context of developing countries (Soeters et al., 2006). In 2012, an editorial of the Bulletin of the World Health Organization about PBF in low- and middle-income countries advocated the use of other robust evaluation designs if conducting an RCT is not feasible. One option is an interrupted time series, in which outcome data are collected at regular intervals during baseline and post-intervention periods (Fretheim et al., 2012). We take this approach in the present paper, using a panel dataset from a network of private not-for-profit health facilities in Uganda to estimate the effects of introducing PBF on healthcare output and – by extension – on the allocative efficiency of healthcare service delivery.

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. The results are presented in Section 5 and discussed in Section 6. Section 7 concludes.

2 Background

2.1 PBF: a policy to improve healthcare provision

The growing evidence that the health of its population is an important determinant of a country's economic growth (Bloom and Canning, 2000; Weil, 2007) has provided an additional argument – besides the ethical ones – for the need for functional and accessible healthcare provision. This in turn is essentially a function of structural inputs (including people, infrastructure, knowledge, drugs, material, equipment, and technology) and 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 et al., 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 (Hecht et al., 2004; Brenzel, 2009; Eldridge and Palmer, 2009; Honda, 2013). However, while the principal-agent problem has been successfully reduced by conditioning payment on performance in many other contexts, it is rather difficult in processes with such multi-dimensional output as healthcare, the production of which is complicated to quantify.

Even if performance is understood in its most limited sense as output, 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 of produced output naturally complicates the matter even further. Some PBF schemes have nonetheless made attempts to condition funding on performance in quality indicators (Perrot et al., 2010). In Basinga et al. (2011) for example, quality is measured as an index of observable structural and process measures, and enters the payment formula as a multiplicative factor that may lower the output-based payment. Soeters et al. (2011) instead try 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.

Besides the various ways in which they define and measure performance, PBF schemes differ along three major lines (Perrot et al., 2010):

1. performance targets and associated payments may apply to individual employees or to entire facilities;

2. conditional payments may make up the entire funding of the facility (or salary of the individual) or there may be a fixed component;
3. and payments may be conditional on fixed targets or incremental.

Each PBF program is thus quite unique, and operates in a unique setting. It is therefore hard – if not impossible – to agree on best practice (Ireland et al., 2011). 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.

Existing literature 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. Other concerns include the risk of increased gaming, i.e. systematic reporting bias (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 et al., 2011). Finally, the bureaucratization of healthcare delivery, which is – to an extent – necessary to implement a PBF program, may end up crowding out intrinsic motivation (Frey and Jegen, 2001), inducing a decline in physician professionalism and morale (Wharam et al., 2009).

Despite these issues, some of which we touch upon in the discussion, PBF is receiving increasing attention from academics as well as policy-makers. Several studies document positive effects of PBF on partial outcomes (Meessen et al., 2006, 2007; Soeters et al., 2011; Basinga et al., 2011; Sekabaraga et al., 2011; Bonfrer et al., 2014), while others find no lasting effects (Banerjee et al., 2008; Morgan, 2010). The evidence to date is thus inconclusive (Eijkenaar et al., 2013), with even the most rigorous studies suffering from various methodological shortcomings (Fretheim et al., 2012).

2.2 Setting

The health sector in Uganda is characterized by a high degree of fragmentation with a mixture of public, private not-for-profit and private for-profit healthcare providers (Björkman and Svensson, 2009), with government health spending representing only 30% of the total (Hunt, 2010). 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 (Governemt 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 (Morgan, 2010).

In this paper, we focus on one of the largest not-for-profit 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 (Governemt 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 mid-sized health centers[2] to test the practical feasibility of this new approach

[2]There are several levels of healthcare facilities in Uganda: dispensaries and aid posts (HC I), health centers (HC II-IV from the smallest to the largest ones), and hospitals (HC V) (Björkman and Svensson, 2009). The pilot PBF scheme targets solely the mid-level HCs III and IV. We consequently limit our analysis to units of the same size.

before possibly extending the scheme to all its health units. The Diocese of Jinja was selected for its manageable size, for its long cooperation with an international NGO which helped set the scheme up, and for its historically good performance.[3]

The scheme was introduced in the 2008/2009 fiscal year. For each health unit, it set an output target based on its performance in the previous 3 years. The targets are defined in terms of the Standard Unit of Output (SUO) – a weighted average of the most commonly performed procedures and services (see Section 3 for full details). Starting in fiscal year 2009/2010, each of the participating facilities would receive an incremental bonus, conditional on reaching its pre-specified target. On average, the bonus would increase the income of the facilities – about a third of which comes from user fees and the remainder from government grants – by about 5%. The targets were set sufficiently low for all units to reach them every year, effectively making the scheme a fully incremental one. Although the targets are set and the bonuses paid at the unit level, the heads of the units typically use 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 Data

To gauge the effect of PBF on the allocative 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.

Output is measured using the Standard Unit of Output (SUO) – a unit in which the PBF targets are defined. The SUO is constructed from common services typically provided by small health facilities, taking into account their relative input requirements in terms of cost and time:

$$SUO = outpatients + 5 \times inpatients + 2 \times deliveries + 0.3 \times ANC + 0.2 \times immunizations \qquad (1)$$

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. Total $expenditures$ are measured in millions of 2014 Ugandan Shillings (USh.). Panel descriptive statistics for 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. Figure 1 shows the trends in output of PBF and control facilities. It plots a two-year moving average of output for PBF and control facilities after standardizing each facility's output at 100 at the beginning of the observed period.[4] The dotted vertical line marks the introduction of the PBF scheme.

[3]Due to the historically good performance of the pilot facilities, parallel trends cannot be assumed, precluding the possibility of using standard difference-in-differences approaches when estimating the effect of PBF. We instead use a panel-data structural break approach (see Section 4).

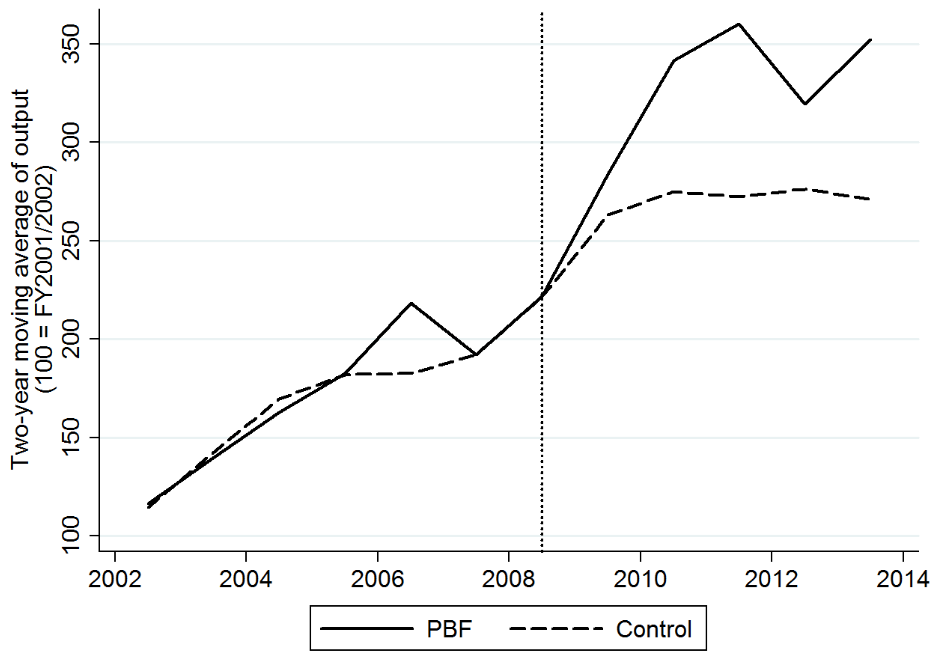
[4]Using a moving average instead of a simple one smooths the plots, especially in the case of the PBF facilities, which are few in number. Standardizing the output of each facility to 100 in fiscal year 2001/2002 prevents gaps in the data from affecting the average.

Table 1: Panel descriptive statistics

year	Obs.	PBF	*SUO*	*expenditures*	*staff*	*beds*
2001/02	79	no	10,471	78.1	11.7	19.8
			(8,753)	(95.3)	(7.4)	(16.7)
2002/03	60	no	10,892	79.9	12.0	22.6
			(10,309)	(56.9)	(8.0)	(16.8)
2003/04	120	no	11,982	86.7	11.7	19.4
			(8,760)	(88.1)	(7.8)	(15.6)
2004/05	122	no	13,587	93.9	12.1	20.7
			(9,543)	(91.2)	(8.3)	(16.0)
2005/06	145	no	12,798	86.4	10.8	19.8
			(9,917)	(89.3)	(7.4)	(15.7)
2006/07	151	no	13,180	79.1	11.0	20.3
			(10,012)	(75.6)	(7.7)	(15.7)
2007/08	170	no	12,956	72.3	10.4	19.5
			(10,729)	(77.2)	(7.4)	(16.9)
2008/09	167	yes	15,533	75.9	9.8	18.3
			(12,370)	(79.9)	(6.6)	(15.2)
2009/10	184	yes	18,361	97.7	10.7	21.3
			(14,918)	(127.0)	(6.7)	(16.4)
2010/11	184	yes	15,635	91.1	11.0	20.8
			(12,963)	(105.2)	(7.6)	(16.0)
2011/12	173	yes	14,660	91.3	11.2	22.0
			(11,803)	(135.4)	(8.3)	(16.0)
2012/13	184	yes	15,063	85.0	11.5	21.9
			(11,799)	(111.6)	(6.9)	(16.1)
2013/14	164	yes	15,042	92.4	12.1	22.3
			(10,295)	(116.1)	(7.7)	(15.8)

Notes: Standard deviations in parentheses.

Figure 1: Output trends



4 Empirical strategy

When asked to estimate the effect of the introduction of a new technology – such as PBF – on productivity and efficiency, a development microeconomist or a student of industrial organization might instinctively turn to estimating the parameters of a production function (with an additional technology-shifting term indicating whether or not PBF is in place) by means of regression analysis or some more general parametric approach such as the generalized method of moments (GMM). Yet, most such analyses of healthcare production typically involve data envelopment analysis (DEA) or stochastic frontier analysis (SFA) approaches (Hollingsworth, 2008). Each approach has its advantages as well as drawbacks.

DEA does not require one to make 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 data are often sketchy at the best. SFA and various regression methods are considerably less sensitive to outliers and measurement error than DEA, but require one to impose assumptions on the distribution of efficiency or the error term and about the form of the production function. DEA and SFA are commonplace in the field of productivity and efficiency analysis, while the results of standard regression and GMM approaches can be more readily interpreted by a traditionally trained microeconomist. Moreover, GMM allows for healthcare production to be modeled as the dynamic process that it is (Scott and Coote, 2010), bringing the econometric model closer to reality.

In what follows, we start with a simple pooled ordinary least squares (OLS) model, gradually relaxing various assumptions to arrive to a dynamic system GMM model which fully accounts for the dynamics inherent to the healthcare production process. We then compare the results to those obtained from DEA and SFA methods.

4.1 Panel regression analysis and dynamic system GMM

We first estimate a simple pooled OLS model, regressing output on PBF and the equivalent of a translog production function with expenditures, number of staff and capital (proxied for bed capacity) as factors of production. 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:

$$\ln SUO_{it} = \alpha + \beta PBF_{it} + \boldsymbol{\gamma' factors_{it}} + \eta_i + \tau_t + \varepsilon_{it} \tag{2}$$

where PBF_{it} is a dummy equal to one if PBF is in place in the given facility and fiscal year. $\boldsymbol{factors_{it}}$ is a vector representing a translog production function with $expenditures_{it}$ (total expenditures in millions of USh.), $staff_{it}$ (total number of staff), $beds_{it}$ (bed capacity proxying for capital) as production factors. η_i are facility-level fixed effects, τ_t are fiscal year fixed effects, and ε_{it} is a stochastic error term.[5]

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 β 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 $\boldsymbol{\gamma}$, leaving β unaffected. We choose the present specification for the ease of its interpretation as a translog production function.

[5]The full model is thus specified as follows:

$$\begin{aligned}\ln SUO_{it} =& \alpha + \beta PBF_{it} + \gamma_1 \ln expenditures_{it} + \gamma_2 \ln staff_{it} + \gamma_3 \ln beds_{it} \\ &+ \gamma_4 (\ln expenditures)^2_{it} + \gamma_5 (\ln staff)^2_{it} + \gamma_6 (\ln beds)^2_{it} \\ &+ \gamma_7 (\ln expenditures \times \ln staff)_{it} + \gamma_8 (\ln expenditures \times \ln beds)_{it} \\ &+ \gamma_5 (\ln staff \times \ln beds)_{it} + \eta_i + \tau_t + \varepsilon_{it}\end{aligned}$$

Since it can be expected that performance in one period is heavily influenced by performance in previous time periods (Scott and Coote, 2010), we reestimate the model using a three-step feasible generalized least squares estimator (FGLS) to correct for autocorrelation in the data.[6]

The static models developed so far may be further biased by imbalance of the panel—48% of observations are missing. This would not be a major issue if the observations were missing at random. It is however quite conceivable that better-administered units enter the dataset early on, potentially biasing down any time trend. Indeed, the unbalancedness of the panel stems mainly from an expansion of the dataset through time (there are 79 observations in 2001 against 164 in 2014), and to a much lesser extent through attrition and time-series gaps. We verify the possibility of non-random censoring using the procedure proposed by Nijman and Verbeek (1992) to test for non-random attrition in panels. The non-random imbalance of the panel makes it imperative to introduce dynamics into the model. Our final estimator is therefore a panel-robust two-step system GMM, which allows for correcting both autocorrelation and imbalances in the panel.[7]

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 (Oxman and Fretheim, 2008; Maynard, 2012).

4.2 Data envelopment analysis

To verify the robustness of the results, we also estimate the effect of PBF on productivity following the more traditional approaches of DEA- and SFA-based two-stage procedures. 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 θ 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, θ 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:

$$\begin{aligned} &\min_{\theta,\lambda} \theta, \\ &\quad s.t. -\frac{1}{\theta} SUO_i + \boldsymbol{SUO'\lambda} \geq 0, \\ &\quad \boldsymbol{x_i} - \boldsymbol{X'\lambda} \geq 0, \\ &\quad \boldsymbol{N'\lambda} = 1, \\ &\quad \boldsymbol{\lambda} \geq 0 \end{aligned} \tag{3}$$

[6]We test for serial autocorrelation following Wooldridge (2002). We 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 et al. (2003).

[7]In order to eliminate selection bias (and serial autocorrelation), we follow Roodman (2009), using a panel-robust two-step Blundel-Bond system GMM estimator 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.

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

In the second stage, we estimate the effects of the introduction of PBF on θ, 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:

$$\theta_{it} = \alpha + \beta PBF_{it} + \delta_i + \tau_t + \varepsilon_{it} \tag{4}$$

where PBF_{it} equals 1 if the i-th facility had a PBF program in place in fiscal year t and 0 otherwise, δ_i are spatial fixed effects on the diocese level, and τ_t are fiscal year fixed effects.[8]

4.3 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. We use the same translog production function as in (2):

$$\ln SUO_{it} = \alpha + \boldsymbol{\gamma' factors_{it}} + (v_{it} - u_{it}) \tag{5}$$

where 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}:

$$u_{it} = \zeta + \beta PBF_{it} + \delta_i + \tau_t + w_{it} \tag{6}$$

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

$$\ln SUO_{it} = \alpha + \boldsymbol{\gamma' factors_{it}} + (v_{it} - (\zeta + \beta PBF_{it} + \delta_i + \tau_t + w_{it})) \tag{7}$$

5 Results

To estimate the effects of introducing PBF on healthcare output and, as we control for factors of production, also on the allocative efficiency of healthcare service delivery, we start with a naïve pooled OLS model (Table 2, column 1), which yields a large, positive and statistically very significant effect of PBF on output, but is potentially biased as it ignores the panel structure of the dataset. We therefore proceed with estimating a fixed effects model (Table 2, column 2),[9] additionally controlling for a possible time trend by including fiscal year fixed effects (Table 2, column 3).[10] To correct for first-order autocorrelation,[11] we also estimate the model using an FGLS estimator (Table 2, column 4).

[8] To estimate the second stage double-bootstrapped maximum likelihood truncated regression, we adapt Stata code written by Wolszczak-Derlacz and Parteka (2011) and kindly provided by the authors.

[9] A Wald test confirms the presence of facility-level fixed effects ($F(215, 1677) = 5.07$ is significant at the 1% level). 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).

[10] A Wald test confirms the joint significance of fiscal year fixed effects ($F(12, 215) = 12.65$ is significant at the 1% level).

[11] 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

Table 2: PBF effects on output and efficiency

	(1) OLS	(2) FE	(3) FE	(4) FGLS	(5) sysGMM
Dep. var.	$\ln SUO$	$\ln SUO$	$\ln SUO$	$\ln SUO$	$\ln SUO$
PBF	0.734***	0.737***	0.492***	0.509***	0.266**
	(0.129)	(0.123)	(0.131)	(0.095)	(0.130)
$\ln SUO_{it-1}$					0.377***
					(0.053)
Translog inputs	yes	yes	yes	yes	yes
Facility FE	no	yes	yes	yes	no
Diocese FE[13]	no	no	no	no	yes
Time FE	no	no	yes	yes	yes
SE clustered by facility	yes	yes	yes	no	no
Mean dep. var.	9.321	9.321	9.321	9.321	9.372
Std. dev. dep. var.	(0.699)	(0.699)	(0.699)	(0.699)	(0.684)
Obs.	1903	1903	1903	1896	1572
R^2	0.594	0.276	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*. See Table A.1 in Appendix A for full results.

The static models produce large, significant effects of PBF on output and productivity. However. the panel is heavily unbalanced and autocorrelated.[12] To address these issues, our final estimator is a system GMM (Table 2, column 5). Moving from a naïve static pooled OLS model to a fully specified dynamic GMM one, assumptions are progressively relaxed and various problems corrected for, and the magnitude of the estimated effect of the introduction of PBF on healthcare output decreases drastically from a hardly believable value of 73% in model (1) to a more realistic 27% in model (5). Its statistical significance also decreases in the process, but always remains below the 5% level.

Result: The introduction of PBF increases output through improved allocative efficiency of healthcare provision by 27%, or over a third of a standard deviation.

As a robustness check, we also estimate the effect of PBF on productivity using the more traditional two-stage DEA and SFA approaches. The linear program in the first DEA stage produces largely varied efficiency scores θ across the sample of 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.[14] In the second double-bootstrapped maximum-likelihood stage, we estimate that the introduction of PBF increases efficiency θ by 20.1 percentage points (i.e. by 45.8%) – or almost one standard deviation (Table 3, column 1).

Though estimated simultaneously, the SFA approach can effectively be thought of as a two-stage

reject the null hypothesis of non-stationarity at the 1% significance level.

[12] Applying the procedure proposed by Nijman and Verbeek (1992), we find that the censoring of the data is non-random (see Table A.3 in Appendix A for details).

[13] 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.

[14] It is important to note that the magnitude of θ can easily be biased down by outliers. This does not however affect the relative changes due to technology shifters such as PBF.

Table 3: PBF effects on output and efficiency – robustness

	(1) DEA	(2) SFA	
	2nd stage	1st stage	2nd stage
Dep. var.	θ	$\ln SUO$	u
PBF	0.201***		-0.476**
	(0.038)		(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 χ^2 (df)	1045.71 (31)	2877.92 (9)	

Notes: (Cluster) robust (bootstrapped) standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. See Table A.2 in Appendix A for full results.

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 u by 0.476 (Table 3, column 2), which – by a simple arithmetic manipulation – is an equivalent of a 61.0% increase in θ.

The SFA procedure is more robust to noise in the data than DEA, but both techniques are likely to bias the coefficient estimates upwards due to their static nature. Indeed, they produce statistically significant results similar in direction and magnitude to those from regression estimations, but considerably larger in magnitude than estimates from system GMM.

6 Discussion

As demonstrated in Section 5, PBF led to a statistically and economically significant increase of allocative efficiency – and thus output – in the healthcare facilities where it was introduced. A dynamic model results into a much lower estimate (an increase of around 25%) than static models (an increase between 50% and 75%). 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 output and efficiency. In a panel setting, static models which ignore this positive autocorrelation will overestimate the effects of any positive external shock such as an introduction of a PBF scheme. Dynamic estimators like system GMM can overcome this issue and are thus more suitable for evaluating the impact of such shocks.

Although much smaller in magnitude than static estimation methods might suggest, 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. It is quite plausible that such a sudden rise in output could come at the expense of its quality. To see whether this might be an issue, we conducted a two-wave patient satisfaction survey in the PBF facilities and in a sub-sample of those without PBF.[15] Both survey waves took place after PBF had been implemented (in 2012 and 2014), so we cannot infer much about the direct effect of the introduction of PBF on the quality of services. The data nonetheless clearly show that perceived quality of healthcare did not decline between the two waves. In fact, patient satisfaction increased in both PBF and control facilities, and there is even weak

[15]While patient satisfaction is subjective, several studies endorse its validity as an instrument for measuring quality of healthcare (Davies and Ware, 1988; Andaleeb, 2001; Johansson et al., 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 et al., 2004), making it a good measure of the overall quality of healthcare delivery.

evidence that it rose faster in the PBF units (see Appendix B for the full analysis).

Existing literature raises several other concerns about the potential pitfalls of introducing a PBF scheme:

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. Firstly, the PBF scheme only adds a marginal amount of money to the budget. Most of the facilities' income is derived from fixed government subsidies and fee-for-service user payments. Any doctoring up of data would require compensating the parallel increase in user fees, which makes this type of fraudulent reporting as unlikely as it is expensive. Secondly, UCMB datasets have long been used at the national level as an advocacy device, requiring high levels of reliability, as well as an internal means of performance comparison across facilities, inducing all facilities to minimize unintended under-reporting regardless of the financing mechanism at play. Finally, 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 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 services with negative marginal value (Wynia, 2009). This however seems unlikely in a developing-country context where the supply of healthcare services often falls short of demand. Indeed, 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 (see Table A.4 in Appendix A).

Furthermore, Frey and Jegen (2001) and Wharam et al. (2009) are concerned that the bureaucratization of healthcare delivery, which is to some extent necessary to implement a PBF scheme, may crowd out intrinsic motivation, inducing a decline in the professionalism and morale of medical staff. We do not have any quantitative data on staff motivation, but in-depth interviews we conducted with managerial as well as rank-and-file staff in the PBF facilities suggest quite the opposite—the additional salary bonuses which became possible thanks to PBF made staff feel appreciated for their efforts.

Finally, Oxman and Fretheim (2008) warn against the danger of widening the already existing gap between poorly- and well-performing facilities. This could become an issue in cases where PBF is used to reallocate an existing pot of funds in what could be described as a zero-sum game. In our case, however, the performance bonuses come from an external source, precluding the necessity to decrease funding of under-performing facilities in order to top up that of the more successful ones. Any improvements should therefore be Pareto-efficient in our case.[16] It is nevertheless impossible to verify this in our present study given the small number of treated facilities. A larger trial – if not an implementation across the board – of the scheme would be needed in order to assert that no facility is worse off as a result of PBF.

7 Conclusion

Using a panel of output and expenditure data from small private not-for-profit healthcare facilities in Uganda, we estimate the contribution of performance-based financing towards achieving greater allocative efficiency. We rely on a dynamic panel estimation, verifying its robustness through more traditional two-stage DEA and SFA procedures to find that healthcare providers respond strongly to targets by

[16]This highlights a trade-off between Pareto-efficiency and financial sustainability of many PBF schemes—a topic that is very policy-relevant, but beyond the scope of this paper.

increasing output through improved allocative efficiency. In our case, output rose by over 25%. The result, though lower than those from more traditional static models which are likely to be biased upwards, is economically significant, and the increased efficiency does not come at the expense of quality. This shows the potential of PBF in improving the performance of underfunded healthcare systems in developing countries.

References

Andaleeb, S. S. (2001). Service Quality Perceptions and Patient Satisfaction: a Study of Hospitals in a Developing Country. *Social Science & Medicine 52*(9), 1359–70.

Arellano, M. (1993). On the testing of correlated effects with panel data. *Journal of Econometrics 59*(1-2), 87–97.

Banerjee, A. V., R. Glennerster, and E. Duflo (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., P. J. Gertler, A. Binagwaho, A. L. B. Soucat, J. Sturdy, and C. M. J. Vermeersch (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*(9775), 1421–1428.

Battese, G. E. and T. J. Coelli (1995). A Model for Technical Inefficiency Effects in a Stochastic Frontier Production Function for Panel Data. *Empirical Economics 20*, 325–332.

Björkman, M. and J. Svensson (2009). Power to the People: Evidence from a Randomized Field Experiment on Community-Based Monitoring in Uganda. *The Quarterly Journal of Economics 124*(2), 735 –769.

Bloom, D. E. and D. Canning (2000). The Health and Wealth of Nations. *Science 287*(5456), 1207–1209.

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

Brenzel, L. (2009). *Taking Stock: World Bank Experience with Results-Based Financing for Health.* Washington, D.C.: World Bank.

Centers for Medicare and Medicade Services (2011). CAHPS Hospital Survey (HCAPS): Quality Insurance Guidelines Version 6.0. Accessed 2 April 2016. `http://www.hcahpsonline.org`.

Choi, I. (2001, apr). Unit root tests for panel data. *Journal of International Money and Finance 20*(2), 249–272.

Davies, A. R. and J. E. Ware (1988). Involving Consumers in Quality of Care Assessment. *Health Affairs 7*(1), 33–48.

Eichler, R. (2006). Can "Pay for Performance" Increase Utilization by the Poor and Improve the Quality of Health Services. Technical report, Center for Global Development.

Eijkenaar, F., M. Emmert, M. Scheppach, and O. Sch??ffski (2013). Effects of pay for performance in health care: A systematic review of systematic reviews. *Health Policy 110*(2-3), 115–130.

Eldridge, C. and N. Palmer (2009). Performance-Based Payment: Some Reflections on the Discourse, Evidence and Unanswered Questions. *Health Policy and Planning 24*(3), 160–6.

Fretheim, A., S. Witter, A. K. Lindahl, and I. T. Olsen (2012). Performance-Based Financing in Low- and Middle-Income Countries: Still More Questions than Answers. *Bulletin of the World Health Organization 90*(8), 559–559A.

Frey, B. S. and R. Jegen (2001). Motivation Crowding Theory. *Journal of Economic Surveys 15*(5), 589–611.

Governemt of Uganda - Ministry of Health (2010). Health Sector Strategic Plan III: 2010/11-2014/15.

Hecht, R., A. Batson, and L. Brenzel (2004). Making Health Care Accountable. *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*(10), 1107–1128.

Honda, A. (2013). 10 Best Resources on ... Pay for Performance in Low- and Middle-Income Countries. *Health Policy and Planning 28*(5), 454–7.

Hsiao, C. (2003). *Analysis of Panel Data.* Cambridge: Cambridge University Press.

Hunt, J. (2010). Bribery in Health Care in Uganda. *Journal of Health Economics 29*(5), 699–707.

Im, K. S., M. H. Pesaran, and Y. Shin (2003, jul). Testing for unit roots in heterogeneous panels. *Journal of Econometrics 115*(1), 53–74.

Ireland, M., E. Paul, and B. Dujardin (2011). Can Performance-Based Financing be Used to Reform Health Systems in Developing Countries? *Bulletin of the World Health Organization 89*(9), 695–8.

Johansson, P., M. Oléni, and B. Fridlund (2002). Patient Satisfaction with Nursing Care in the Context of Health Care: a Literature Study. *Scandinavian Journal of Caring Sciences 16*(4), 337–44.

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), 444–59.

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., D. A. Collier, and S. Meyer Goldstein (2004). The Role of Clinical and Process Quality in Achieving Patient Satisfaction in Hospitals. *Decision Sciences 35*(3), 349–369.

Maynard, A. (2012). The Powers and Pitfalls of Payment for Performance. *Health Economics 12*, 3–12.

Meessen, B., J.-P. I. Kashala, and L. Musango (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., L. Musango, J.-P. I. Kashala, and J. Lemlin (2006). Reviewing Institutions of Rural Health Centres: the Performance Initiative in Butare, Rwanda. *Tropical medicine & international health 11*(8), 1303–1317.

Morgan, L. (2010). Some Days are Better than Others: Lessons Learned from Uganda's First Results-Based Financing Pilot. Technical report, World Bank.

Nijman, T. and M. Verbeek (1992). Nonresponse in panel data: The impact on estimates of a life cycle consumption function. *Journal of Applied Econometrics 7*(3), 243–257.

Oxman, A. D. and A. Fretheim (2008). An Overview of Research on the Effects of Results-Based Financing. Technical Report 16, Norwegian Knowledge Centre for Health Services.

Peabody, J. W., F. Tozija, J. a. Muñoz, R. J. Nordyke, and J. Luck (2004). Using Vignettes to Compare the Quality of Clinical Care Variation in Economically Divergent Countries. *Health Services Research 39*(6), 1951–70.

Perrot, J., E. de Roodenbeke, L. Musango, and G. Fritsche (2010). Performance Incentives for Health Care Providers. Technical report, World Health Organization.

Roodman, D. (2009). How to Do xtabond2: An Introduction to Difference and System GMM in Stata. *The Stata Journal 9*(1), 86–136.

Rusa, L., M. Schneidman, G. Fritsche, and L. Musango (2009). Rwanda: Performance-Based Financing in the Public Sector. In R. Eichler and R. Levine (Eds.), *Performance incentives for global health: potential and pitfalls*. Washington, D.C.: Center for Global Development.

Ryan, A. (2009). Hospital-Based Pay-for-Performance in the United States. *Health Economics 18*, 1109–1113.

Sahn, D. and D. Stifel (2003). Exploring Alternative Measures of Welfare in the Absence of Expenditure Data. *Review of Income and Wealth 49*(4), 463–489.

Scheffler, R. M. (2010). Pay For Performance (P4P) Programs in Health Services: What is the Evidence? Technical report, World Health Organization.

Scott, A. and W. Coote (2010). Do Regional Primary-Care Organisations Influence Primary-Care Performance? A Dynamic Panel Estimation. *Health Economics 19*, 716–729.

Sekabaraga, C., F. Diop, and A. Soucat (2011). Can Innovative Health Financing Policies Increase Access to MDG-Related Services? Evidence from Rwanda. *Health Policy and Planning 26 Suppl 2*, ii52–ii62.

Simar, L. and P. W. Wilson (2007). Estimation and Inference in Two-Stage, Semi-Parametric Models of Production Processes. *Journal of Econometrics 136*(1), 31–64.

Soeters, R., C. Habineza, and P. Bob (2006). Performance-Based Financing and Changing the District Health System: Experience from Rwanda. *Bulletin of the World Health Organization 84*(11), 884–889.

Soeters, R., P. B. Peerenboom, P. Mushagalusa, and C. Kimanuka (2011). Performance-Based Financing Experiment Improved Health Care in the Democratic Republic of Congo. *Health Affairs 30*(8), 1518–1527.

Ware, J. E. J., M. Kosinski, and S. D. Keller (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. *The Quarterly Journal of Economics* (August), 1265–1306.

Wharam, J. F., M. K. Paasche-Orlow, N. J. Farber, C. Sinsky, L. Rucker, K. J. Rask, M. K. Figaro, C. Braddock, M. J. Barry, and D. P. Sulmasy (2009). High Quality Care and Ethical Pay-for-Performance: a Society of General Internal Medicine Policy Analysis. *Journal of General Internal Medicine 24*(7), 854–9.

Wolszczak-Derlacz, J. and A. Parteka (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 Bank (2016). World Development Indicators. Accessed 2 June 2016. `http://data.worldbank.org/products/wdi`.

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), 884–7.

Appendices

A Tables

Table A.1: PBF effects on output – full regression specifications

	(1) OLS	(2) FE	(3) FE	(4) FGLS	(5) sysGMM
Dep. var.	$\ln SUO$	$\ln SUO$	$\ln SUO$	$\ln SUO$	$\ln SUO$
PBF	0.734***	0.737***	0.492***	0.509***	0.266**
	(0.129)	(0.123)	(0.131)	(0.0945)	(0.130)
$\ln SUO_{it-1}$					0.377***
					(0.053)
$\ln expenditures$	0.655***	0.402**	0.488**	0.500***	0.594***
	(0.149)	(0.158)	(0.195)	(0.073)	(0.111)
$\ln staff$	0.237	-0.334	-0.336*	-0.351***	0.077
	(0.202)	(0.210)	(0.184)	(0.108)	(0.143)
$\ln capacity$	0.011	-0.032	-0.074	-0.085	0.058
	(0.154)	(0.172)	(0.170)	(0.085)	(0.115)
$(\ln expenditures)^2$	-0.035	-0.047*	-0.054*	-0.055***	-0.046***
	(0.025)	(0.026)	(0.029)	(0.012)	(0.017)
$(\ln staff)^2$	0.097**	0.016	0.044	0.060*	-0.005
	(0.047)	(0.050)	(0.051)	(0.035)	(0.049)
$(\ln capacity)^2$	0.029	0.031	-0.001	0.018	-0.013
	(0.032)	(0.034)	(0.031)	(0.016)	(0.021)
$\ln expenditures \times \ln staff$	-0.046	0.110*	0.065	0.060*	0.013
	(0.056)	(0.062)	(0.059)	(0.036)	(0.053)
$\ln expenditures \times \ln capacity$	0.031	0.022	0.036	0.026	0.008
	(0.043)	(0.035)	(0.035)	(0.021)	(0.028)
$\ln staff \times \ln capacity$	-0.051	0.018	0.041	0.015	0.029
	(0.053)	(0.054)	(0.056)	(0.031)	(0.044)
Time FE	no	no	yes	yes	yes
Diocese FE	no	no	no	no	yes
SE clustered by facility	yes	yes	yes	no	no
Mean dep. var.	9.321	9.321	9.321	9.321	9.372
Std. dev. dep. var.	(0.699)	(0.699)	(0.699)	(0.699)	(0.684)
Obs.	1903	1903	1903	1896	1572
R^2	0.594	0.276	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. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.2: PBF effects on output – full frontier analysis specifications

	(2) SFA	
	1^{st} stage	2^{nd} stage
Dep. var.	$\ln SUO$	u
PBF		-0.476**
		(0.241)
$\ln expenditures$	0.670***	
	(0.094)	
$\ln staff$	0.094	
	(0.082)	
$\ln beds$	0.170	
	(0.125)	
$(\ln expenditures)^2$	-0.041**	
	(0.016)	
$(\ln staff)^2$	0.110**	
	(0.046)	
$(\ln beds)^2$	-0.013	
	(0.015)	
$\ln expenditures \times \ln staff$	-0.053	
	(0.044)	
$\ln expenditures \times \ln beds$	0.057**	
	(0.029)	
$\ln staff \times \ln beds$	-0.045	
	(0.045)	
Time FE	no	yes
Diocese FE	no	yes
Mean dep. var.	9.321	0.817
Std. dev. dep. var.	(0.699)	(0.228)
Obs.	1903	
Wald $\chi^2(9)$	2877.92	

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

Table A.3: Nijman-Verbeek tests for non-random censoring

	(1) FGLS	(2) FGLS	(3) sysGMM	(4) sysGMM
Dep. var.	$\ln SUO$	$\ln SUO$	$\ln SUO$	$\ln SUO$
PBF	0.505***	0.421***	0.268**	0.264**
	(0.092)	(0.123)	(0.131)	(0.130)
In previous FY	0.034**		-0.007	
	(0.017)		(0.072)	
Number of FY		0.008*		-0.002
		(0.004)		(0.005)
Translog inputs	yes	yes	yes	yes
SE clustered by facility	yes	no	no	no
SE clustered by diocese	no	yes	yes	yes
Mean dep. var.	9.337	9.321	9.372	9.372
Std. dev. dep. var.	(0.691)	(0.699)	(0.684)	(0.684)
N	1817	1896	1572	1572
Wald χ^2	11118***	3452***		
Wald χ^2 df	229	41		
Instruments			53	53
$AR(2)$ z			0.75	0.74
Sargan $\chi^2(8)$			7.35	7.66
Hansen $\chi^2(8)$			7.78	8.04
Difference in Hansen $\chi^2(1)$			1.91	2.29

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

Table A.4: PBF effects on individual output indicators

	(1) sysGMM	(2) sysGMM	(3) sysGMM	(4) sysGMM
Dep. var.	$\ln SUO$	$\ln outpatients$	$\ln inpatients$	$\ln immunizations$
PBF	0.266**	0.276***	0.245	0.320***
	(0.130)	(0.100)	(0.197)	(0.120)
Dep. var. $_{it-1}$	0.377***	0.405***	0.475***	0.531***
	(0.053)	(0.042)	(0.074)	(0.060)
Dep. var. $_{it-2}$			0.103**	
			(0.043)	
Translog inputs	yes	yes	yes	yes
Time FE	yes	yes	yes	yes
Diocese FE	yes	yes	yes	yes
Mean dep. var.	9.319	8.705	6.513	8.333
Std. dev. dep. var.	(0.705)	(0.689)	(1.078)	(0.869)
Obs.	1572	1572	1380	1572
Instruments	52	52	51	52
$AR(2)$ z	0.76	0.72	0.19	-0.22
Sargan χ^2 (df)	7.30 (8)	5.15 (8)	4.02 (6)	11.43 (8)
Hansen χ^2 (df)	7.64 (8)	4.01 (8)	3.55 (6)	9.55 (8)
Difference in Hansen $\chi^2(1)$	1.66	0.35	0.02	0.06

Notes: SE in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Inputs: *expenditures*, *staff*, *capacity*. An additional lag of the dependent variable ($\ln inpatients_{it-1}$) is needed in model (3) to satisfy identification requirements.

B Quality

To complement the input and output panel data, we also collected information about the quality of healthcare delivery in a sub-sample of the facilities. The data was only collected after the introduction of the PBF scheme, and it is therefore not suited for analyzing the impact of PBF on healthcare quality. It can however help us address some of the reservations one might have about an introduction of a scheme that only incentivizes quantity, not quality. The relevant analysis is presented in this appendix.

B.1 Data

While our sample contains facilities belonging to the same size class, we further 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, as measured in the last year prior to the introduction of the scheme.[17] 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 and Medicade Services, 2011) to the Ugandan environment and administered it in two waves in 2012 and in 2014. While patient satisfaction is subjective, several studies endorse its validity as an instrument for measuring quality of healthcare (Davies and Ware, 1988; Andaleeb, 2001; Johansson et al., 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 et al., 2004), making it a good measure of the overall quality of healthcare delivery. To address concerns that 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, which – excluding respondents who had not visited the catholic facilities in the three years prior to the interview – resulted in 440 incidentally truncated interviews. From answers to a set of questions regarding various aspects of 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 ($health$), an asset index ($assets$) approximating wealth,[18] 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*.[19] As can be seen in Table B.1, the sample was well balanced across all the confounding characteristics

[17] The matching characteristics include the facilities' income and expenditures, catchment population, number of staff, bed capacity, number of carried out diagnostic procedures and minor surgical operations, average length of stay of admitted patients, number of fatalities, and the availability of mental counseling.

[18] The physical health index is based on the SF-12 Health Survey as proposed by Ware et al. (1996). The asset index is obtained by principal factor analysis following Sahn and Stifel (2003)

[19] We use the mean perceived quality of the main public facility rather than the quality perceived by each respondent, de facto controlling for the relative reputation of the public competitor.

Table B.1: Survey sample balance

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

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

(except for *govt*) in 2012, but completely unbalanced in 2014. We will address these issues below.

B.2 Empirical strategy

It is evident from the previous section that we only have data on perceived quality of 12 selected facilities from two survey waves which took place several years after the introduction of PBF in 6 of the units. Lacking a baseline survey, we clearly 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 the trends in the PBF and non-PBF facilities after the new financing system had been introduced and after the management and staff of the facilities had presumably gotten accustomed to it.

We already outlined above our choice to investigate quality through a HCAHPS-type survey rather than through observable structural and process measures. 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:

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

where FU is a dummy equal to one for all observations from the 2014 follow-up, $\boldsymbol{X}$ is a vector of production factor proxying the size of the private facility (*expenditures*, $staff$ and *beds*), $\boldsymbol{Z}$ is a vector of respondent characteristics comprising *health* (a physical health index), *assets* (an asset index approximating wealth), *primary* (a dummy equal to one when 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, η are location type fixed effects indicating whether the interview took place at the private facility, the public facility, or the village in between the two, and ε is a stochastic error term.

Finally, to account for possible biases resulting from the imbalance in the 2014 sample, we further estimate a kernel propensity score DID model using $\boldsymbol{Z}$ and *govt* to compute the propensity scores.

B.3 Results

In this section, we compare the trends in the perceived quality of healthcare provision in PBF and non-PBF facilities following the introduction of the new financing system in the former. Before proceeding to a full DID model specification, 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 B.2, 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.

Table B.2: PBF effects on trend in perceived quality of healthcare – regressions

	(1) Heckman		(2) OLS	(3) OLS	(4) OLS
Dep. var.	*private*	*quality*	*quality*	*quality*	*quality*
$PBF \times FU$		0.037	0.038**	0.035**	0.038**
		(0.022)	(0.016)	(0.016)	(0.015)
PBF		-0.045***	-0.047	-0.040	-0.035
		(0.017)	(0.026)	(0.026)	(0.026)
FU		0.031*	0.030**	0.028*	0.031*
		(0.016)	(0.011)	(0.013)	(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 λ		-0.027			
		(0.050)			
R^2			0.069	0.082	0.085

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

Starting with a naïve specification which only includes the DID terms and size controls (Table B.2, column 2), we gradually include respondent characteristics (Table B.2, column 3), the mean level of perceived quality of the nearest similarly-sized public facility (Table B.2, column 4), and controls for the location and timing of the interview (Table B.2, column 5). To address the imbalance in the 2014 sample, we also estimate a kernel propensity score DID model using respondent characteristics and public competitor quality to compute the propensity scores and controlling for the remaining covariates from the full DID OLS model (Table B.3).

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 B.2). There is also some evidence that the increase was much more pronounced in the case of PBF facilities (see coefficients on $PBF \times FU$ in Table B.2), though the statistical significance of this relationship falls below the 10% level when we account for self-selection into attending a private facility

Table B.3: PBF effects on trend in perceived quality of healthcare – PSM

	2012			2014			DID
	Control	PBF	Diff	Control	PBF	Diff	
quality	0.851	0.889	-0.038	0.928	0.920	-0.008	0.030
Std. error			(0.030)			(0.023)	(0.019)

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

by means of a PSM DID estimation (see column DID in Table B.3). It is nevertheless clear that PBF did not lead to a decline in patient satisfaction between the two survey waves.