The Medium-Run and Scale-Up Effects of Performance-Based Financing: An Extension of Rwanda's 2006 Trial Using Secondary Data

Diana Ngo and Sebastian Bauhoff

Abstract

Objective: To extend the initial evaluation of Rwanda's performance-based financing program in order to identify medium-run and scale-up effects of incentives and unconditional financing relative to one another and a new "business as usual" counterfactual.

Methods: We use secondary data from the Demographic and Health Surveys from Rwanda and its East African neighbors from 2001 to 2010. We identify a relevant set of controls using neighboring regions that are similar to Rwanda based on pre-intervention trends in covariates and outcomes. We then use difference-in-differences regressions to measure the program's impacts on key maternal health service indicators.

Findings: In the first two years and relative to no intervention, performance-based and unconditional financing raised institutional delivery rates by 21 and 13 percentage points, respectively, and performance-based financing increased completion of four antenatal visits by 6 percentage points. After two years, relative to no intervention and in addition to the initial short-run impacts, performance-based incentives resulted in further improvements of 11 percentage points for institutional deliveries and 10 percentage points for completion of four antenatal visits. Program scale-up was effective, with no differences between intervention arms after all areas received performance-based incentives. We find few effects on antenatal tetanus prophylaxis.

Conclusions: Rwanda's performance-based incentives were effective for some indicators, but unconditional financing also induced improvements. The incentive effects persisted in the medium-run and as the program was scaled-up. Additionally, the analysis demonstrates how observational research methods and secondary data can generate new insights on existing evaluations.

Keywords: Performance-based financing, health care finance, health care providers, difference-in-difference analysis, secondary data

Center for Global Development
www.cgdev.org

Working Paper 497
November 2018

The Medium-Run and Scale-Up Effects of Performance-Based Financing: An Extension of Rwanda's 2006 Trial Using Secondary Data

Diana Ngo
Department of Economics, Occidental College

Sebastian Bauhoff
Department of Global Health and Population, Harvard TH Chan School of Public Health
Center for Global Development

The Center for Global Development is grateful for contributions from the Australian Department of Foreign Affairs and Trade (grant agreement 72518) and Johnson & Johnson (C2017001214) in support of this work. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Diana Ngo and Sebastian Bauhoff, 2018. "The Medium-Run and Scale-Up Effects of Performance-Based Financing: An Extension of Rwanda's 2006 Trial Using Secondary Data." CGD Working Paper 497. Washington, DC: Center for Global Development. https://www.cgdev.org/publication/medium-run-and-scale-effects-performance-based-financing-extension-rwandas-2006-trial

Center for Global Development
2055 L Street NW
Washington, DC 20036

202.416.4000
(f) 202.416.4050

www.cgdev.org

Contents

Introduction

Rigorous impact evaluations are increasingly used to learn about health policies in low- and middle-income countries (LMIC), including alternative ways to pay for health care services. These efforts include national governments and donors, such as the World Bank's 420 million USD Health Results Innovation Trust Fund (HRITF), which supports dozens of randomized-controlled trials (RCTs) and field experiments of performance-based financing (PBF) programs (1–3).

PBF involves a package of interventions centered around providing monetary incentives for achieving targeted indicators for health service utilization, quality, and/or outcomes. Operationally, PBF programs also involve coaching, initial infrastructure upgrades, and intensive monitoring and performance verification. To date, PBF programs have generated positive results for some indicators and in some countries but no improvements for others (4), suggesting that both contextual factors and program design are important. However, despite the substantial investment of donors into piloting, evaluating and diffusing PBF, as well as interest among LMIC governments in alternatives to input-based financing, there continue to be evidence gaps, particularly concerning medium-run and scale-up changes as well as the merits of PBF relative to other financing approaches.

There are several reasons for these gaps. First, many PBF programs have only recently been introduced, and there is a time lag in evidence generation and dissemination. The existing evaluations tend to focus on short-run outcomes, with exposure times ranging between 18 and 24 months. For example, evaluations of Rwanda's seminal program have all focused on the initial two year period (5–8). Relatedly, there have not been follow-up evaluations to determine how PBF performs when it is scaled-up. These medium-run and scale-up impacts are important to identify, particularly if there are initial implementation challenges associated with program administration, path-dependency in transitioning from one type of financing to another, or changes in oversight after the evaluation period.

Second, the evaluations generally test a small number of interventions, limiting their ability to compare different financing schemes or to separate out the effects of different program components. For example, Rwanda's influential 2006 trial was designed to isolate the incentive effect of PBF, comparing the effect of two alternative treatments, performance-based incentives versus additional unconditional financing (5). However, it lacked a "business as usual" scenario and therefore did not identify the effect of either financing modality relative to no intervention. Several ongoing evaluations are similarly designed to test competing interventions against one another rather than against the status-quo approach, while others include a no-intervention control group but only test a single intervention (9).

In this paper, we answer additional substantive questions central to the current PBF debates and illustrate how observational research methods and secondary data can be used to extend existing policy evaluations. Specifically, we extend Rwanda's 2006 PBF trial in two ways, by applying difference-in-difference methods to data from Rwanda and neighboring countries. First, we identify the effects of performance-based incentives and unconditional financing

relative to no intervention by using the neighboring countries to generate a counterfactual "synthetic" Rwanda that received no intervention. The RCT only analyzed the alternative treatments against each other. Second, we measure the medium-run (52 months) and scale-up effects of the program using the new control group. The RCT's data ended as the program was scaled-up, limiting the exposure time to 23 months, and additional evaluation was not possible using data from Rwanda only, since the national expansion left no remaining control group from within Rwanda. Our findings indicate that, for a subset of incentivized indicators, both incentives and unconditional financing had positive impacts in the short-run, incentives had additional positive impacts in the medium-run, and program scale-up was effective.

Background on Rwanda's Performance-Based Financing Trial

Rwanda implemented various nationwide health reforms in the mid-2000s, including a large multi-sector program of performance-based contracting with local governments (*imihigo*) and a health insurance expansion (5,10). In conjunction with these reforms, Rwanda randomly phased in a PBF program (5). In addition to the traditional input-based budgets, facilities under the PBF program received varying unit payments for a set of incentivized service indicators in the domains of maternal health, child health, family planning and HIV/AIDS. These bonus payments were adjusted according to a measure of overall facility quality, constructed using indicators for the availability of key inputs, management processes, and adherence to clinical protocols. Districts in the trial's arm 1 received performance bonuses beginning in June 2006, while arm 2 districts received additional unconditional financing equal to the average facility bonus. This design aimed to isolate the effect of incentives from the effect of additional resources. The incentives were scaled up to arm 2 districts in April 2008.

Several studies have evaluated the program's short-run impact. Basinga et al. collect and analyze primary data and report, among other findings, that rates of institutional deliveries and antenatal tetanus prophylaxis increased under PBF relative to unconditional financing; they find no differential impact on completion of four antenatal visits (5). Using data from the Rwandan Demographic and Health Surveys (DHS), Sherry et al. find a similar positive effect on institutional deliveries but no effect on the other two indicators (11). Other studies examine HIV testing and counseling (6), heterogeneous effects (7), and health systems changes (8).

Data

To measure the program's impact, we compare outcomes in Rwanda to those in similar neighboring countries. We use DHS data, which are standardized household surveys used for monitoring and evaluating indicators related to population, health, and nutrition. For outcomes, we focus on maternal service outcomes because 1) these services were a primary target of the Rwandan PBF and showed the largest improvements in previous studies, and 2) the DHS collects retrospective data on all births five years prior to the survey, allowing us to

construct a pseudo-panel of annual data and perform analyses that consider trends in pre-intervention outcomes. Specifically, we analyze key incentivized indicators for maternal services that are observed annually in the DHS: institutional deliveries, antenatal tetanus prophylaxis, and completion of four antenatal visits. We use a set of mother's characteristics as covariates in our analyses. These include demographics, health insurance status, wealth indicators, and measures of community health worker access and are very similar to those considered in the initial RCT (5).

For Rwanda, we include the DHS from 2005, 2007, and 2010, allowing us to create a pseudo-panel from 2001-2010, covering the pre-intervention period until mid-2010, when Rwanda implemented additional subnational PBF reforms (12). To identify potential control countries, we reviewed published literature on health policies in all neighboring countries to identify and exclude areas that were affected by similar policies during the relevant period. Following this review, we excluded Zambia as a potential control country because Zambia also implemented a PBF program in this period. While other neighboring countries were also making efforts to improve their health systems, other concurrent policies outside Rwanda were subnational or likely to improve outcomes in the control group, which would attenuate any positive estimates of Rwanda's PBF impact. We use all remaining neighboring countries with sufficiently frequent DHS data, including Ethiopia (2005, 2011), Kenya (2003, 2008, 2014), Malawi (2004, 2010), Uganda (2006, 2011), Tanzania (2004, 2010, 2015), and Zimbabwe (2005, 2010). See appendix A for additional details on survey inclusion, coding consistency across surveys, and other programs.

Methods

To identify causal estimates, we combine difference-in-differences regressions with a pre-processing step. The pre-processing step identifies control areas that are similar to the treatment group according to covariates and pre-treatment outcomes, attempting to improve balance for unobserved confounders (13). The difference-in-differences analyses then control for time-invariant residual biases (14). We interpret our findings in the context of the broader reforms, by controlling for increasing insurance status and using the terms "incentives" and "unconditional financing" to also include *imihigo*, the multi-sector performance-based contracting reforms in Rwanda.

Pre-processing

For our pre-processing, we use the synthetic control method (SCM) to identify a relevant set of control areas from neighboring countries. SCM is a data-driven algorithm that generates a "synthetic" control group using a weighted linear combination of the potential control areas. The weights are chosen to minimize the distance between the treatment group and synthetic control according to pre-intervention trends in outcomes and key covariates (15).

SCM has been used to study a range of topics including the impacts of state-level anti-smoking legislation (15), trade liberalization (16), and hospital PBF schemes (17). It has been primarily been used in comparative case studies, using group-level data to construct a time-

series for a single synthetic control and compare changes in outcomes for the treated and synthetic control units (15). It has also been used to generate weights that are then applied in regression analyses (18,19). We follow this second approach to exploit the individual level variation in our data, using the weights to enforce balance in pre-intervention outcome trends and observed covariates, strengthening the likelihood that, in the absence of the policy, treatment and control groups would have experienced similar changes in the outcomes over time. Our approach is similar to multivariate matching combined with difference-in-differences methods (20); we use the synthetic control method because it performs well in generating balance (13).

Specifically, we create the synthetic control group in two steps. First, we generate aggregate annual statistics for the treatment and control areas. For the control countries, we use the smallest geographical areas for which the respective DHS are representative. For Rwanda, we pool data from arm 1 and 2 districts, aggregating to a single treatment area (15). The two sets of districts were designed to be comparable under the initial randomization. In addition, having a single control group allows us to test the relative effects of each scenario in a single regression (see below). Following Basinga et al. (5), we account for changes in administrative boundaries that affected the treatment assignment; we also drop Rwandan districts that received PBF prior to 2006.

Second, we create a "lagged" synthetic control by applying SCM to all pre-intervention lagged outcomes from 2001 through 2005 (17). After including all lagged outcomes, additional covariates have no predictive power (21). For robustness, we generate another "covariate" synthetic control using pre-intervention outcomes and covariate predictors (21). Specifically, we split the pre-intervention time period into two (2001-2003 and 2004-2005) and use average outcomes and covariates within each period as predictors. The covariates are demographics and indicators for health insurance and the availability of community health workers.

Under the rationale that PBF affects the entire health system rather than individual indicators, we generate a single control group based on an aggregate system measure. Specifically, we create the synthetic control based on the maternal service rate, defined as the simple average of the rates of institutional deliveries, four antenatal visits, and tetanus prophylaxis. As predictors, we include the three outcomes separately. In the appendix, we also generate synthetic control groups using each outcome separately and find similar results.

Difference-in-Differences Regressions

We transform the weighted combination of synthetic control regions to individual-level weights by normalizing the sampling weights. Within each year, the sum of weights for all synthetic control observations equals one. Similarly, within each year, the weights for Rwanda's arm 1 and arm 2 areas sum to one, separately.

We then run two sets of difference-in-differences regressions using individual birth data. The first set identifies the impact of the performance incentives relative to unconditional financing examined by earlier studies. We first replicate Basinga et al.'s design by using data

from 2005 and 2007, their baseline and endline years (5). We then use data from 2001 through 2008, ending before the scale-up in 2008. We report results from the two synthetic control approaches and results from a specification using data from Rwanda only. The specification is:

$$Outcome_{irt} = \tau\ Arm\ 1_r\ x\ post\ 2006_t + \delta\ Rwanda_r\ x\ post\ 2006_t + \alpha_r + \gamma_t + X_{irt}\delta + \varepsilon_{irt} \quad (1)$$

where $Arm\ 1_r$ is an indicator for Rwanda arm 1 districts, $Rwanda_r$ is an indicator for Rwanda arm 1 or arm 2 districts, $post\ 2006_t$ is an indicator for being born after the initial roll-out, α_r are regional fixed effects (i.e., districts in Rwanda and regions in other countries), γ_t are birth month-year fixed effects, and X_{irt} are individual covariates. Standard errors are clustered at the regional level. τ represents the causal effect of incentives relative to unconditional financing under the standard parallel trends assumption. We drop $Rwanda_r\ x\ post\ 2006_t$ in the Rwanda-only specification.

The second set of regressions expands the data series to 2010, which allows us to newly examine the medium-term effect of incentives and the scale-up in arm 2 areas:

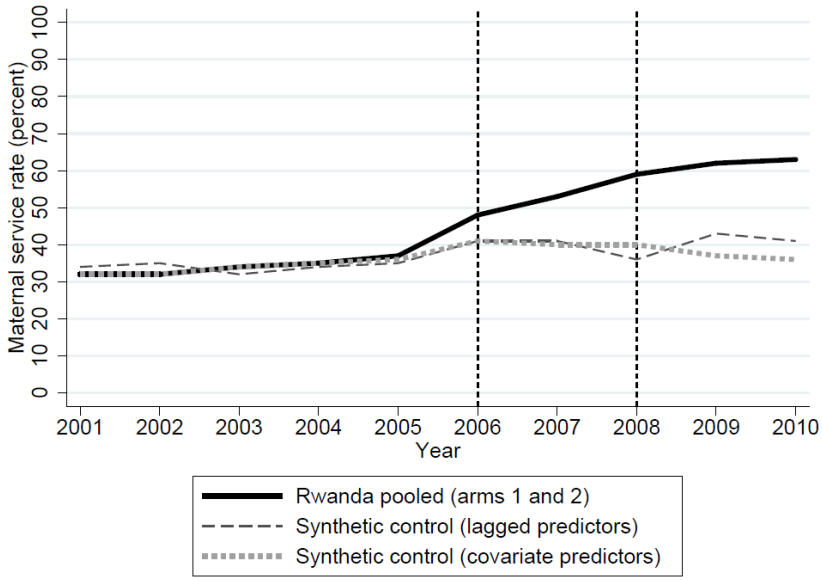
$$Outcome_{irt} = \beta_1\ Arm\ 1_r\ x\ post\ 2006_t + \beta_2\ Arm\ 2_r\ x\ post\ 2006_t + \beta_3\ Arm\ 1_r\ x\ post\ 2008_t + \beta_4\ Arm\ 2_r\ x\ post\ 2008_t + \alpha_r + \gamma_t + X_{irt}\delta + \varepsilon_{irt} \quad (2)$$

where $Arm\ 2_r$ is an indicator for Rwanda arm 2 districts, $post\ 2008_t$ is an indicator for being born after the scale-up phase, and all other descriptors are as above. Here, β_1 represents the initial effect of incentives relative to no program, β_2 represents the initial effect of unconditional financing relative to no program, β_3 represents the additional medium-run effect of incentives beyond the initial effects relative to no program, and β_4 represents the scale-up effect of transitioning from unconditional financing to incentives. We also test combinations of coefficients that have substantive interpretations.

Results

Figure 1 plots the maternal service rates by year for Rwanda and the two synthetic controls. Both controls are similar to the treatment areas in pre-intervention maternal service rates, suggesting that SCM works well in generating a counterfactual for Rwanda. Appendix figures B.1, B.2, and B.3 show the plots disaggregated by outcome and treatment arm, and appendix table B.1 shows the weights for each control region.

Figure 1: Maternal service rates by year



Note: The maternal service rate is defined as the simple average of the rates of institutional deliveries, four antenatal visits, and antenatal tetanus prophylaxis.

Table 1: Predictor balance, regional level averages

Lagged predictors					Covariate predictors				
	Year	Rwanda	Synth	Unweighted controls		Year	Rwanda	Synth	Unweighted controls
Institutional deliveries	2001	0.22	0.20	0.46	Institutional deliveries	2001-03	0.23	0.22	0.46
	2002	0.22	0.18	0.46		2004-05	0.30	0.24	0.49
	2003	0.25	0.20	0.47	Four antenatal visits	2001-03	0.12	0.28	0.55
	2004	0.28	0.20	0.47		2004-05	0.13	0.30	0.49
	2005	0.32	0.23	0.51	Antenatal tetanus shot	2001-03	0.63	0.48	0.76
Four antenatal visits	2001	0.11	0.23	0.59		2004-05	0.64	0.53	0.73
	2002	0.11	0.24	0.55	Birth order	2001-03	3.99	4.02	3.54
	2003	0.13	0.20	0.52		2004-05	3.88	3.94	3.49
	2004	0.13	0.24	0.49	Urban	2001-03	0.08	0.07	0.22
	2005	0.13	0.26	0.49		2004-05	0.08	0.08	0.22
Ante. tetanus shot	2001	0.62	0.57	0.77	Has comm. health worker	2001-03	0.58	0.71	0.79
	2002	0.62	0.62	0.76		2004-05	0.61	0.74	0.82
	2003	0.65	0.57	0.74	Primary education	2001-03	0.68	0.37	0.70
	2004	0.63	0.59	0.73		2004-05	0.70	0.39	0.73
	2005	0.65	0.56	0.73	Household size	2001-03	5.68	5.97	6.09
						2004-05	5.55	5.95	6.14
					Health insurance, household	2001-03	0.54	0.00	0.02
						2004-05	0.63	0.00	0.04
					Age†	2001-03	28.71	26.63	26.15
						2004-05	28.78	26.86	26.23
					Num. unsatisfied basic needs†	2001-03	2.50	2.57	2.25
						2004-05	2.34	2.55	2.20
					Num. durable assets†	2001-03	0.02	0.04	0.26
						2004-05	0.02	0.05	0.26

†Note: Age, total unsatisfied basic needs, and total durable assets are shown for conciseness for the covariate predictors. Instead, the synthetic control analysis includes indicators for age under 20, age over 35, indicators for one, two, three, and four unsatisfied basic needs, and indicators for ownership of televisions, landline telephones, cars/trucks, and refrigerators.

Table 1 presents the averages for all predictors included in the SCM at the regional level. Overall, the synthetic controls are more similar to Rwanda than the unweighted set of controls. Compared to Rwanda, both synthetic control regions have lower levels of institutional deliveries and tetanus prophylaxis and higher levels of four antenatal visits. While there are some treatment and control differences in predictor levels, the time trends are generally similar with the exception of household health insurance. We include all controls in the individual-level regressions. Appendix table B.2 shows the complete balance tables using the individual-level data.

Table 2 shows the results from the first regression. Overall, the results are comparable to those identified in the RCT, which are reproduced in column 2. Relative to unconditional financing, incentives increased institutional delivery rates by 8-9 percentage points, had no effect on completion of four antenatal visits, and increased tetanus prophylaxis by 4-5 percentage points (marginally significant in some specifications).

Table 2: Effects of incentives versus unconditional finance, comparison with RCT (percentage points)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
		2005, 2007			2001-2008		
	Basinga RCT	Rwanda only	Lagged synth	Covar synth	Rwanda only	Lagged synth	Covar synth
Institutional deliveries							
Arm 1 x post 2006 (incentives vs. uncond. finance)	8.1**	9.1**	8.7**	8.6**	8.1***	8.0***	7.9***
	[$p = 0.02$]	(3.7)	(3.5)	(3.6)	(2.6)	(2.5)	(2.6)
Observations	2,108	3,064	4,249	6,839	11,184	15,330	23,269
2005 mean (Rwanda arm 1)	35	32.2	32.2	32.2	32.2	32.2	32.2
Four antenatal visits							
Arm 1 x post 2006 (incentives vs. uncond. finance)	0.8	1.6	1.0	1.7	2.6	2.4	2.4
	[$p = 0.83$]	(4.5)	(4.4)	(4.5)	(3.3)	(3.3)	(3.3)
Observations	2,223	2,173	2,967	4,412	6,749	9,526	13,985
2005 mean (Rwanda arm 1)	18	13.4	13.4	13.4	13.4	13.4	13.4
Antenatal tetanus shot							
Arm 1 x post 2006 (incentives vs. uncond. finance)	5.1*	-2.1	-1.8	-2.4	4.3	4.7*	4.6*
	[$p = 0.06$]	(5.3)	(5.1)	(5.3)	(2.6)	(2.6)	(2.6)
Observations	2,856	2,156	2,940	4,383	6,698	9,457	13,887
2005 mean (Rwanda arm 1)	71	66.5	66.5	66.5	66.5	66.5	66.5

Note: * p < 0.1, ** p < 0.05, *** p < 0.01. Data from column (1) were independently collected by Basinga et al.; all other columns use DHS data. Coefficients are from difference-in-differences regressions, multiplied by 100 for exposition. All specifications include covariates, regional fixed effects, and birth month-year fixed effects. Standard errors are clustered at the regional level and are shown in parenthesis.

Table 3 shows the results from the second regression. With the exception of the initial effects for four antenatal visits, the results using the lagged and covariate synthetic controls are very similar in magnitude and significance; we cannot reject the equality of coefficients across models (appendix tables B.3 and B.4). Hereafter, we describe the lagged synthetic controls coefficients, which are generally smaller.

Table 3: Differential effects by arm and phase based on synthetic control method, 2001-2010 (percentage points)

	(1)	(2)	(3)	(4)	(5)	(6)
	Institutional deliveries		4 antenatal visits		Ante tetanus shot	
	Lagged	Covar	Lagged	Covar	Lagged	Covar
β_1: Arm 1 x post 2006 (initial incentive effect)	21.1***	23.3***	5.6*	13.2***	1.6	-3.4
	(1.9)	(2.6)	(3.1)	(3.3)	(2.0)	(3.4)
β_2: Arm 2 x post 2006 (uncond. finance effect)	13.1***	15.4***	3.3	10.9***	-2.9	-7.9**
	(2.9)	(3.3)	(2.8)	(2.9)	(1.9)	(3.2)
β_3: Arm 1 x post 2008 (med. run incentive effect)	10.9***	12.8***	9.9***	8.7***	3.3*	5.7*
	(2.9)	(1.9)	(2.8)	(2.6)	(1.9)	(3.0)
β_4: Arm 2 x post 2008 (scale-up effect)	15.3***	17.2***	12.9***	11.8***	4.2	6.5*
	(3.2)	(2.3)	(2.0)	(1.6)	(2.5)	(3.5)
Observations	19,804	30,673	13,341	20,552	13,066	20,397
2005 mean (full sample)	28.5	29.5	18.0	19.5	61.5	60.1
Hypothesis testing						
β_1 - β_2 = 0 (incentives vs. uncond. financing)	8.0	7.9	2.3	2.3	4.5	4.5
p-value	[0.01]	[0.01]	[0.49]	[0.48]	[0.10]	[0.09]
β_1 + β_3 = 0 (med. run incentives, total)	32.0	36.0	15.5	21.9	4.9	2.3
p-value	[0.00]	[0.00]	[0.00]	[0.00]	[0.08]	[0.64]
(β_1 + β_3) - (β_2 + β_4) = 0 (arm 1 vs. 2, post 2008)	3.5	3.5	-0.7	-0.8	3.6	3.7
p-value	[0.19]	[0.21]	[0.83]	[0.81]	[0.18]	[0.17]

Note: * p < 0.1, ** p < 0.05, *** p < 0.01. Coefficients are from difference-in-differences regressions, multiplied by 100 for exposition. All specifications include covariates, regional fixed effects, and birth month-year fixed effects. Standard errors are clustered at the regional level and are shown in parenthesis.

Compared to no intervention, incentives and unconditional financing increased institutional deliveries by 21 and 13 percentage points, respectively. After two years, incentives generated an additional increase of 11 percentage points for areas continuing with incentives, and areas transitioning from unconditional financing to incentives experienced a 15-percentage point increase. The total medium-term effect ($\beta_1 + \beta_3$) of the incentives is 32 percentage points. Finally, two years after the scaling-up of incentives, the improvements in arm 2 areas are not statistically distinguishable from the total effect in the early adopter areas.

While there were no initial differences between incentives and unconditional financing for four antenatal visits (β_1 vs. β_2), we find that the incentives increased completion by 6 percentage points relative to no program (marginally significant). Thus, the lack of a statistically significant difference between the incentives and unconditional financing is due to an improvement in both groups. After two years, incentives generated an additional 10 percentage point increase (β_3), and areas transitioning from unconditional financing to incentives experienced a 13-percentage point increase (β_4).

Neither trial arm had systematic effects on tetanus prophylaxis. In the medium run, incentives increased tetanus prophylaxis by 5 percentage points ($\beta_1 + \beta_3$, marginally significant). The comparison of the coefficients on the incentives (β_1) and unconditional financing (β_2) shows that the initial difference between them (marginally significant) is due to a relative decrease for the latter group.

Appendix tables B.5 and B.6 present results using the outcome-specific synthetic controls. The coefficients are similar.

Discussion

Substantively, we find that, for a subset of incentivized indicators, both incentives and unconditional financing had positive impacts in the short-run, incentives had additional positive impacts in the medium-run, and program scale-up was effective. The positive effects of unconditional financing align with findings from evaluations of subsequent PBF pilots. In particular, a study in Zambia found that PBF and unconditional financing had similar impacts and that both were improvements over no intervention (22). Similarly, an evaluation in Cameroon found that both PBF and unconditional financing paired with additional supervision increased some indicators (4).

With respect to medium-run outcomes, there is little evidence on the medium-run and scale-up effects of PBF programs outside of one experiment in Argentina. The Argentina study found that temporarily increasing targeted incentives increased the rate of timely antenatal care and that the effect persisted even after the study ended (23). Further experimental and/or observational studies like this analysis remain important in light of the overall mixed and conflicting evidence of PBF in LMICs.

Methodologically, we show that observational methods can be applied to secondary data to complement or substitute for dedicated research studies. In our application to Rwanda's PBF program, this approach produced new insights, including a new counterfactual representing no intervention and estimates of unconditional financing effects, medium-term impacts, and scale-up effects. These parameters are substantively important for PBF and health policy in general.

The approach can be implemented at low cost where secondary data is available but also has limitations, including the risk of confounding due to concurrent events. Additionally, there are often time-lags until secondary data is available, suggesting the approach may be best suited to examine previously uninvestigated programs and to generate additional insights from completed studies.

Bibliography

1. Donato K, Miller G, Mohanan M, Truskinovsky Y, Vera-Hernández M. Personality Traits and Performance Contracts: Evidence from a Field Experiment among Maternity Care Providers in India. American Economic Review. 2017 May;107(5):506–10.

2. Wang H, Zhang L, Yip W, Hsiao W. An Experiment In Payment Reform For Doctors In Rural China Reduced Some Unnecessary Care But Did Not Lower Total Costs. Health Affairs. 2011 Dec 1;30(12):2427–36.

3. Yip W, Powell-Jackson T, Chen W, Hu M, Fe E, Hu M, et al. Capitation Combined With Pay-For-Performance Improves Antibiotic Prescribing Practices In Rural China. Health Aff. 2014 Mar 1;33(3):502–10.

4. de Walque D, Robyn PJ, Saidou H, Sorgho G, Steenland M. Looking into the Performance-Based Financing Black Box [Internet]. World Bank; 2017. (Policy Research Working Paper). Report No.: 8162. Available from: http://documents.worldbank.org/curated/en/834601502391015068/Looking-into-the-performance-based-financing-black-box-evidence-from-an-impact-evaluation-in-the-health-sector-in-Cameroon

5. Basinga P, Gertler PJ, Binagwaho A, Soucat AL, Sturdy J, Vermeersch CM. Effect on maternal and child health services in Rwanda of payment to primary health-care providers for performance: an impact evaluation. The Lancet. 2011;377(9775):1421–8.

6. de Walque D, Gertler PJ, Bautista-Arredondo S, Kwan A, Vermeersch C, de Dieu Bizimana J, et al. Using provider performance incentives to increase HIV testing and counseling services in Rwanda. Journal of Health Economics. 2015 Mar;40:1–9.

7. Skiles MP, Curtis SL, Basinga P, Angeles G. An equity analysis of performance-based financing in Rwanda: are services reaching the poorest women? Health Policy Plan. 2013 Dec 1;28(8):825–37.

8. Ngo DKL, Sherry TB, Bauhoff S. Health system changes under pay-for-performance: the effects of Rwanda's national programme on facility inputs. Health Policy Plan. 2017 Feb;32(1):11–20.

9. RBFHealth website. Retrieved on July 9, 2018 [Internet]. 2018. Available from: https://www.rbfhealth.org/

10. Saksena P, Antunes AF, Xu K, Musango L, Carrin G. Mutual health insurance in Rwanda: Evidence on access to care and financial risk protection. Health Policy. 2011 Mar;99(3):203–9.

11. Sherry TB, Bauhoff S, Mohanan M. Multitasking and Heterogeneous Treatment Effects in Pay-for-Performance in Health Care: Evidence from Rwanda. American Journal of Health Economics. 2017;3(2):192–226.

12. Shapira G, Kalisa I, Condo J, Humuza J, Mugeni C, Nkunda D. Effects of Performance Incentives for Community Health Worker Cooperatives in Rwanda [Internet]. Washington, D.C.: World Bank; 2017. Report No.: 8059. Available from: https://elibrary.worldbank.org/doi/abs/10.1596/1813-9450-8059

13. O'Neill S, Kreif N, Grieve R, Sutton M, Sekhon JS. Estimating causal effects: considering three alternatives to difference-in-differences estimation. Health Services and Outcomes Research Methodology. 2016 Jun;16(1–2):1–21.

14. Blundell R, Dias MC. Alternative Approaches to Evaluation in Empirical Microeconomics. J Human Resources. 2009 Jul 1;44(3):565–640.

15. Abadie A, Diamond A, Hainmueller J. Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program. Journal of the American Statistical Association. 2010;105(490):493–505.

16. Billmeier A, Nannicini T. Assessing Economic Liberalization Episodes: A Synthetic Control Approach. The Review of Economics and Statistics. 2012 Oct 10;95(3):983–1001.

17. Kreif N, Grieve R, Hangartner D, Turner AJ, Nikolova S, Sutton M. Examination of the Synthetic Control Method for Evaluating Health Policies with Multiple Treated Units: Synthetic Control Method for Evaluating Health Policies. Health Economics. 2016 Dec;25(12):1514–28.

18. Hackmann MB, Kolstad JT, Kowalski AE. Health Reform, Health Insurance, and Selection: Estimating Selection into Health Insurance Using the Massachusetts Health Reform. American Economic Review (Papers and Proceedings). 2012;102(3):498–501.

19. Bauhoff S. The Effect of School Nutrition Policies in California on Dietary Intake and Obesity: A Synthetic Control Approach. Economics & Human Biology. 2014;12:45–55.

20. A. Smith J, E. Todd P. Does matching overcome LaLonde's critique of nonexperimental estimators? Journal of Econometrics. 2005;125(1–2):305–53.

21. Kaul A, Klößner S, Pfeifer G, Schieler M. Synthetic Control Methods: Never Use All Pre-Intervention Outcomes Together With Covariates. MPRA Working Paper 83790 [Internet]. 2018 [cited 2018 May 14]; Available from: https://econpapers.repec.org/paper/pramprapa/83790.htm

22. Zeng W, Shepard DS, Nguyen H, Chansa C, Kumar A, Qamruddin J, et al. Cost–effectiveness of results-based financing, Zambia: a cluster randomized trial. Bulletin of the World Health Organization. 2018;Online First.

23. Celhay P, Gertler P, Giovagnoli P, Vermeersch C. Long-Run Effects of Temporary Incentives on Medical Care Productivity [Internet]. World Bank; 2015. Report No.: 7348. Available from: http://documents.worldbank.org/curated/en/736861467997636007/Long-run-effects-of-temporary-incentives-on-medical-care-productivity

Appendices

You can find the appendices at **cgdev.org/sites/default/files/Ngo-Bauhoff_Performance-Based-Financing_APPEND.pdf**.