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

Performance-based financing (PBF) programs have been introduced in numerous developing countries to increase the provision and quality of health services through financial incentives. Despite growing evidence about short-term impacts of PBF, less is known about medium-run impacts and scale-up effects, and about how PBF compares to other financing approaches. In this paper, we extend the initial evaluation of Rwanda's PBF program to identify medium-run and scale-up effects of incentives and unconditional financing relative to a new "business as usual" counterfactual. We use data from the Demographic and Health Surveys from Rwanda and several Sub-Saharan African countries from 2001 to 2010, using two control group strategies: all available control regions, and a subset of 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 Rwandan program's impacts on four indicators: institutional deliveries, antenatal tetanus prophylaxis, completion of any antenatal visits, and completion of four antenatal visits. The results are similar using the various control groups and in additional robustness checks. In the short-run and relative to no intervention, both performance-based and unconditional financing raised institutional delivery rates and completion of four antenatal visits. In the medium-run, relative to no intervention and in addition to the initial short-run impacts, performance-based incentives resulted in further improvements for institutional deliveries. Program scale-up was effective, with few differences between intervention arms after all areas received performance-based incentives. There were few effects on antenatal tetanus prophylaxis or on completion of any antenatal visits. Together, the results suggest that PBF can have persistent effects for some indicators, but unconditional financing can also be effective. Moreover, the analysis demonstrates how observational research methods and secondary data can generate new insights on completed trials.

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

* This paper was originally published in November 2018. It was revised in May 2021. The original version can be viewed here: https://www.cgdev.org/sites/default/files/Ngo-Bauhoff_Medium-Run_Scale-Up_Effects_Performance-Based_Financing.pdf

Center for Global Development
www.cgdev.org

Working Paper 497
May 2021*

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, and Center for Global Development

We thank Mirabel Levine for excellent research assistance.

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.

Ngo would like to thank the Occidental College Robinson Scholars Program for support.

Diana Ngo and Sebastian Bauhoff. 2018, revised 2021. "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

1. 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 include efforts by 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 (Donato et al., 2017; Wang et al., 2011; Yip et al., 2014).

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 (de Walque et al., 2017), 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 focused on the initial two year period (Basinga et al., 2011; de Walque et al., 2015; Ngo et al., 2017; Skiles et al., 2013). 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 particularly important 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. Indeed, an evaluation of Mozambique's PBF was only able to detect impacts after the program had been effective for 18 months (Rajkotia et al., 2017).

Second, published PBF 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 by comparing the effect of two alternative treatments, performance-based incentives versus additional unconditional financing (Basinga et al., 2011). 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 financing approach (RBFHealth, 2018). Comparisons against the status quo are important; several evaluations with multiple arms have found positive effects from

unconditional financing relative to a pure control, raising concerns about the relative cost-effectiveness of PBF (Borghi et al., 2015; de Walque et al., 2017; Zeng et al., 2018).

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, by applying difference-in-differences methods to data from Rwanda and control countries, we extend Rwanda's 2006 PBF trial in two ways. First, we identify the effects of performance-based incentives and additional unconditional financing relative to no intervention by constructing a control group using all countries in Sub-Saharan Africa that had comparable data and did not implement PBF or similar interventions. The 2006 trial only analyzed the alternative treatments against each other. Second, we measure the medium-run (52 months) and scale-up effects of the PBF program using the new "business as usual" 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 program was expanded nationally and left no remaining control group from within Rwanda. We obtain similar estimates using two separate strategies for identifying relevant control groups, using all available control regions and using a "synthetic control group" that consists of regions with similar pre-trends in outcomes and covariates. In addition, the results are robust to two additional checks, one in which we sequentially drop individual countries from the control group and the other where we generate alternative synthetic control groups based on different specifications. We interpret these results in the context of broader health reforms in Rwanda while directly controlling for expansions of health insurance and community health workers.

We find that PBF increased institutional delivery rates in the short-run and had additional positive effects in the medium-run. Unconditional financing also increased institutional delivery rates in the short run but had a smaller effect than incentives. For the completion of four antenatal visits, both incentives and unconditional financing increased completion relative to no program in the short-run; both types of financing had similar effect sizes. For antenatal tetanus prophylaxis and completion of any antenatal visits, we do not find any consistent effects from incentives or from unconditional financing in the short-run or medium-run. Program scale-up was effective for the outcomes tested, with few differences between intervention arms after the performance-based incentives were adopted nationally.

Together, the results suggest that PBF may generate positive impacts for some, but not all indicators, and PBF can generate additional improvements in the medium-run. However, the analysis also suggests that unconditional financing can be effective for some indicators, suggesting that increased resources are important in and of themselves.

While there were some positive results from Rwanda's program, there is need for more research to understand whether or not these impacts are specific to Rwanda's context or more broadly generalizable. To this end, the analysis also highlights the potential for observational research methods and secondary data to generate new insights on existing evaluations. This is particularly important when a limited number of interventions can be tested in an RCT and when program expansion prevents identification of medium-run and scale-up effects.

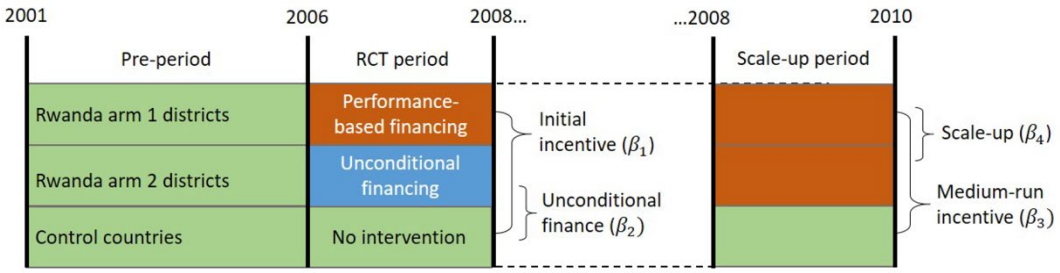
The paper proceeds as follows. Section 2 provides background information on Rwanda's PBF intervention, and Section 3 describes the data for the analysis. Section 4 lays out the analytical framework and empirical strategy, and Section 5 presents the results. Section 6 concludes.

2. Background on Rwanda's PBF trial

In 2006, Rwanda introduced the PBF program as a strategy towards accelerating progress toward the Millennium Development Goals (Basinga et al., 2011). The program was implemented in public and "government-assisted" facilities (e.g., faith-based or non-profit facilities); these account for the vast majority of health facilities in Rwanda. In addition to the traditional input-based budgets, program facilities received varying unit payments for 24 service indicators in the domains of maternal health, child health, family planning and HIV/AIDS (Appendix Table B.1). 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. This basic structure of Rwanda's PBF—and the focus on maternal and child health—has served as the basis for subsequent PBF programs in other LMICs (Fritsche and Peabody, 2018).

The Rwanda PBF trial was phased in a quasi-randomized fashion to allow for rigorous evaluation and designed to isolate the effect of incentives from the effect of additional resources. 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. The incentives were scaled up to arm 2 districts in April 2008. Figure 1 presents a timeline of the program rollout.

Figure 1: Intervention timeline



Note: ßs refer to coefficients from the medium-term regressions.

Several studies have evaluated the program's short-run impact. Basinga et al. (2011) 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 any antenatal visits or on four antenatal visits (Basinga et al., 2011). 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 antenatal tetanus prophylaxis or completion of four antenatal visits (Sherry et al., 2017). Outside of maternal indicators, Basinga et al. identify positive effects on child preventative care visits, and de Walque et al. find positive effects on HIV testing (de Walque et al., 2015). In addition, Sherry et al. report positive spillovers on some unrewarded services but find no effects on health outcomes. Other studies have explored heterogeneous effects and report that the policy was neither pro-rich nor pro-poor (Skiles et al., 2013), and have studied the corresponding health systems changes, finding that the intervention may have increased provider attendance and improved management (Ngo et al., 2017).

More broadly, Rwanda's PBF program was implemented in the context of large and ongoing governance reforms in the mid-2000s, including decentralization and a program of performance-based contracting called *imihigo*. *Imihigo* was a multi-sector program in which local governments made public commitments to particular targets; within the health sector, this included a goal of expanding health insurance (Basinga et al., 2011; Saksena et al., 2011). In evaluating the effects of PBF relative to a counterfactual of no intervention, these concurrent reforms are difficult to fully disentangle from the PBF program. (Within-Rwanda comparisons should not be affected by these national efforts.) However, *imihigo's* effects on the health sector were reportedly low due to lack of financial resources and accountability mechanisms (Hasselskog, 2016), and thus overshadowed by the PBF program, which came with formal accountability structures and substantial monetary resources. As discussed below, we also directly control for household health insurance coverage, accounting for the insurance expansion over the period. For these reasons we focus on the PBF program when discussing our findings while acknowledging that the PBF program was implemented in a particular context and amidst other reforms. We provide additional discussion in Appendix A.

3. Data

Our analytical strategy to estimate the PBF program's impact compares trends in outcomes between Rwanda and its control regions. We use DHS data, standardized household surveys used for monitoring and evaluating indicators related to population, health, and nutrition. As described below, this approach requires annual data; this requirement determines which control countries and outcomes are included in the analysis.

Treatment areas in Rwanda

For Rwanda, we include the DHS from 2005, 2007, and 2010. With the five year recall period in the DHS, these surveys allow us to create a pseudo-time series from 2001 to 2010, covering the pre-intervention period until mid-2010, when Rwanda implemented additional subnational PBF reforms (Shapira et al., 2018). We use the same treatment assignment as in Basinga et al. (2011) and Sherry et al. (2017), which accounts for a change in administrative boundaries after the randomization. As result of this change, this and the earlier studies are quasi-randomized. We provide details in Appendix A on classifying treatment and control

areas using geographic location identifiers. We also drop Rwandan districts that received PBF prior to 2006.

Control countries

To identify regions that could serve as controls, we begin with all countries in Sub-Saharan Africa that conducted standard DHS and interim DHS surveys, omitting more specialized surveys such as the DHS-supported Malaria Indicator Surveys (MIS).

We then restrict the pool of potential controls on substantive and data grounds. First, we reviewed published literature on health policies to identify and exclude areas that were affected by similar policies during the relevant period, with the goal of identifying a control group that represents "business as usual." Following this review, described in detail in Appendix A, we excluded Zambia and Senegal as potential control countries because they implemented large scale PBF and free delivery policies, respectively, in the same period. Second, we include all remaining neighboring countries with data for every year from 2001 through 2010. This allows us to measure multiple pre-treatment years (2001 through 2005) and the relevant post-treatment periods (2006 through 2007 for the initial phase and 2008 through 2010 for scale-up). Given the five-year recall period, this data requirement implies that each control country must have two or three DHS surveys in the relevant time period, with an early survey fielded between 2001 and 2006, a late survey fielded between 2010 and 2015, and a maximum of six years between surveys. We consider a year of data complete if there is information on outcomes for six or more months, with at least 10 observations per month. Appendix Table B.2 lists all DHS surveys from Sub-Saharan Africa and notes whether or not a country is eligible for inclusion.

Following this process, we include nine control countries that did not have any relevant large-scale programs during the study period and meet the data requirements: Benin (2001, 2006, 2011–12), Ethiopia (2005, 2011), Kenya (2003, 2008–09, 2014), Lesotho (2004, 2009, 2014), Malawi (2004, 2010), Nigeria (2003, 2008, 2013), Tanzania (2004–05, 2010, 2015–16), Uganda (2006, 2011), and Zimbabwe (2005–06, 2010–11).

Outcomes

For outcomes, we use all incentivized indicators that are available at the annual level and measured consistently in the DHS data, as described in Appendix Table B.1. Some incentivized indicators, like the HIV/AIDS indicators, are not observed at all in the DHS. Others, like curative care visits, are available but not at the annual level. Instead, these indicators refer to one specific point in time e.g., curative care visits in the last two weeks before the interview date. We exclude these indicators because we cannot compare outcomes measured in different years. For example, we cannot compare outcomes in Rwanda 2010 to outcomes in Nigeria in 2008.

There are six incentivized indicators that are available at the annual level: institutional deliveries, completion of any antenatal visits, completion of four antenatal visits, antenatal

tetanus prophylaxis, antenatal malaria prophylaxis, and on-time completion of child vaccinations. Specifically, the DHS collects information about children born in the five years prior to the survey and the associated care during pregnancy and delivery. By matching the outcomes to the year of care provided, we can construct a pseudo-time series of annual data, observing outcomes in the same geography (region) over time. For example, a 2010 DHS contains information on children born between 2005 through 2010 for each region in the country. This cross-sectional survey therefore allows us to observe outcomes in the same regions for a five-year period.

We exclude antenatal malaria prophylaxis due to inconsistencies in survey coding. Specifically, the malaria items are asked separately for each drug, and the listed drugs differ both between countries and across surveys within countries. For example, Rwanda's 2005 survey asks about fansidar, amodiaquine, quinine, other unspecified drugs, and unknown drugs, while Rwanda's 2010 survey asks about coartem, quinine, and other unspecified drugs. A visual check shows large discrepancies in malaria prophylaxis across Rwanda's survey waves; similar discrepancies occur in other countries.

We also exclude childhood vaccinations as we cannot construct the PBF indicator (completed vaccinations by 12 months of age) for every country and year. That is because the outcome is not realized for children that are younger than 12 months at the interview date. For example, Ethiopia has surveys from 2005 and 2011, capturing data on births from 2000 through 2005, and from 2006 to 2011. In this context, we cannot measure 12-month vaccinations in 2006: in the 2005 survey, children born in 2005 are not observed when they are 12-months old in 2006, and in the 2010 survey, the earliest data come from births in 2006, which are only associated with 12-month vaccinations in 2007. Similar issues occur in other control countries.

We therefore focus on the remaining four incentivized outcomes: institutional deliveries (with an incentive of U.S. $4.59 per unit), completion of one antenatal visit ($0.09), completion of four antenatal visits ($0.37), and antenatal tetanus prophylaxis ($0.46).

We define institutional deliveries as delivery in any type of health care facility, including public, private, religious/volunteer hospitals, health centers, and clinics. We include in our definition facilities that are not part of the PBF (such as private facilities) because, conceptually, we focus on population-level impacts of the PBF. This definition of institutional deliveries allows us to compare overall health system utilization in Rwanda to utilization in the control countries, where different facility classifications may exist, and avoids issues arising from switching between facility types. It also aligns with that of earlier and related papers on Rwanda's PBF (Basinga et al., 2011; Sherry et al., 2017; Skiles et al., 2013). Empirically, almost all facility deliveries in Rwanda in our study period are at public/government-assisted facilities, which are included in the PBF program.

In contrast to deliveries, data on antenatal care are only available for the most recent birth, resulting in smaller sample sizes.

In total, our analysis covers four of the eight maternal and child health indicators studied in the initial Basinga et al. study. The initial trial reported large (deliveries), modest (tetanus prophylaxis), and no improvements (any and four antenatal visits) for these outcomes, so the outcomes we study span the range of initial effect sizes.

Covariates

In addition to the four outcome variables, we use a set of mother's characteristics as covariates in our analyses. These include demographics and wealth indicators; these covariates are very similar to those considered in the initial RCT (Basinga et al., 2011). For the variables we include, basic demographic variables and antenatal care outcomes are already consistently defined across surveys. We do some additional data cleaning to generate consistent definitions of institutional deliveries, household health insurance, and the availability of community health workers. See Appendix A for details.

Health insurance and community health workers

We also control for Rwanda's expansion of health insurance and community health worker program during the study period, as both expansions could have affected utilization. As Basinga et al. (2011) note, these nationwide programs are unlikely to bias comparisons of the two trial arms within Rwanda, but they could affect our comparisons with the control countries. Controlling for these expansions reduces the residual risk of confounding and also accounts for similar efforts in the control countries. Time-invariant differences between countries and across regions within countries, such as fixed differences in health systems, are controlled for in the regressions using region fixed-effects (described below).

We define insurance availability at the household level, aggregating information on all insurance-related questions for all members of the household and for all types of care. Specifically, we define a household as insured if any respondent reported health insurance coverage for any related question. For community health workers, we aggregate availability up to the level of the primary sampling unit (village). Specifically, we define a village as having a community health worker if anyone in the village reported receiving any services or information (e.g., family planning or treatment advice) from a community health worker.

Socioeconomic status

We control for household socioeconomic status with eight variables that are likely to represent the same levels of wealth across different countries and years (Rutstein and Staveteig, 2014). These variables are: having a television, having a landline phone, having a refrigerator, having a car/truck, and the number of unsatisfied basic needs as measured by four separate indicators. The basic needs are defined as having non-dirt flooring, having an adequate toilet (an improved latrine or better), having adequate drinking water (piped or bottled for urban areas, any protected water for rural areas), and low economic dependency (three or fewer household members per working individual).

4. Analytical framework

To identify causal estimates, we estimate difference-in-differences regressions that compare changes in the Rwandan arm 1 and 2 districts over time with changes in control countries. Since the Rwandan program was implemented separately in different regions, we use regions within each of the control countries as comparable units of analysis. We use two strategies for identifying control groups from control country data. For both strategies, the identification assumption is that outcomes in the Rwandan districts would have evolved similarly to those in the control group, in the absence of the PBF intervention. Using these controls, the counterfactual is a "business as usual" scenario, where "business as usual" consists of updated five-year health plans, a mixture of smaller regional interventions, and a few national, non-PBF interventions.

As noted above, we interpret our findings in the context of the broader reforms in Rwanda, by controlling for increasing insurance status and availability of community health workers, and using the terms "incentives" and "unconditional financing" to also include *imihigo*, the multi-sector performance-based contracting reforms.

Strategy 1: Full sample controls

First, we use data from all available control countries (Benin, Ethiopia, Lesotho, Kenya, Malawi, Nigeria, Tanzania, Uganda, and Zimbabwe) as one control group. This method is the most direct and straightforward. While our evidence review did not identify large-scale policies that would have affected the control countries' trajectories, the control countries continued to operate with goals of improving their health systems. These other ongoing efforts are likely to improve outcomes in the control group, attenuating any positive estimates of Rwanda's PBF impact.

To guard against the risk that some country had other relevant policy changes that we do not observe, we also re-estimate the difference-in-differences regressions by removing countries from our control group one at a time. If the results do not change meaningfully when dropping each of the control countries, this suggests that the results are less likely to be driven by confounding policies in any given control country.

Strategy 2: Synthetic controls

Second, using a newer method drawn from the comparative case study literature, we construct a "synthetic control group" that consists of control countries' regions with similar pre-trends in outcomes and covariates. This control group is specifically constructed to resemble Rwanda in the pre-intervention period and strengthens the likelihood that, in the absence of the policy, treatment and control groups would have experienced similar changes in the outcomes over time. The difference-in-differences analyses then control for time-invariant residual biases (Blundell and Dias, 2009).

Specifically, we use the synthetic control method (SCM) to identify a relevant set of control areas from the full set of control regions. 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 the outcome of interest, where the outcome is predicted by a set of variables that can include both pre-treatment outcomes and key covariates (Abadie et al., 2010).

SCM has been used to study a range of topics including the impacts of state-level anti-smoking legislation (Abadie et al., 2010), trade liberalization (Billmeier and Nannicini, 2012), and hospital PBF schemes (Kreif et al., 2016). It has 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 (Abadie et al., 2010). It has also been used to generate weights that are then applied in regression analyses (Bauhoff, 2014; Hackmann et al., 2012). 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. Our approach is similar to multivariate matching combined with difference-in-differences methods (Smith and Todd, 2005); we use the synthetic control method because it performs well in generating balance (O'Neill et al., 2016).

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 is representative. For Rwanda, we pool pre-intervention data from arm 1 and 2 districts, aggregating to a single treatment area, since 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).

To construct the synthetic control group, we must identify the outcome of interest on which to match the pre-intervention trends. Under the rationale that PBF affects the entire health system rather than individual indicators, our preferred strategy matches on an aggregate outcome that reflects multiple indicators. We call this aggregate system measure the maternal service rate and construct it by equally weighting rates of delivery care (institutional delivery) and antenatal care. We define the antenatal care rate as the simple average of the rates of any antenatal visits, four antenatal visits, and the antenatal tetanus shot.

Using the maternal service rate as the synthetic control outcome, we use two strategies for choosing predictors. First, we create a "lagged" synthetic control by applying SCM to all pre-intervention lagged outcomes from 2001 through 2005 (Kreif et al., 2016). Specifically, we include the rates of institutional deliveries, any and four antenatal visits, and tetanus prophylaxis as separate predictors. Including outcomes separately for every year prior to the intervention generates the control group that best matches Rwanda's pretreatment outcomes. After including all lagged outcomes, additional covariates have no predictive power. However, covariates are often important predictors of the outcomes and trends, suggesting that they should be included in the synthetic control algorithm (Kaul et al.,

2018). Thus, we generate a second "covariate" synthetic control using a combination of pre-intervention outcomes and covariates as predictors (Kaul et al., 2018). 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.

Since the maternal service rate is an average of several outcomes, the synthetic controls specified above do not guarantee that the control groups will match the levels and trends of Rwanda for each separate outcome. Therefore, as a robustness check, we also generate synthetic control groups using each outcome separately. For example, we construct a synthetic control group based on the pre-intervention trends in institutional deliveries and a separate synthetic control group based on antenatal tetanus prophylaxis. Within each of these outcome-specific synthetic controls, we generate two separate control groups as above.

Difference-in-differences regressions

Using these various control groups, we use the individual data to run two sets of difference-in-differences regressions, including subnational region fixed-effects to account for time-invariant differences in health system contexts. The main identifying assumption is that Rwanda's outcomes would have evolved similarly to those of the control units, in the absence of the PBF program.

The first set of regressions 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 (Basinga et al., 2011). We then use data from 2001 through 2008, ending before the scale-up in 2008. The specification for the short-term effect of performance incentives is:

$$Outcome_{irt} = \tau\, Arm\,1_r\, x\, post\,2006_t + \delta\, Rwanda_r\, x\, post\,2006_t$$

$$+\alpha_r + \gamma_t + X_{irt}\delta + \varepsilon_{irt} \quad \text{(short-term regression)}$$

where $Rwanda_r$ is an indicator for Rwanda arm 1 or arm 2 districts. This term identifies the average effect of additional financing relative to the control regions, since both arms in Rwanda received additional financing. $Arm\,1_r$ is an indicator for Rwanda arm 1 districts and identifies the additional effect of incentives above the unconditional financing; this was the effect identified in the initial RCT. $post\,2006_t$ indicates the time after the initial roll-out. α_r are subnational region fixed effects, γ_t are month-year fixed effects, and X_{irt} are individual covariates. Standard errors are clustered at the region 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 a specification using only Rwanda.

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:

$$Outcome_{irt} = \beta_1\ Arm\ 1_r\ x\ post\ 2006_t + \beta_2\ Arm\ 2_r\ x\ post\ 2006_t + \beta_3\ Arm\ 1_r\ x\ post\ 2008_t$$

$$+\beta_4\ Arm\ 2_r\ x\ post\ 2008_t + \alpha_r + \gamma_t + X_{irt}\delta + \varepsilon_{irt} \quad \text{(medium-term regression)}$$

where $Arm\ 2_r$ is an indicator for Rwanda arm 2 districts, $post\ 2008_t$ indicates the time after the scale-up phase, and all other variables are as above.

For both the short- and medium-term regressions, we consider deliveries as treated if the birth occurred during the treatment period. We consider antenatal outcomes as treated if more than half of the pregnancy occurred during the treatment period (i.e., the pregnancy was five months or less at the beginning of the treatment period). Five months is also the approximate median date of the first visit in Rwanda in our study period. We date the outcomes analogously for the construction of the synthetic control groups described above.

In the medium-term regression, 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, including the comparison of PBF and additional unconditional financing in the short-run, the total effects on each arm after the initial and scale-up phases, and the final differences between the two arms. Figure 1 displays the coefficients of interest within the intervention rollout and identifies the relevant groups for comparison for each coefficient.

Weighting

To implement the difference-in-differences regressions with individual-level data, we normalize the sample weights under the assumption that the control regions represent the same number of people as Rwanda's treatment arm areas. Specifically, within each year, we normalize the weights for Rwanda's arm 1 and arm 2 areas to sum to one, separately. For the full sample control group, we transform the individual weights so that the sum of weights within each control region sum to one within each year. For the synthetic control group, we transform the weights, multiplying the individual weights and synthetic control region weights, so that the sum of weights for all synthetic control observations equals one within each year.

5. Results

Table 1 presents the summary statistics for Rwanda and the three control groups: all controls, the synthetic control with lagged outcomes as predictors, and the synthetic control with covariates and outcomes as predictors. The table displays the average levels for 2001 through 2005 as well as the average trends from 2001 through 2005.

Table 1: Summary statistics, 2001 to 2005

Outcome/ covariate	Year	Group averages Mean/(s.e.)				Difference from Rwanda Diff/[p-val]		
		Rwanda	All controls	Lagged synth	Covar synth	All controls	Lagged synth	Covar synth
Institutional deliveries	2001–2005 avg	0.26	0.50	0.28	0.28	-0.24	-0.03	-0.03
		(0.46)	(0.50)	(0.50)	(0.50)	[0.00]	[0.01]	[0.02]
	Annual trend	0.02	0.01	0.01	0.01	0.02	0.01	0.01
		(0.00)	(0.00)	(0.01)	(0.01)	[0.00]	[0.24]	[0.23]
Any antenatal visits	2001–2005 avg	0.95	0.84	0.69	0.71	0.11	0.27	0.24
		(0.23)	(0.36)	(0.52)	(0.51)	[0.00]	[0.00]	[0.00]
	Annual trend	0.00	0.00	0.03	0.01	0.01	-0.03	-0.01
		(0.00)	(0.00)	(0.01)	(0.01)	[0.07]	[0.00]	[0.34]
Four antenatal visits	2001–2005 avg	0.12	0.56	0.32	0.34	-0.43	-0.18	-0.21
		(0.36)	(0.50)	(0.52)	(0.53)	[0.00]	[0.00]	[0.00]
	Annual trend	0.00	-0.02	0.01	0.01	0.04	0.01	0.01
		(0.00)	(0.00)	(0.01)	(0.01)	[0.00]	[0.39]	[0.42]
Antenatal tetanus shot	2001–2005 avg	0.64	0.75	0.65	0.61	-0.09	0.00	0.05
		(0.53)	(0.43)	(0.53)	(0.55)	[0.00]	[0.86]	[0.00]
	Annual trend	0.01	-0.01	0.02	0.01	0.02	-0.01	0.01
		(0.01)	(0.00)	(0.01)	(0.01)	[0.00]	[0.57]	[0.40]
Birth order	2001–2005 avg	3.93	3.46	3.94	3.86	0.47	-0.01	0.07
		(2.60)	(2.39)	(2.85)	(2.82)	[0.00]	[0.85]	[0.23]
	Annual trend	-0.05	-0.02	-0.11	-0.03	-0.03	0.06	-0.02
		(0.03)	(0.01)	(0.03)	(0.04)	[0.23]	[0.16]	[0.67]
Age under 20	2001–2005 avg	0.05	0.15	0.16	0.16	-0.10	-0.11	-0.11
		(0.23)	(0.36)	(0.41)	(0.41)	[0.00]	[0.00]	[0.00]
	Annual trend	-0.01	-0.00	0.00	0.00	-0.00	-0.01	-0.01
		(0.00)	(0.00)	(0.01)	(0.01)	[0.24]	[0.12]	[0.10]
Age over 35	2001–2005 avg	0.18	0.11	0.13	0.11	0.07	0.06	0.07
		(0.41)	(0.31)	(0.37)	(0.36)	[0.00]	[0.00]	[0.00]
	Annual trend	-0.01	-0.00	-0.01	-0.00	-0.00	0.00	-0.01
		(0.00)	(0.00)	(0.00)	(0.00)	[0.29]	[0.66]	[0.40]

Primary education	2001–2005 avg	0.69	0.70	0.39	0.58	-0.01	0.30	0.11
		(0.49)	(0.46)	(0.55)	(0.55)	[0.07]	[0.00]	[0.00]
	Annual trend	0.01	0.01	0.01	0.00	0.00	-0.00	0.01
		(0.00)	(0.00)	(0.01)	(0.01)	[0.62]	[0.64]	[0.24]
Household size	2001–2005 avg	5.63	6.31	6.46	5.78	-0.68	-0.84	-0.15
		(2.08)	(3.05)	(3.51)	(2.41)	[0.00]	[0.00]	[0.00]
	Annual trend	-0.07	0.00	0.01	-0.03	-0.07	-0.09	-0.04
		(0.02)	(0.01)	(0.04)	(0.03)	[0.00]	[0.04]	[0.30]
Health insurance, household	2001–2005 avg	0.57	0.02	0.01	0.00	0.54	0.56	0.56
		(0.52)	(0.15)	(0.09)	(0.08)	[0.00]	[0.00]	[0.00]
	Annual trend	0.03	0.01	0.00	0.00	0.03	0.03	0.03
		(0.01)	(0.00)	(0.00)	(0.00)	[0.00]	[0.00]	[0.00]
Urban	2001–2005 avg	0.08	0.22	0.11	0.06	-0.15	-0.03	0.02
		(0.28)	(0.42)	(0.35)	(0.26)	[0.00]	[0.00]	[0.00]
	Annual trend	0.00	0.00	0.01	-0.00	-0.00	-0.01	0.00
		(0.00)	(0.00)	(0.00)	(0.00)	[0.98]	[0.22]	[0.51]
Has comm. health worker	2001–2005 avg	0.58	0.80	0.84	0.86	-0.22	-0.26	-0.27
		(0.52)	(0.40)	(0.41)	(0.39)	[0.00]	[0.00]	[0.00]
	Annual trend	0.01	0.01	0.01	0.02	0.00	0.00	-0.01
		(0.01)	(0.00)	(0.00)	(0.00)	[0.56]	[0.56]	[0.42]
Num. of unsatisfied basic needs†	2001–2005 avg	2.44	2.14	2.66	2.53	0.30	-0.22	-0.09
		(0.89)	(1.10)	(0.91)	(0.86)	[0.00]	[0.00]	[0.00]
	Annual trend	-0.07	-0.02	-0.04	0.00	-0.05	-0.03	-0.07
		(0.01)	(0.00)	(0.01)	(0.01)	[0.00]	[0.10]	[0.00]
Num. of durable assets†	2001–2005 avg	0.07	0.45	0.14	0.12	-0.38	-0.07	-0.05
		(0.36)	(0.94)	(0.56)	(0.49)	[0.00]	[0.00]	[0.00]
	Annual trend	0.01	0.07	0.08	0.02	-0.05	-0.06	-0.01
		(0.00)	(0.00)	(0.01)	(0.01)	[0.00]	[0.00]	[0.20]

Note: The statistics are calculated using linear regressions; standard errors are shown in parenthesis and p-values are shown in brackets. Statistics are weighted. Within each year, weights for Rwanda's arm 1 and arm 2 areas sum to one, separately. For the full sample control group, weights within each control region sum to one within each year. For the synthetic control group, weights for all synthetic control observations sum to one within each year. The synthetic control groups are matched according to the maternal service rate using outcomes as predictors (lagged specification) and outcomes and covariates as predictors (covariate specification).

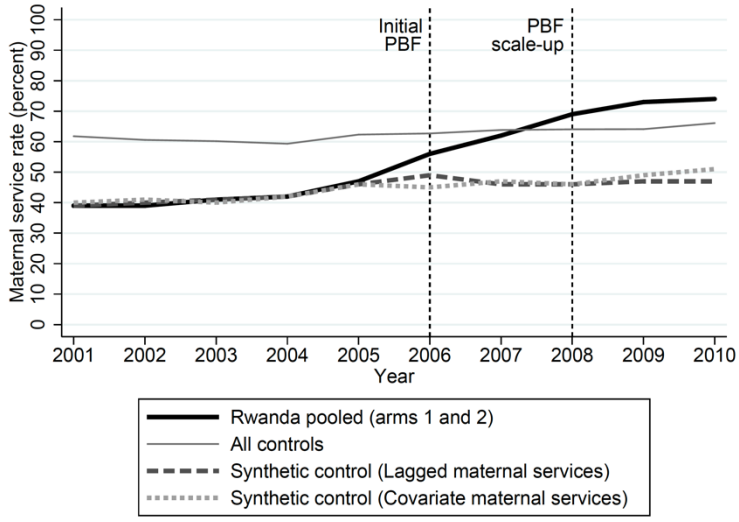
† Total unsatisfied basic needs and total durable assets are shown for conciseness for the covariate predictors. Instead, the analysis includes indicators for one, two, three, and four unsatisfied basic needs, and indicators for ownership of televisions, landline telephones, cars/trucks, and refrigerators.

Overall, compared to Rwanda, the full set of control regions have similar trends in most covariates, higher levels of institutional deliveries, four antenatal visits and antenatal tetanus prophylaxis, and lower levels of any antenatal visits. The full set of controls have slowly increasing rates of institutional deliveries and modest decreases in four antenatal visits and antenatal tetanus prophylaxis.

The synthetic control groups are combinations of various regions across different countries that are weighed to resemble Rwanda in the pre-period. (See Appendix Table B.3 for the specific weights on each region). The synthetic control groups are generally more similar to Rwanda in the levels and trends of most outcomes and covariates than the full set of controls. However, some differences remain, likely because the synthetic controls in Table 1 are constructed using the maternal service rate, our aggregate health system measure. Summary statistics using the synthetic control groups constructed separately by outcome are discussed below and in Appendix Table B.4.

Figure 2 presents the evolution of maternal service rates for Rwanda and the three control groups. Relative to the full set of controls, both synthetic control groups are more similar to Rwanda in the pre-period as expected, since the synthetic control groups were constructed to minimize differences in pre-intervention maternal service rates between Rwanda and the controls. Both the levels and trends in maternal service rates are very similar to the treatment areas in the pre-intervention period, suggesting that SCM works well in generating a counterfactual for Rwanda. Appendix Figures B.1 through B.4 show the plots disaggregated by outcome and treatment arm.

Figure 2: Maternal services by year



Note: The maternal service rate is defined as the simple average of the rates of institutional deliveries and antenatal care. Within this, antenatal care is defined as the simple average of any antenatal visits, four antenatal visits, and antenatal tetanus prophylaxis. Rates of each outcome are first calculated for each region and then averaged across regions.

Table 2 shows the results from the short-term regression that uses data prior to the PBF scale-up and focuses on the effect of PBF relative to additional unconditional financing. Overall, the results are generally comparable to those identified in Basinga et al.'s RCT, which are reproduced in column 1 and only use data from the years 2005 and 2007. Relative to unconditional financing, incentives increased institutional delivery rates by 8 to 9 percentage points and had no effect on completion of four antenatal visits. For any antenatal visits and antenatal tetanus prophylaxis, the 2001 through 2008 results (right-hand

columns) align with the initial RCT, while the results using only 2005 and 2007 are less consistent. There is no consistent effect for any antenatal visits, with marginally negative effects in some specifications. There are marginally significant positive effects on antenatal tetanus prophylaxis using 2001 through 2008; these are not significant using only 2005 and 2007 data.

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

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	2005, 2007					2001–2008			
	Basinga RCT	Rwanda only	All controls	Lagged synth	Covar synth	Rwanda only	All controls	Lagged synth	Covar synth
Institutional deliveries									
Arm 1 x post 2006 (incentives vs. uncond. finance)	8.1**	8.9**	8.8**	8.4**	8.5**	8.2***	8.2***	8.0***	8.0***
	[p = 0.02]	(3.7)	(4.0)	(3.7)	(3.7)	(2.5)	(2.6)	(2.5)	(2.5)
Observations	2,108	3,064	42,834	5,047	4,897	11,184	137,247	18,265	17,648
2005 mean (Rwanda arm 1)	35	32.2	32.2	32.2	32.2	32.2	32.2	32.2	32.2
Any antenatal visits									
Arm 1 x post 2006 (incentives vs. uncond. finance)	0.2	-2.9*	-3.5**	-3.2*	-3.5*	-0.2	-0.2	-0.1	0.0
	[p = 0.88]	(1.6)	(1.5)	(1.7)	(1.7)	(1.0)	(0.9)	(0.9)	(0.9)
Observations	2,309	1,930	28,794	3,093	3,166	6,962	88,310	11,552	11,305
2005 mean (Rwanda arm 1)	95	98.2	98.2	98.2	98.2	98.2	98.2	98.2	98.2
Four antenatal visits									
Arm 1 x post 2006 (incentives vs. uncond. finance)	0.8	-2.9	-4.8	-3.8	-3.3	1.2	0.5	1.0	1.0
	[p = 0.83]	(2.9)	(3.0)	(3.1)	(3.2)	(3.4)	(3.2)	(3.4)	(3.5)
Observations	2,223	1,930	28,794	3,093	3,166	6,962	88,310	11,552	11,305
2005 mean (Rwanda arm 1)	18	22.9	22.9	22.9	22.9	22.9	22.9	22.9	22.9
Antenatal tetanus shot									
Arm 1 x post 2006 (incentives vs. uncond. finance)	5.1*	-4.0	0.1	-2.6	-2.4	4.4	5.0**	4.9*	4.8*
	[p = 0.06]	(4.7)	(5.2)	(4.8)	(4.6)	(2.6)	(2.4)	(2.5)	(2.5)
Observations	2,856	1,916	29,322	3,075	3,145	6,910	89,262	11,477	11,242
2005 mean (Rwanda arm 1)	71	70.7	70.7	70.7	70.7	70.7	70.7	70.7	70.7

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, region fixed effects, and month-year fixed effects. Standard errors are clustered by region and are shown in parenthesis. Weights are applied as described in the text and Table B.3. The synthetic control groups are matched according to the maternal service rate using outcomes as predictors (lagged specification) and outcomes and covariates as predictors (covariate specification).

Table 3: Differential effects by arm and phase, 2001-2010 (percentage points)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Institutional deliveries			Any antenatal visits			Four antenatal visits			Antenatal tetanus shot		
	All controls	Lagged synth	Covar synth	All controls	Lagged synth	Covar synth	All controls	Lagged synth	Covar synth	All controls	Lagged synth	Covar synth
β1: Arm 1 x post 2006 (initial incentive effect)	23.9***	21.2***	21.6***	-1.4	-4.4**	-3.2	15.3***	13.0***	15.2***	6.3***	2.4	-1.3
	(1.4)	(4.7)	(2.7)	(1.0)	(2.0)	(2.1)	(2.3)	(2.3)	(3.4)	(1.9)	(3.3)	(4.2)
β2: Arm 2 x post 2006 (uncond. finance effect)	15.7***	13.2**	13.6***	-1.2	-4.2**	-3.1	14.8***	12.1***	14.2***	1.4	-2.3	-6.0
	(2.4)	(5.0)	(3.2)	(1.0)	(2.0)	(2.0)	(2.6)	(2.8)	(3.7)	(2.0)	(3.5)	(4.5)
β3: Arm 1 x post 2008 (med. run incentive effect)	12.2***	11.7***	9.9***	0.7	1.2	-1.8**	4.5*	7.7**	0.1	2.4	4.5	1.9
	(1.6)	(1.8)	(2.1)	(0.7)	(1.8)	(0.7)	(2.6)	(3.2)	(3.7)	(1.6)	(4.1)	(2.3)
β4: Arm 2 x post 2008 (scale-up effect)	16.8***	16.2***	14.2***	-1.9**	-1.6	-4.4***	6.8***	10.5***	2.9	4.2*	5.8	3.5
	(2.0)	(2.3)	(2.5)	(0.9)	(1.7)	(0.8)	(1.6)	(2.6)	(3.5)	(2.4)	(4.5)	(3.2)
Observations	195,266	23,826	22,818	127,945	15,894	15,496	127,945	15,894	15,496	127,480	15,530	15,201
2005 mean (full sample)	52.4	31.3	31.4	85.9	86.5	87.7	53.7	27.3	28.0	75.2	72.1	67.8
Hypothesis testing												
β1 - β2 = 0 (incentives vs. uncond. financing)	8.2	8.0	8.0	-0.2	-0.2	-0.1	0.5	1.0	1.0	4.9	4.7	4.6
p-value	[0.00]	[0.00]	[0.00]	[0.79]	[0.87]	[0.93]	[0.88]	[0.78]	[0.78]	[0.04]	[0.06]	[0.07]
β1 + β3 = 0 (med. run incentives, total)	36.1	32.9	31.6	-0.7	-3.2	-5.0	19.8	20.7	15.3	8.7	7.0	0.5
p-value	[0.00]	[0.00]	[0.00]	[0.56]	[0.20]	[0.04]	[0.00]	[0.00]	[0.00]	[0.00]	[0.08]	[0.91]
β2 + β4 = 0 (med. run arm 2, total)	32.5	29.4	27.8	-3.1	-5.8	-7.5	21.6	22.6	17.1	5.6	3.5	-2.5
p-value	[0.00]	[0.00]	[0.00]	[0.04]	[0.03]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.32]	[0.56]
(β1 + β3) - (β2 + β4) = 0 (arm 1 vs. 2, post 2008)	3.6	3.5	3.8	2.4	2.6	2.5	-1.8	-1.8	-1.8	3.1	3.5	3.0
p-value	[0.19]	[0.19]	[0.16]	[0.09]	[0.03]	[0.04]	[0.55]	[0.56]	[0.57]	[0.17]	[0.14]	[0.21]

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, region fixed effects, and month-year fixed effects. Standard errors are clustered by region and are shown in parenthesis. Weights are applied as described in the text and Table B.3. The synthetic control groups are matched according to the maternal service rate using outcomes as predictors (lagged specification) and outcomes and covariates as predictors (covariate specification).

Table 3 shows the results from the medium-term regression. The results for institutional deliveries are positive and significant for all coefficients across all three control groups. Compared to no intervention, incentives increased institutional deliveries by 21 to 24 percentage points, while unconditional financing increased institutional deliveries by 13 to 16 percentage points. After two years, incentives generated an additional increase of 10 to 12 percentage points for areas continuing with incentives, and areas transitioning from unconditional financing to incentives in the scale-up period experienced a 14 to 17 percentage point increase. Overall, the total medium-term effect ($\beta_1 + \beta_3$) of the incentives is 32 to 36 percentage points, and the total effect on arm 2 areas ($\beta_2 + \beta_4$) is 28 to 33 percentage points. Finally, by 2010, 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 (($\beta_1 + \beta_3$)- ($\beta_2 + \beta_4$)).

For completion of four antenatal visits, there are consistent and positive short-run effects for both arms of the intervention. Relative to no intervention, vs. β_2, also reported in the RCT and replicated in Table 2) is due to a similar improvement in both groups. At the scale-up phase, there are some positive and significant effects for both arms, but these are not consistently significant across specifications. Over the full time period, the results indicate that completion of four antenatal visits increased by 15 to 21 percentage points in arm 1 ($\beta_1 + \beta_3$) and by 17 to 23 percentage points in arm 2 ($\beta_2 + \beta_4$). There are no differences between the arms at the end of the period (($\beta_1 + \beta_3$)- ($\beta_2 + \beta_4$)).

For any antenatal visits and tetanus prophylaxis, the estimated effects are not consistent across the regressions. For any antenatal visits, there are a few negative and significant effects while for tetanus prophylaxis there are some positive and significant effects using the full control group, but there are no consistent effects across the regressions for any arm or roll-out phase. For any antenatal care, these results may reflect the already very high levels of any antenatal care visits in Rwanda prior to the program (95 to 98 percent, Table 2) that left little room for improvement. They could also reflect the small incentive, 9 U.S. cents for each first antenatal visit, which is the lowest unit payment among all PBF indicators. For tetanus prophylaxis, the comparison of the coefficients on the incentives and unconditional financing (β_1 vs. β_2, marginally significant in all regressions) provides a useful check on the findings of a positive effect of the incentive arm relative to unconditional financing in the initial RCT (Table 2). In particular, the results from the synthetic control groups (columns 11 and 12) suggest that the relatively better performance of the incentive arm may not represent an improvement in the incentive arm relative to no intervention but instead is due to a decrease in the unconditional financing arm relative to no intervention.

Table 4: Robustness using alternate synthetic controls matched by outcome, differential effects by arm and phase, 2001-2010 (percentage points)

Outcome	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Institutional deliveries		Any antenatal visits		Four antenatal visits		Antenatal tetanus shot	
	Lagged Outcome	Covar Outcome	Lagged Outcome	Covar Outcome	Lagged Outcome	Covar Outcome	Lagged Outcome	Covar Outcome
β1: Arm 1 x post 2006 (initial incentive effect)	21.6***	20.2***	-0.0	1.3	12.7***	12.5***	-0.2	1.0
	(1.4)	(2.4)	(0.8)	(1.0)	(2.6)	(2.5)	(2.6)	(3.5)
β2: Arm 2 x post 2006 (uncond. finance effect)	13.7***	12.2***	0.1	1.5	11.6***	11.5***	-4.8	-3.2
	(2.0)	(3.0)	(1.0)	(1.2)	(3.2)	(3.1)	(2.9)	(3.7)
β3: Arm 1 x post 2008 (med. run incentive effect)	13.3***	10.2***	1.0**	2.7*	3.5	2.1	1.6	1.6
	(1.7)	(2.4)	(0.5)	(1.4)	(2.9)	(3.0)	(2.3)	(3.1)
β4: Arm 2 x post 2008 (scale-up effect)	17.6***	14.6***	-1.8***	0.2	6.4***	5.1**	2.7	2.3
	(2.1)	(2.7)	(0.7)	(1.5)	(2.1)	(2.4)	(3.0)	(3.6)
Observations	195,266	28,455	127,006	14,825	13,680	15,682	127,480	15,278
2005 mean (full sample)	31.3	31.1	97.5	97.3	19.4	19.0	71.1	70.9
Hypothesis testing								
β1 - β2 = 0 (incentives vs. uncond. financing)	7.9	8.1	-0.2	-0.2	1.0	1.0	4.6	4.2
p-value	[0.00]	[0.00]	[0.85]	[0.82]	[0.76]	[0.78]	[0.07]	[0.11]
β1 + β3 = 0 (med. run incentives, total)	34.8	30.5	0.9	4.0	16.1	14.6	1.4	2.6
p-value	[0.00]	[0.00]	[0.12]	[0.04]	[0.00]	[0.00]	[0.72]	[0.59]
β2 + β4 = 0 (med. run arm 2, total)	31.2	26.7	-1.6	1.6	18.0	16.6	-2.1	-0.9
p-value	[0.00]	[0.00]	[0.18]	[0.46]	[0.00]	[0.00]	[0.56]	[0.84]
(β1 + β3) - (β2 + β4) = 0 (arm 1 vs. 2, post 2008)	3.6	3.8	2.6	2.3	-1.9	-2.0	3.4	3.5
p-value	[0.15]	[0.16]	[0.02]	[0.04]	[0.53]	[0.51]	[0.13]	[0.14]

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, region fixed effects, and month-year fixed effects. Standard errors are clustered by region and are shown in parenthesis. Weights are applied as described in the text and Table B.3. The outcome synthetic control is matched by each outcome, e.g., matching deliveries for the deliveries regression. The lagged specification uses outcomes as predictors, while the covariate specification uses outcomes and covariates as predictors.

Table 4 presents the results from robustness checks using alternate synthetic control groups constructed using lagged outcome and covariate predictors for each outcome separately (for details, see Appendix B.2 and B.3). Compared to the synthetic control groups identified using the aggregate maternal services outcome, the outcome-specific synthetic control groups are more similar to Rwanda in the levels and trends for each outcome, as expected. Overall, the medium-run and scale-up effects using the alternative synthetic controls are similar to those in Table 3. In Table 4, there are again large effects for institutional deliveries that are quite similar in magnitude to the full set of controls and the maternal service synthetic controls. For four antenatal visits, there continue to be positive and significant effects for both incentives and unconditional financing in the short term; these are of similar magnitude to the main results. For any antenatal visits, there are some statistically significant coefficients using the outcome-specific synthetic controls, but overall, no consistent results emerge across all five control groups (Table 3 and Table 4). Again, there are no effects for antenatal tetanus prophylaxis.

The results from the remaining robustness checks also align with our main findings and suggest that the effects we find are unlikely to be due to confounding changes in any specific control region, particularly as the different synthetic controls are composed of various and distinct combinations of regions. Appendix Table B.5 replicates the initial RCT effects using the alternative controls and broadly reproduces the conclusions from Table 2. Appendix Tables B.6 through B.9 present the differential effects of the program by arm and phase, dropping each country from the full set of controls one at a time as a robustness check. The coefficients are quite similar in magnitude and significance to the estimates from the full set of controls.

6. 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 (Zeng et al., 2018). Similarly, an evaluation in Cameroon found that both PBF and unconditional financing paired with additional supervision increased some indicators (de Walque et al., 2017).

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 two years after the study ended (Celhay et al., 2019). 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 research studies using primary data collection. In our application to Rwanda's PBF program, the combination of standard observational approaches and secondary data—including from countries that were not part of the initial trial—produced new insights, including 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 usual limitations of observational studies apply, including the risk of confounding due to concurrent events and the issue of time-lags until secondary data is available. This suggests that this approach may be best suited to examine previously uninvestigated programs and to generate additional insights from completed studies.

Our analysis has several limitations. First, we cannot cleanly attribute the differences between Rwanda and other countries to the PBF because of concurrent national reforms in Rwanda. However, the PBF had substantially more resources and accountability structures than the broader *imihigo* plan, and we directly control for the health insurance and community health worker expansion. Second, our identifying assumption is that Rwanda's outcomes would have evolved similar to those of the control countries in absence of the PBF, but it is possible that Rwanda may have improved anyway. This is a general concern with the difference-in-differences design that we mitigate by using the synthetic control group, which is specifically constructed to match Rwanda's pre-intervention trends. Visually, we see that while Rwanda was improving prior to the program, its improvement accelerated substantially at the time of the program's introduction. Third, although we carefully scrutinized the control countries, it is possible that there remain unobserved interventions that could confound our estimates. For instance, our literature search did not turn up a 2007 ban on traditional birth attendants in Malawi that could have impacted institutional deliveries (Godlonton and Okeke, 2016).[1] To guard against the risk of having missed some—possibly not documented—interventions, we conducted a robustness check in which we dropped each individual country from the set of control countries. Finally, our findings may not generalize to other countries, contexts or programs. However, Rwanda's PBF served as a blueprint for designs of PBF programs in other countries and remains an influential case study.

[1] We are grateful to an anonymous reviewer for this example.

Bibliography

Abadie, A., Diamond, A., Hainmueller, J., 2010. Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program. Journal of the American Statistical Association 105, 493–505. https://doi.org/10.1198/jasa.2009.ap08746

Basinga, P., Gertler, P.J., Binagwaho, A., Soucat, A.L., Sturdy, J., Vermeersch, C.M., 2011. Effect on maternal and child health services in Rwanda of payment to primary health-care providers for performance: an impact evaluation. The Lancet 377, 1421–1428. https://doi.org/10.1016/S0140-6736(11)60177-3

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

Billmeier, A., Nannicini, T., 2012. Assessing Economic Liberalization Episodes: A Synthetic Control Approach. The Review of Economics and Statistics 95, 983–1001. https://doi.org/10.1162/REST_a_00324

Blundell, R., Dias, M.C., 2009. Alternative Approaches to Evaluation in Empirical Microeconomics. J. Human Resources 44, 565–640. https://doi.org/10.3368/jhr.44.3.565

Borghi, J., Little, R., Binyaruka, P., Patouillard, E., Kuwawenaruwa, A., 2015. In Tanzania, The Many Costs Of Pay-For-Performance Leave Open To Debate Whether The Strategy Is Cost-Effective. Health Aff 34, 406–414. https://doi.org/10.1377/hlthaff.2014.0608

Celhay, P.A., Gertler, P.J., Giovagnoli, P., Vermeersch, C., 2019. Long-Run Effects of Temporary Incentives on Medical Care Productivity. American Economic Journal: Applied Economics 11, 92–127. https://doi.org/10.1257/app.20170128

de Walque, D., Gertler, P.J., Bautista-Arredondo, S., Kwan, A., Vermeersch, C., de Dieu Bizimana, J., Binagwaho, A., Condo, J., 2015. Using provider performance incentives to increase HIV testing and counseling services in Rwanda. Journal of Health Economics 40, 1–9. https://doi.org/10.1016/j.jhealeco.2014.12.001

de Walque, D., Robyn, P.J., Saidou, H., Sorgho, G., Steenland, M., 2017. Looking into the Performance-Based Financing Black Box (No. 8162), Policy Research Working Paper. World Bank.

Donato, K., Miller, G., Mohanan, M., Truskinovsky, Y., Vera-Hernández, M., 2017. Personality Traits and Performance Contracts: Evidence from a Field Experiment among Maternity Care Providers in India. American Economic Review 107, 506–510. https://doi.org/10.1257/aer.p20171105

Fritsche, G., Peabody, J., 2018. Methods to improve quality performance at scale in lower- and middle-income countries. Journal of Global Health 8. https://doi.org/10.7189/jogh.08.021002

Godlonton, S., Okeke, E.N., 2016. Does a ban on informal health providers save lives? Evidence from Malawi. J Dev Econ 118, 112–132. https://doi.org/10.1016/j.jdeveco.2015.09.001

Hackmann, M.B., Kolstad, J.T., Kowalski, A.E., 2012. Health Reform, Health Insurance, and Selection: Estimating Selection into Health Insurance Using the Massachusetts Health Reform. American Economic Review (Papers and Proceedings) 102, 498–501.

Hasselskog, M., 2016. Participation or What? Local Experiences and Perceptions of Household Performance Contracting in Rwanda. Forum for Development Studies 43, 177–199. https://doi.org/10.1080/08039410.2015.1090477

Kaul, A., Klößner, S., Pfeifer, G., Schieler, M., 2018. Synthetic Control Methods: Never Use All Pre-Intervention Outcomes Together With Covariates. MPRA Working Paper 83790.

Kreif, N., Grieve, R., Hangartner, D., Turner, A.J., Nikolova, S., Sutton, M., 2016. Examination of the Synthetic Control Method for Evaluating Health Policies with Multiple Treated Units: Synthetic Control Method for Evaluating Health Policies. Health Economics 25, 1514–1528. https://doi.org/10.1002/hec.3258

Ngo, D.K.L., Sherry, T.B., Bauhoff, S., 2017. Health system changes under pay-for-performance: the effects of Rwanda's national programme on facility inputs. Health Policy Plan 32, 11–20. https://doi.org/10.1093/heapol/czw091

O'Neill, S., Kreif, N., Grieve, R., Sutton, M., Sekhon, J.S., 2016. Estimating causal effects: considering three alternatives to difference-in-differences estimation. Health Services and Outcomes Research Methodology 16, 1–21. https://doi.org/10.1007/s10742-016-0146-8

Rajkotia, Y., Zang, O., Nguimkeu, P., Gergen, J., Djurovic, I., Vaz, P., Mbofana, F., Jobarteh, K., 2017. The effect of a performance-based financing program on HIV and maternal/child health services in Mozambique—an impact evaluation. Health Policy Plan 32, 1386–1396. https://doi.org/10.1093/heapol/czx106

RBFHealth, 2018. RBFHealth website. Retrieved on July 9, 2018.

Rutstein, S.O., Staveteig, S., 2014. Making the Demographic and Health Surveys Wealth Index Comparable (No. 9), DHS Methodological Reports. IFC International.

Saksena, P., Antunes, A.F., Xu, K., Musango, L., Carrin, G., 2011. Mutual health insurance in Rwanda: Evidence on access to care and financial risk protection. Health Policy 99, 203–209. https://doi.org/10.1016/j.healthpol.2010.09.009

Shapira, G., Kalisa, I., Condo, J., Humuza, J., Mugeni, C., Nkunda, D., Walldorf, J., 2018. Going beyond incentivizing formal health providers: Evidence from the Rwanda Community Performance-Based Financing program. Health Economics 27, 2087–2106. https://doi.org/10.1002/hec.3822

Sherry, T.B., Bauhoff, S., Mohanan, M., 2017. Multitasking and Heterogeneous Treatment Effects in Pay-for-Performance in Health Care: Evidence from Rwanda. American Journal of Health Economics 3, 192–226. https://doi.org/10.1162/AJHE_a_00072

Skiles, M.P., Curtis, S.L., Basinga, P., Angeles, G., 2013. An equity analysis of performance-based financing in Rwanda: are services reaching the poorest women? Health Policy Plan. 28, 825–837. https://doi.org/10.1093/heapol/czs122

Smith, J., Todd, P., 2005. Does matching overcome LaLonde's critique of nonexperimental estimators? Journal of Econometrics 125, 305–353. https://doi.org/10.1016/j.jeconom.2004.04.011

Wang, H., Zhang, L., Yip, W., Hsiao, W., 2011. An Experiment In Payment Reform For Doctors In Rural China Reduced Some Unnecessary Care But Did Not Lower Total Costs. Health Affairs 30, 2427–2436. https://doi.org/10.1377/hlthaff.2009.0022

Yip, W., Powell-Jackson, T., Chen, W., Hu, Min, Fe, E., Hu, Mu, Jian, W., Lu, M., Han, W., Hsiao, W.C., 2014. Capitation Combined With Pay-For-Performance Improves Antibiotic Prescribing Practices In Rural China. Health Aff 33, 502–510. https://doi.org/10.1377/hlthaff.2013.0702

Zeng, W., Shepard, D.S., Nguyen, H., Chansa, C., Das, A.K., Qamruddin, J., Friedman, J., 2018. Cost–effectiveness of results-based financing, Zambia: a cluster randomized trial. Bulletin of the World Health Organization 96, 760–771. https://doi.org/10.2471/BLT.17.207100

Appendix A. Details on coding consistency across surveys and concurrent programs

Coding consistency

The DHS instruments are generally standardized across countries and waves, allowing researchers to generate comparable estimates across surveys. For this analysis, basic demographic variables and antenatal care outcomes are already consistently defined across surveys. We do some additional data cleaning to generate consistent definitions of institutional deliveries, household health insurance, and the availability of community health workers.

Institutional deliveries

For institutional deliveries, we recode the categorical survey item "Where did you give birth to (NAME)?" into a dichotomous indicator for delivering in any type of health care facility. Specifically, we define an institutional birth as any category that is not "home," "respondent's home," "other home," "traditional birth attendant (home, premise)," or "other." The specific health facility categories differ from survey to survey but broadly include all public, private, and religious/volunteer hospitals, health centers, and clinics.

Health insurance

For health insurance, there is a substantial variation from survey to survey on the inclusion of insurance related questions. This is due to the fact that insurance availability was very low in most of Sub-Saharan Africa during the early years of the study period and likely considered irrelevant for the majority of households. For example, coverage was less than 1% in Tanzania's 2004 survey. However, at the beginning of the period, health insurance coverage in Rwanda was relatively high (53%) and increased over the period.

Specifically, we use a very broad definition of insurance availability, compiling information on all insurance-related questions and aggregating insurance availability up to the household level, since different items were only asked in individual surveys. We searched all datasets within the DHS instruments (household roster, individual recode, female recode, birth recode, male recode, child recode, and couple recode) for all insurance related variables (Stata command: lookfor_all insur, vlabs). These variables included dichotomous indicators for insurance, categorical variables for insurance type, and use of insurance to pay for various types of care. Some surveys do not include any insurance-related questions, so we treat all births from those surveys as having no health insurance. While this is a strong assumption, the coverage trends within each country are plausible. For example, in Uganda, we assume that coverage is 0% in 2006, and coverage is less than 2% in 2011 when insurance-related questions are first included.

Community Health Workers

There is also substantial variation between surveys on questions related to community health workers. Again, we use a very broad definition of availability of community health workers, compiling information on all items related to community health workers and aggregating availability up to the level of the primary sampling unit, since community health workers are available at the village level. Specifically, we search the birth recode data files for the following keywords: "community," "worker," "field," "comm.," "wrkr," "chw," "hsa," and "hew," and manually identify all relevant results. The identified variables contain information on whether community health workers provided services or information related to family planning, antenatal care, delivery care, postpartum care, and treatment of various issues including fevers, diarrhea, and sexually transmitted infections. Using this definition, the overall availability levels are high for most countries, ranging from 59% (Rwanda 2005) to 100% (Tanzania 2009). While Rwanda's availability levels were among the lowest in the early years, they increased to among the highest in the later years.

Geographic and treatment assignment

As reported by Basinga et al., Rwanda changed several district boundaries after the initial randomization for the PBF trial. Some districts were merged - including those that were initially assigned to treatment and control. As a result, some of the (initially random) assignment was changed for a few districts. We follow Basinga et al. and Sherry et al. and use the same final treatment assignments in our analysis.

We use geographic identifiers provided by DHS to match the Rwandan household data to the new districts. The DHS geo-coordinates contain random positional error to maintain confidentiality. Urban and rural clusters are displaced by up to 2km and 5km, respectively; an additional 1 percent of rural clusters are randomly displaced by up to 10km. This could lead to some misclassification, e.g., if apparent control enumeration areas are in fact located in treatment areas. Any errors should not be correlated with program assignment, so this should not generate any bias, except for attenuation, in the results.

Concurrent programs in Rwanda

Rwanda's PBF program was implemented in the context of large and ongoing governance reforms. Specifically, the country underwent a process of decentralization, with various phases (2000–2003, 2004–2008, and ongoing) (MINALOC, 2006), where redistricting in late 2005/early 2006 affected the definition of Rwanda's arm 1 and arm 2 districts (Basinga et al., 2011)[2].

[2] We account for the redistricting in our analysis, using the new district definitions in our arm 1 and arm 2 classifications.

In line with this effort, Rwanda implemented a large multi-sector program of performance-based contracting in early 2006, formalizing a traditional practice referred to as *imihigo*. Under *imihigo*, local governments made public commitments to particular actions, including improving health-related indicators (Bucagu et al., 2012). Concurrently, there were concerted efforts to increase health insurance coverage. Many mutual health insurance schemes were created between 2000 and 2003, there was an effort to scale up in 2005 with external funding, and coverage was legally mandated in 2008 (Saksena et al., 2011).

To the extent that Rwanda's PBF program was rolled out concurrently with these policies, the results that we observe when comparing Rwanda to other countries are the combined effects of PBF and these programs. We control for health insurance in our regressions, which attenuates the coefficients in the analyses identifying the medium-run and scale-up effects.

Similarly, the 2006 effects we observe for arm 1 and arm 2 relative to the synthetic control (i.e., the counterfactual state of "no intervention") are the combined effects of *imihigo* and monetary performance incentives and *imihigo* and unconditional financing, respectively. In other words, we observe effects for a broader program of performance-based contracting, which consisted of public commitments and monetary incentives.

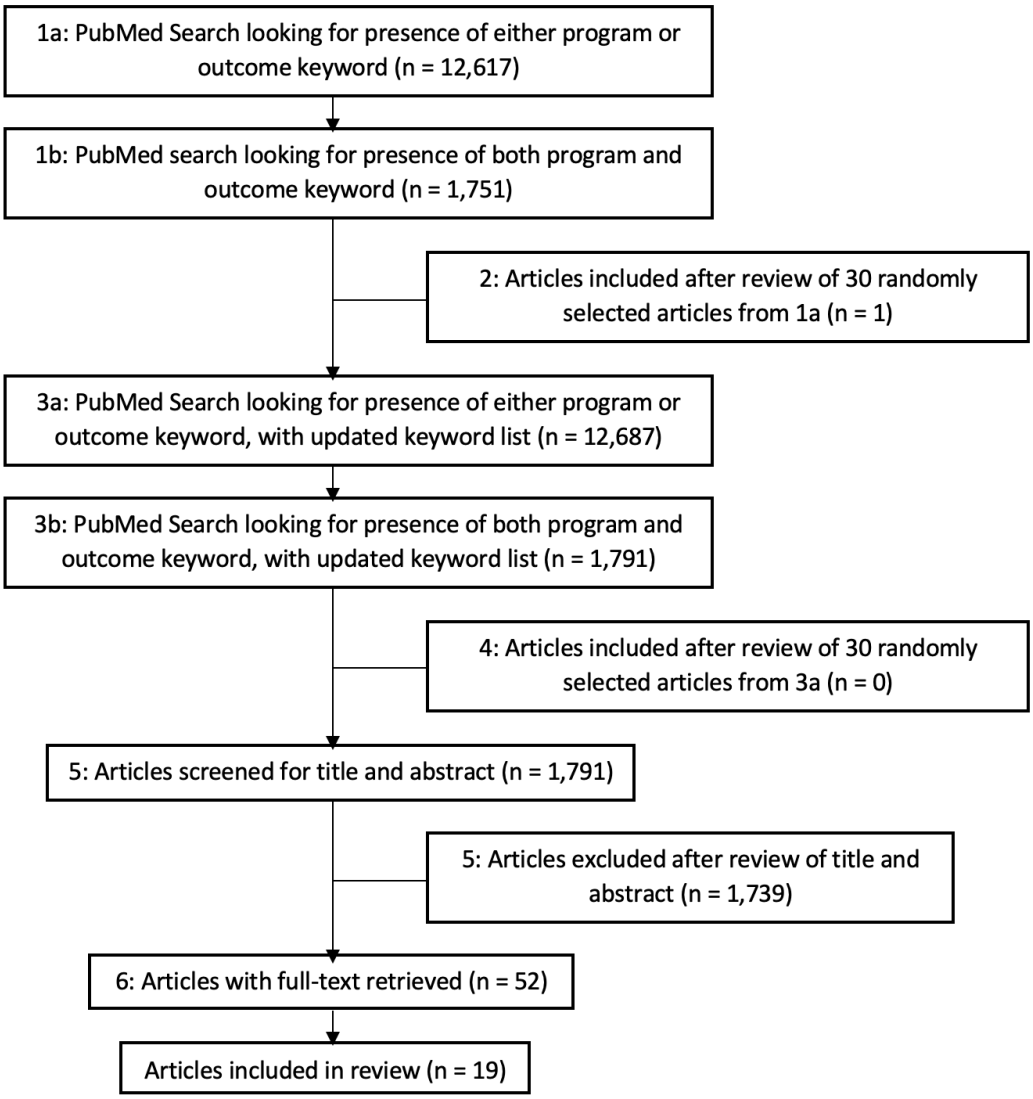
We focus on and highlight the monetary incentives for two reasons. First, observational evaluations of *imihigo* suggest that within the health sector, the main target was increased health insurance coverage, which we discuss above and control for. Moreover, studies suggest that *imihigo's* effectiveness was limited due to lack of financial resources and lack of sufficient accountability mechanisms (Hasselskog, 2016; MINALOC, 2012, 2011, 2010).

Given the formal monetary accountability and large influx of financial resources associated with PBF (an average increase in expenditures of 22% above 2006 levels (Basinga et al., 2011), we consider it plausible that a large fraction of the effects we observe are associated specifically with the financing aspect of the treatment. This is also in line with studies in Cameroon and Zambia, contexts without the public commitments component, that find sizable positive effects of performance incentives and unconditional financing for various outcomes (de Walque et al., 2015; Friedman et al., 2017).

Concurrent health policies in the control countries

We conducted a systematic search for health policies and reforms in the study countries, including Rwanda. The overall process is described in Figure A.1.

Figure A.1: Search flow and results



First, we conducted a keyword search on the PubMed database in April 2018 and March 2020.[3] This search consisted of four categories of keywords: program keywords, outcomes keywords, countries, and further restrictions (see below for the full list of keywords in each category). We conducted two searches, both searching only in titles and abstracts and both

[3] In April 2018, we focused on Rwanda, Uganda, Kenya, Tanzania, Malawi, Ethiopia and Zimbabwe. In March 2020 we expanded the search to 13 additional countries, as listed below.

restricted by publication dates between 2000 and 2020. In the first, the search looked for the presence of either a program keyword or an outcome keyword, in addition to the presence of the country and further restriction. To make the review process more manageable, we conducted a second search, which instead looked for the presence of both a program keyword and an outcome keyword.

Second, to check the credibility of this second search, we randomly selected thirty articles from the first search and reviewed. We compared these articles against the second search and reviewed, with only one article meriting further consideration. Based on this search, we added one keyword to the original program keywords list. We repeated this step once and added eight additional terms.

Third, we conducted the two searches again using the expanded list of keywords.

Fourth, we repeated the randomized process and again reviewed thirty articles. In these random articles, one was determined as relevant to review and was added to the list returned from the second search.

In the fifth step, we reviewed the titles and abstracts of this second search list for relevance. In this case, we defined relevance as whether the article had the potential to mention health-related policies or reforms during the period of question in its text. The two most common reasons for exclusion were 1) too isolated of a sample size (for instance, at the single district or facility level) and 2) a focus on trials and experimentally imposed conditions rather than policy-related situations.

Lastly, we obtained and reviewed the full texts of the relevant articles. We collected and consolidated all mentions of policies in the countries of interest and initiated during the relevant years of our study period in Table A1 below. Countries for which we searched but did not find relevant policies are omitted from Table A1 but listed below. Table A2 shows the corresponding sources.

Table A1: Overview of concurrent health policies in the control countries

Country	Program Description	Year Initiated	Year Updated	Coverage	Source
Burkina Faso	Obstetric Care Subsidy Policy	2006		National	1
Ethiopia	Health care financing reforms	2005		Regional	2
Ghana	Free Maternal Health Policy	2008		National	3
Kenya	Output-based-aid (OBA) voucher program for safe motherhood (SF), family planning (FP), and Gender-Violence Recovery Services (GBVRS)	2006	2014 program end (expected)	Regional	4
Kenya	Policy emphasis on community health	2007		National	5
Malawi	Community-based maternal and newborn care package	2008	Still active in 2011	Regional (parts of 3 of 28 districts)	6
Malawi	User-fee exemptions coupled with service reimbursements for services performed	2006	Expanded 2015	National (subset of private, not-for-profit clinics)	7
Mali	Free Caesarean Policy	2002	2005	National	8
Rwanda	Pay for performance	2006		National	9
Rwanda	Community-based health insurance scheme	2000		National	10
Rwanda	Government subsidized benefit package of health insurance	2008		National	10
Senegal	Free Delivery and Caesarean Policy	2005	Expanded 2006	Five regions in 2005 and all regions in 2006 except capital (Dakar)	11
Tanzania	Transfer of community health insurance programs (voluntary, informal) to national insurance program (mandatory, formal)	2009		National	12
Tanzania	Pay for performance	2011		Regional	13
Tanzania	Decentralization	2000	2013 program end (expected)	National	14
Tanzania	Antenatal care (ANC) program	2002	Guidelines revised 2012	National	14
Tanzania	Life-saving skills program	2000		National	14
Tanzania	Maternal and perinatal death reviews (MPDR) program	2006	Guidelines revised 2013	National	14
Tanzania	Kangaroo mother care (KMC) program	2008		National	14
Tanzania	Essential newborn care (ENC) program	2007		National	14
Tanzania	Integrated management of childhood illness (IMCI) program	2000	Program revision to include neonatal illnesses in 2006	National	14
Tanzania	Insecticide-treated bednet (ITN) program	2003		National	14
Tanzania	Insecticide-treated bednet (ITN) voucher	2004		National	14

Tanzania	Prevention of mother-to-child transmission of HIV (PMTCT) program	2006	Revisions in 2010	National	14
Tanzania	Eliminating mother-to-child HIV transmission (EMCT) program	2012	Expanded in 2013	National	14
Tanzania	Oral rehydration salts (ORS) and zinc program	2010		National	14
Tanzania	Maternal and child health (MCH) user-fee exemption	At least by 2003		National	15
Uganda	Vaccine (Gavi Vaccine introduction grant)	2002	2006 program end (expected)	National (?)	16
Uganda	Reproductive health (RH) voucher	2006	Expanded in 2008, 2012 program end (expected)	Regional	17
Uganda	Essential health services package	1999	Second version in 2010	National	18
Zambia	User Fee Removal Policy	2006	2011	Rural areas in 2006 and nationally in 2011	19

Table A2. Articles included in review

#	Citation
1	Ganaba, R., Ilbouda P., Cresswell, J., Yaogo, M., Diallo, C., Richard, F., Cunden, N., Filippi, V., Witter, S., 2016. The obstetric care subsidy policy in Burkina Faso: what are the effects after five years of implementation? Findings of a complex evaluation. BMC Pregnancy and Childbirth 16:84, 1–14. http://doi.org/10.1186/s12884-016-0875-2.
2	Pearson, L., Gandhi, M., Admasu, K., Keyes, E.B., 2011. User fees and maternity services in Ethiopia. International Journal of Gynecology & Obstetrics 115.3, 210–315. https://doi.org/10.1016/j.ijgo.2011.09.007
3	Dalinjong, P., Wang, A., Homer, C., 2018. The implementation of the free maternal health policy in rural Northern Ghana: synthesised results and lessons learnt. BMC Res Notes 11:341, 1–6. http://doi.org/10.1186/s13104-018-3452-0.
4	Njuki, R., Abuya, T., Kimani, J., Kanya, L., Korongo, A., Mukanya, C., Bracke, P., Bellows, B. and Warren, C.E., 2015. Does a voucher program improve reproductive health service delivery and access in Kenya?. BMC health services research 15(1), 206. https://doi.org/10.1186/s12913-015-0860-x
5	Wangalwa, G., Cudjoe, B., Wamalwa, D., Machira, Y., Ofware, P., Ndirangu, M. and Ilako, F., 2012. Effectiveness of Kenya's Community Health Strategy in delivering community-based maternal and newborn health care in Busia County, Kenya: non-randomized pre-test post test study. The Pan African Medical Journal 13(1).
6	Greco, G., Daviaud, E., Owen, H., Ligowe, R., Chimbalanga, E., Guenther, T., Gamache, N., Zimba, E. and Lawn, J.E., 2017. Malawi three district evaluation: Community-based maternal and newborn care economic analysis. Health policy and planning 32(1), i64–i74. https://doi.org/10.1093/heapol/czw079
7	Manthalu, G., Yi, D., Farrar, S. and Nkhoma, D., 2016. The effect of user fee exemption on the utilization of maternal health care at mission health facilities in Malawi. Health policy and planning 31(9), 1184–1192. https://doi.org/10.1093/heapol/czw050
8	Ravit, M., Philibert, A., Tourigny, C., Traore, M., Coulibaly, A., Dumont, A., Fournier, P., 2015. The Hidden Costs of a Free Caesarean Section Policy in West Africa (Kayes Region, Mali). Matern Child Health J 19:1734–1743. http://doi.org/10.1007/s10995-015-1687-0.
9	Basinga, P., Gertler, P.J., Binagwaho, A., Soucat, A.L., Sturdy, J. and Vermeersch, C.M., 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. https://doi.org/10.1016/S0140-6736(11)60177-3
10	Sekabaraga, C., Diop, F. and Soucat, A., 2011. Can innovative health financing policies increase access to MDG-related services? Evidence from Rwanda. Health policy and planning 26(2), ii52–ii62. https://doi.org/10.1093/heapol/czr070
11	Witter, S., Dieng, T., Mbengue, D., Moreira, I., De Brouwere, V., 2010. The national free delivery and caesarean policy in Senegal: evaluating process and outcomes. Health Policy and Planning 25, 384–392. http://doi.org/10.1093/heapol/czq013.
12	McIntyre, D., Ranson, M.K., Aulakh, B.K. and Honda, A., 2013. Promoting universal financial protection: evidence from seven low-and middle-income countries on factors facilitating or hindering progress. Health research policy and systems 11(1), 36. https://doi.org/10.1186/1478-4505-11–36
13	Anselmi, L., Binyaruka, P. and Borghi, J., 2017. Understanding causal pathways within health systems policy evaluation through mediation analysis: an application to payment for performance (P4P) in Tanzania. Implementation Science 12(1), 10. https://doi.org/10.1186/s13012-016-0540-1

14 Afnan-Holmes, H., Magoma, M., John, T., Levira, F., Msemo, G., Armstrong, C.E., Martínez-Álvarez, M., Kerber, K., Kihinga, C., Makuwani, A. and Rusibamayila, N., 2015. Tanzania's Countdown to 2015: an analysis of two decades of progress and gaps for reproductive, maternal, newborn, and child health, to inform priorities for post-2015. The Lancet Global Health 3(7), e396–e409. https://doi.org/10.1016/S2214-109X(15)00059-5

15 Kruk, M.E., Mbaruku, G., Rockers, P.C. and Galea, S., 2008. User fee exemptions are not enough: out-of-pocket payments for 'free'delivery services in rural Tanzania. Tropical medicine & international health 13(12), 1442–1451. https://doi.org/10.1111/j.1365-3156.2008.02173.x

16 Roberts, D.A., Ng, M., Ikilezi, G., Gasasira, A., Dwyer-Lindgren, L., Fullman, N., Nalugwa, T., Kamya, M. and Gakidou, E., 2015. Benchmarking health system performance across regions in Uganda: a systematic analysis of levels and trends in key maternal and child health interventions, 1990–2011. BMC medicine 13(1), 285. https://doi.org/10.1186/s12916-015-0518-x

17 Kanya, L., Obare, F., Warren, C., Abuya, T., Askew, I. and Bellows, B., 2013. Safe motherhood voucher programme coverage of health facility deliveries among poor women in South-western Uganda. Health policy and planning 29(1), i4–i11. https://doi.org/10.1093/heapol/czt079

18 Zikusooka, C.M., Kyomuhang, R., Orem, J.N. and Tumwine, M., 2009. Is health care financing in Uganda equitable?. African health sciences 9(2).

19 Masiye, F., Kaonga, O., Kirigia, J., 2016. Does User Fee Removal Policy Provide Financial Protection from Catastrophic Health Care Payments? Evidence from Zambia. PLoS ONE 11(1), 1–15. http://doi.org/10.1371/journal.pone.0146508.

Countries and keywords used for search

Countries (April 2018 search)
Rwanda
Uganda
Kenya
Tanzania
Malawi
Ethiopia
Zimbabwe

Countries (March 2020 search)
Benin
Burkina Faso
Cameroon
Congo
Ghana
Guinea
Lesotho
Mali
Mozambique
Namibia
Niger
Nigeria
Senegal

Program Keywords
Decentralization
Health care financing
Pay for performance
Performance based payment
Results based financing
Health insurance
Conditional cash transfer
Voucher
Community health worker
Health plan benefits
Universal health coverage
Co-payments
Out of pocket costs
User fee

Outcome Keywords
Family planning
Antenatal care
Prenatal care
Tetanus
Facility delivery
Institutional delivery
Assisted Delivery
Co-payments
Out of pocket costs
Out of pocket payments

Further restrictions
Policy
Reform
Intervention
Project
Evaluation
Evaluate
Trial
Experiment

Bibliography

Basinga, P., Gertler, P.J., Binagwaho, A., Soucat, A.L., Sturdy, J., Vermeersch, C.M., 2011. Effect on maternal and child health services in Rwanda of payment to primary health-care providers for performance: an impact evaluation. The Lancet 377, 1421–1428. https://doi.org/10.1016/S0140-6736(11)60177-3

Bucagu, M., Kagubare, J.M., Basinga, P., Ngabo, F., Timmons, B.K., Lee, A.C., 2012. Impact of health systems strengthening on coverage of maternal health services in Rwanda, 2000–2010: a systematic review. Reproductive Health Matters 20, 50–61. https://doi.org/10.1016/S0968-8080(12)39611-0

de Walque, D., Gertler, P.J., Bautista-Arredondo, S., Kwan, A., Vermeersch, C., de Dieu Bizimana, J., Binagwaho, A., Condo, J., 2015. Using provider performance incentives to increase HIV testing and counseling services in Rwanda. Journal of Health Economics 40, 1–9. https://doi.org/10.1016/j.jhealeco.2014.12.001

Friedman, J., Qamruddin, J., Chansa, C., Das, A.K., 2017. Impact Evaluation of Zambia's Health Results-Based Financing Pilot Project (Working Paper No. 120723). World Bank.

Hasselskog, M., 2016. Participation or What? Local Experiences and Perceptions of Household Performance Contracting in Rwanda. Forum for Development Studies 43, 177–199. https://doi.org/10.1080/08039410.2015.1090477

MINALOC, 2012. District Imihigo Evaluation Report 2011–2012. Rwanda Ministry of Local Government.

MINALOC, 2011. District Imihigo Evaluation Report 2010–2011. Rwanda Ministry of Local Government.

MINALOC, 2010. Report on Districts Imihigo Evaluation 2009/2010. Rwanda Ministry of Local Government.

MINALOC, 2006. Making Decentralized Service Delivery Work. Rwanda Ministry of Local Government.

Saksena, P., Antunes, A.F., Xu, K., Musango, L., Carrin, G., 2011. Mutual health insurance in Rwanda: Evidence on access to care and financial risk protection. Health Policy 99, 203–

209. https://doi.org/10.1016/j.healthpol.2010.09.009

Appendix B. Figures and tables

Appendix B provides additional disaggregated figures, the weights underlying the SCM analysis, and the results of the robustness check dropping each control country from the set of full controls.

B.1 Outcome and country inclusion criteria

For outcomes, we use all incentivized indicators that are available at the annual level and measured consistently in the DHS data. We exclude the indicators, like the HIV/AIDS measures, that are not observed in the DHS. We also exclude the indicators, like family planning visits, that are not available at the annual level, because we cannot compare outcomes measured in different years (e.g., we cannot compare Rwanda 2010 to Nigeria 2008). The remaining indicators are available at the annual level due to the recall questions on children born in the previous five years and the associated care during pregnancy and delivery. Of these, we exclude antenatal malaria prophylaxis due to inconsistencies in survey coding. We also exclude child vaccinations, as we are unable to construct the PBF indicator (completed vaccinations by 12 months of age) for every country and year. That is because the outcome is not realized for children that are younger than 12 months at the interview date. We also exclude child vaccinations due to concurrent vaccine programs in Rwanda and note the difficulty of accurately identifying concurrent vaccine campaigns in control countries as these may not be well publicized or documented.

To identify control countries, we require that a country has data on births for every year from 2001 through 2010. This is necessary to control for pre-existing differences between countries (as is standard for difference-in-differences frameworks) and provides multiple pre-treatment years that are necessary to create the synthetic control, which is chosen to match Rwanda's pre-treatment trends. In practice, this means that a control country must have 2 or 3 DHS surveys between 2001 and 2015, with a maximum of 6 years between surveys. With the five-year recall period for the relevant outcomes, this generates annual data from 2001 through 2010. We consider a year of data complete if there are observations for at least 6 months of the year, with at least 10 observations per month.

Table B.1: Rewarded indicators and available in DHS

Rewarded output and quality indicators	Payment/unit (US $)	Available in DHS	Notes
Primary Care			
Emergency referrals during curative treatment	1.83		
Curative care visits	0.18	Once per survey	
Family Planning			
First-time family planning visits	1.83	Once per survey	DHS records "current use of modern method" for family planning.
1-month contraceptive resupply	0.18	Once per survey	Contraceptive resupply can be measured from current use of a modern method requiring regular resupplies.
Maternal Health			
Deliveries in the facility	4.59	Annually for all births in previous five years	
Emergency transfers to hospital for obstetric care during delivery	4.59		
At-risk pregnancies referred to hospital for delivery during prenatal care	1.83		
Women who received 2nd dose of malaria prophylaxis during prenatal care	0.46	Annually for last birth in previous five years	Excluded due to data inconsistencies. Questions on malaria prophylaxis change between survey rounds in Rwanda and differ by country.
Women who received appropriate tetanus vaccination during prenanatal care	0.46	Annually for last birth in previous five years	
Women who completed 4 prenatal care visits	0.37	Annually for last birth in previous five years	
First prenatal care visits	0.09	Annually for last birth in previous five years	We measure whether or not a women had any antenatal visits during the pregnancy.
Child Health			
Malnourished children referred for treatment during preventive care visit	1.83		
Children who completed vaccination on time	1.83	Annually for all births in previous five years	Excluded because we are unable to construct outcome for all years.
Child (0-59 months) preventive care visits	0.18		
HIV/AIDS			
PMTCT: exposed children tested	8.93		
New pediatric clients put on ARVs	6.7		
Prevention of mother-to-child-transmission (PMTCT): partner tested	4.58		
PMTCT: women under treatment with ARVs during labor	4.58		
New adult clients put on ARVs	4.58		
HIV+ clients tested for CD4 count	4.58		
HIV+ women who use modern method of family planning	2.68		
HIV+ clients tested for TB	2.68		
Voluntary counseling and testing	0.89		
HIV+ clients treated with cotrimoxazole each month	0.44		

Note: See the text above for the inclusion criteria.

Table B.2: Demographic and Health Survey availability, Sub-Saharan Africa control countries

Country	Standard and Interim DHS	Included	Surveys included	Notes
Angola	2015, 2020			
Benin	1996, 2001, 2006, 2011-12	Y	2001, 2006, 2011-12	
Botswana	1988			
Burkina Faso	1993, 1998, 2003, 2010, 2020			
Burundi	2010, 2016-17			
Cameroon	1991, 1998, 2004, 2011, 2018			
Cape Verde	2005			
Central African Republic	1994-95			
Chad	1996-97, 2004, 2014-15			
Comoros	1996, 2012			
Congo	2005, 2011-12			Excluded due to insufficient births from 2006 (3 months coverage)
Congo Democratic Republic	2007, 2013-14			
Cote d'Ivoire	1994, 1998-99, 2011-12, 2020			
Equatorial Guinea	2011			
Eritrea	1995, 2002			
Eswatini	2006-07			
Ethiopia	2000, 2005, 2011, 2016, 2019	Y	2005, 2011	
Gabon	2000, 2012, 2019			
Gambia	2013, 2019-20			
Ghana	1988, 1993, 1998, 2003, 2008, 2014			Excluded due to insufficient births from 2009 (3 months coverage)
Guinea	1992, 1999, 2005, 2012, 2018			
Kenya	1989, 1993, 1998, 2003, 2008-09, 2014	Y	2003, 2008-09, 2014	
Lesotho	2004, 2009, 2014	Y	2004, 2009, 2014	
Liberia	1986, 2007, 2013, 2019-20			
Madagascar	1992, 1997, 2003-04, 2008-09, 2020			
Malawi	1992, 2000, 2004, 2010, 2015-16	Y	2004, 2010	Sufficient births from 2010 in the 2010 survey.
Mali	1987, 1995-96, 2001, 2006, 2012-13, 2018			Excluded due to insufficient births from 2007 (1 month coverage)
Mauritania	2000-01, 2019-20			
Mozambique	1997, 2003, 2011, 2021			
Namibia	1992, 2000, 2006-07, 2013			Excluded due to insufficient births in 2007 (2 months coverage)
Niger	1992, 1998, 2006, 2012, 2017			Excluded due to insufficient births in 2006 (4 months coverage)
Nigeria	1990, 1999, 2003, 2008, 2013, 2018	Y	2003, 2008, 2013	
Sao Tome and Principe	2008-09			
Senegal	1986, 1992-93, 1997, 1999, 2005, 2010-11			Excluded due to large scale concurrent maternal health program.
Sierra Leone	2008, 2013, 2019			
South Africa	1998, 2003, 2016			
Sudan	1989-90			
Tanzania	1991-92, 1996, 1999, 2004-05, 2010, 2015-16	Y	2004-05, 2010, 2015-16	
Togo	1988, 1998, 2013-14			
Uganda	1988-89, 1995, 2000-01, 2006, 2011, 2016	Y	2006, 2011	Sufficient 2001 births from the 2006 survey.
Zambia	1992, 1996, 2001-02, 2007, 2013-14, 2018			Excluded due to concurrent PBF program
Zimbabwe	1988, 1994, 1999, 2005-06, 2010-11, 2015, 2020	Y	2005-06, 2010-11	Sufficient 2010 births from the 2010-11 survey.

Note: Yearly data is considered complete if there are 6 or more months covered. Monthly data is considered complete if there are at least 10 births available nationally.

B.2 Additional summary statistics, with alternate control groups

Figure B.1: Institutional deliveries by year

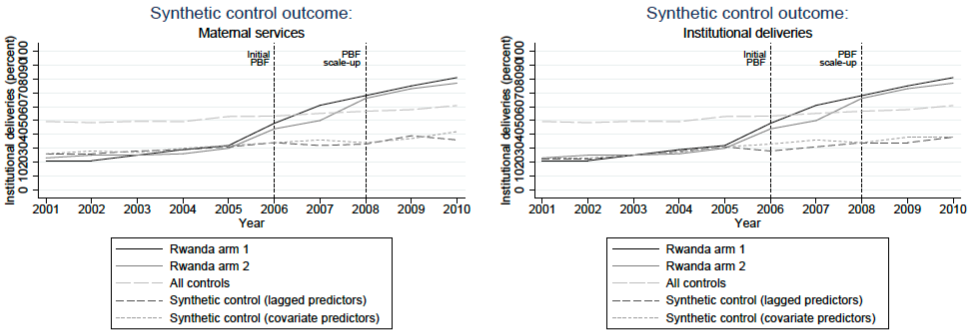


Figure B.2: Any antenatal visits by year

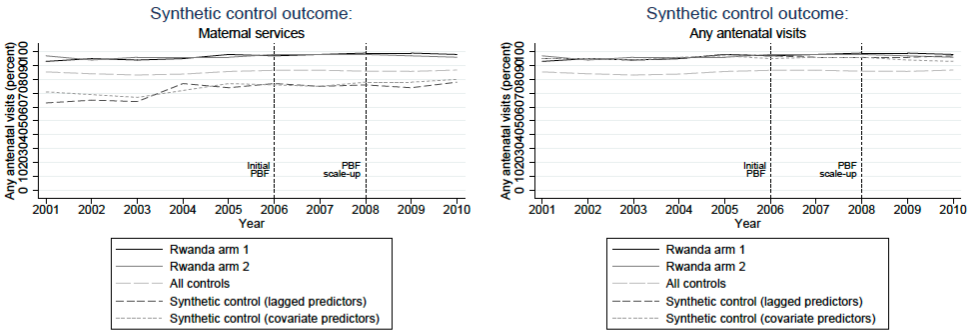


Figure B.3: Four antenatal visits by year

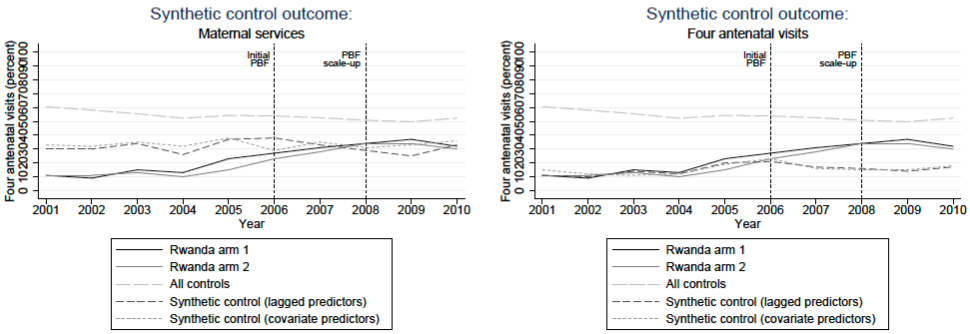
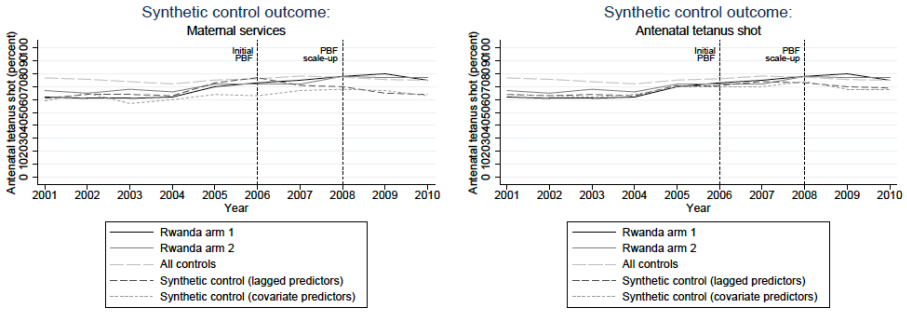


Figure B.4: Antenatal tetanus shot by year



Note: The maternal service rate is defined as the simple average of the rates of institutional deliveries and antenatal care. Within this, antenatal care is defined as the simple average of any antenatal visits, four antenatal visits, and antenatal tetanus prophylaxis. Rates of each outcome are first calculated for each region and then averaged across regions.

Table B.3: Region weights by specification

		Synthetic control outcome									
		Maternal services		**Institutional deliveries**		**Any antenatal visits**		**Four antenatal visits**		**Antenatal tetanus shot**	
Country	Region	Lagged	Covar	Lagged	Covar	Lagged	Covar	Lagged	Covar	Lagged	Covar
Benin	Atacora	-	-	0.002	-	0.002	-	-	-	0.005	-
	Atlantique	-	-	0.001	-	0.003	-	-	-	0.005	-
	Borgou	-	-	0.002	-	0.002	-	-	-	0.005	-
	Mono	-	-	0.002	-	0.004	-	-	-	0.006	-
	Oueme	-	-	0.001	-	0.004	-	-	-	0.005	-
	Zou	-	-	0.001	-	0.005	-	-	0.020	0.004	-
Ethiopia	Tigray	-	-	0.008	-	0.002	-	-	-	0.016	-
	Afar	-	-	0.010	-	0.001	-	0.121	-	0.008	-
	Amhara	0.271	0.349	0.016	0.424	0.002	-	-	0.561	0.010	-
	Oromiya	-	-	0.019	-	0.001	-	-	-	0.012	-
	Somali	-	-	0.005	-	0.001	-	0.371	0.313	0.132	0.134
	Ben-Gumuz	-	-	0.015	-	0.002	-	0.264	0.018	0.022	-
	Southern Nations	-	-	0.029	-	0.001	-	-	-	0.016	0.272
	Gambela	-	-	0.007	-	0.002	-	-	-	0.020	-
	Harari	-	-	0.005	-	0.002	-	-	-	0.006	-
	Addis Abeba	-	-	0.002	-	0.006	-	-	-	0.053	-
	Dire Dawa	-	-	0.008	-	0.002	-	-	-	0.018	-
Kenya	Nairobi	-	-	0.001	-	0.037	-	0.022	-	0.008	0.375
	Central	-	-	0.002	-	0.003	-	-	-	0.004	0.022
	Coast	-	-	0.003	-	0.018	-	-	-	0.046	-
	Eastern	-	-	0.003	-	0.002	-	-	-	0.004	-
	Nyanza	0.029	0.067	0.004	0.057	0.008	-	-	-	0.020	-
	Rift Valley	-	-	0.002	-	0.004	-	-	-	0.007	-
	Western	-	-	0.003	-	0.003	-	-	-	0.004	-
	North Eastern	0.203	-	0.503	0.029	0.002	0.035	0.222	-	0.067	-
Lesotho	Botha-Bothe	-	-	0.001	-	0.004	-	-	-	0.008	-
	Leribe	-	-	0.002	-	0.003	-	-	-	0.006	-
	Berea	-	-	0.026	0.100	0.006	-	-	-	0.006	-
	Maseru	-	-	0.002	-	0.006	-	-	-	0.012	-
	Mafeteng	-	-	0.002	-	0.004	-	-	0.087	0.005	-
	Mohale's Hoek	-	-	0.001	-	0.004	-	-	-	0.004	-
	Quthing	-	-	0.188	-	0.003	-	-	-	0.004	-
	Qacha's-Nek	-	-	0.002	0.039	0.004	0.199	-	-	0.003	-
	Mokhotlong	-	-	0.002	-	0.004	-	-	-	0.004	-
	Thaba Tseka	-	-	0.017	-	0.002	-	-	-	0.004	0.004
Malawi	Blantyre	-	-	0.001	-	0.075	-	-	-	0.006	-
	Kasungu	-	-	0.002	-	0.027	-	-	-	0.012	-
	Machinga	-	-	0.002	-	0.008	-	-	-	0.104	-
	Mangochi	0.253	-	0.003	-	0.089	-	-	-	0.005	-
	Mzimba	-	-	0.001	-	0.003	-	-	-	0.004	-
	Salima	-	-	0.001	-	0.002	-	-	-	0.003	-
	Thyolo	-	0.034	0.001	0.094	0.019	0.570	-	-	0.007	-
	Zomba	-	-	0.002	-	-	-	-	-	0.004	-
	Lilongwe	-	-	0.001	-	0.005	-	-	-	0.006	-
	Mulanje	-	0.189	0.002	-	0.221	-	-	-	0.005	-
	Other Northern	-	-	0.002	-	0.047	-	-	-	0.004	-
	Other Central	-	-	0.003	-	0.007	-	-	-	0.007	-
	Other Southern	-	-	0.002	-	0.012	-	-	-	0.008	-

Note: Continued on next page.

Region weights by specification, continued from previous page

		Synthetic control outcome									
		Maternal services		**Institutional deliveries**		**Any antenatal visits**		**Four antenatal visits**		**Antenatal tetanus shot**	
Country	Region	Lagged	Covar	Lagged	Covar	Lagged	Covar	Lagged	Covar	Lagged	Covar
Nigeria	North Central	-	-	0.002	-	0.002	-	-	-	0.005	-
	North East	-	-	0.003	-	0.002	-	-	-	0.010	-
	North West	-	-	0.003	-	0.001	-	-	-	0.011	-
	South East	-	-	0.001	-	0.003	-	-	-	0.004	-
	South South	-	-	0.001	-	0.002	-	-	-	0.009	-
	South West	-	-	0.001	-	0.003	-	-	-	0.004	-
Tanzania	Central	-	-	0.002	-	0.004	-	-	-	0.005	-
	Northern	-	-	0.002	-	0.012	-	-	-	0.007	-
	Eastern	-	-	0.001	-	0.002	-	-	-	0.005	-
	Southern	-	-	0.004	-	0.154	-	-	-	0.005	-
	Southern Highlands	-	-	0.003	-	0.019	-	-	-	0.005	-
	Western	0.243	-	0.001	-	0.003	-	-	-	0.017	-
	Lake	-	-	0.003	-	0.026	-	-	-	0.005	-
	Zanzibar	-	-	0.001	0.096	0.003	-	-	-	0.006	-
Uganda	Central 1	-	-	0.002	-	0.008	-	-	-	0.006	-
	Central 2	-	-	0.003	-	0.004	-	-	-	0.004	-
	Kampala	-	-	0.001	-	0.004	-	-	-	0.003	-
	East Central	-	-	0.002	-	0.006	-	-	-	0.004	-
	Eastern	-	-	0.003	-	0.005	-	-	-	0.003	-
	North	-	-	0.012	0.161	0.004	-	-	-	0.006	-
	West Nile	-	-	0.004	-	0.003	-	-	-	0.005	-
	Western	-	-	0.004	-	0.003	-	-	-	0.004	-
	Southwest	-	0.361	0.006	-	0.007	-	-	-	0.007	-
Zimbabwe	Manicaland	-	-	0.001	-	0.003	0.023	-	-	0.005	-
	Mashonaland Central	-	-	0.002	-	0.011	-	-	-	0.079	-
	Mashonaland East	-	-	0.001	-	0.006	-	-	-	0.005	-
	Mashonaland West	-	-	0.001	-	0.005	-	-	-	0.005	-
	Matabeleland North	-	-	0.001	-	0.004	0.058	-	-	0.006	-
	Matabeleland South	-	-	0.003	-	0.003	-	-	-	0.005	-
	Midlands	-	-	0.001	-	0.003	-	-	-	0.005	-
	Masvingo	-	-	0.001	-	0.003	0.001	-	-	0.006	0.193
	Harare	-	-	0.001	-	0.004	0.114	-	-	0.005	-
	Bulawayo	-	-	0.001	-	0.003	-	-	-	0.005	-

Note: The maternal synthetic control is matched according to the maternal service rate, while the outcome synthetic control is matched by each outcome, e.g., matching deliveries for the deliveries regression. The lagged specification uses outcomes as predictors, while the covariate specification uses outcomes and covariates as predictors.

Table B.4: Summary statistics: Alternate control groups, 2001 to 2005

		Group averages Mean / (s.e.)					Difference from Rwanda Diff / [p-val]			
Outcome/covariate	Year	Rwanda	Lagged deliveries	Covars deliveries	Lagged any antenatal	Covars any antenatal	Lagged deliveries	Covars deliveries	Lagged any antenatal	Covars any antenatal
Institutional deliveries	2001-2005 avg	0.26	0.26	0.26	0.61	0.59	0.00	-0.00	-0.33	-0.31
		(0.46)	(0.49)	(0.49)	(0.55)	(0.55)	[0.90]	[0.98]	[0.00]	[0.00]
	Annual trend	0.02	0.03	0.02	0.02	0.04	-0.00	0.00	0.02	-0.01
		(0.00)	(0.01)	(0.00)	(0.01)	(0.01)	[0.83]	[0.66]	[0.02]	[0.68]
Any antenatal visits	2001-2005 avg	0.95	0.61	0.63	0.95	0.95	0.35	0.32	0.00	0.00
		(0.23)	(0.59)	(0.58)	(0.23)	(0.23)	[0.00]	[0.00]	[0.83]	[0.82]
	Annual trend	0.00	0.06	-0.01	0.01	0.01	-0.06	0.01	0.00	0.00
		(0.00)	(0.01)	(0.01)	(0.00)	(0.00)	[0.00]	[0.10]	[0.98]	[0.81]
Four antenatal visits	2001-2005 avg	0.12	0.36	0.37	0.59	0.67	-0.23	-0.25	-0.45	-0.53
		(0.36)	(0.57)	(0.58)	(0.55)	(0.53)	[0.00]	[0.00]	[0.00]	[0.00]
	Annual trend	0.00	0.05	-0.03	-0.02	-0.02	-0.04	0.03	0.04	0.04
		(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	[0.00]	[0.00]	[0.00]	[0.00]
Antenatal tetanus shot	2001-2005 avg	0.64	0.57	0.60	0.83	0.80	0.07	0.04	-0.18	-0.14
		(0.53)	(0.59)	(0.59)	(0.42)	(0.45)	[0.00]	[0.00]	[0.00]	[0.00]
	Annual trend	0.01	0.04	-0.02	-0.01	-0.02	-0.03	0.03	0.02	0.03
		(0.01)	(0.01)	(0.01)	(0.00)	(0.01)	[0.02]	[0.00]	[0.00]	[0.01]
Birth order	2001-2005 avg	3.93	3.84	3.86	3.52	3.36	0.09	0.07	0.38	0.54
		(2.60)	(2.88)	(2.88)	(2.65)	(2.50)	[0.27]	[0.18]	[0.00]	[0.00]
	Annual trend	-0.05	-0.19	-0.06	-0.02	0.14	0.14	0.01	-0.05	-0.21
		(0.03)	(0.05)	(0.03)	(0.04)	(0.07)	[0.01]	[0.88]	[0.25]	[0.00]
Age under 20	2001-2005 avg	0.05	0.15	0.16	0.15	0.16	-0.09	-0.11	-0.10	-0.12
		(0.23)	(0.39)	(0.41)	(0.39)	(0.41)	[0.00]	[0.00]	[0.00]	[0.00]
	Annual trend	-0.01	0.01	-0.00	-0.01	-0.02	-0.01	-0.00	0.00	0.02
		(0.00)	(0.01)	(0.00)	(0.00)	(0.01)	[0.08]	[0.38]	[0.53]	[0.17]
Age over 35	2001-2005 avg	0.18	0.13	0.14	0.13	0.11	0.06	0.04	0.05	0.07
		(0.41)	(0.37)	(0.39)	(0.38)	(0.36)	[0.00]	[0.00]	[0.00]	[0.00]
	Annual trend	-0.01	-0.02	-0.01	-0.00	0.01	0.01	0.00	-0.01	-0.02
		(0.00)	(0.01)	(0.00)	(0.00)	(0.01)	[0.28]	[0.91]	[0.38]	[0.08]
Primary education	2001-2005 avg	0.69	0.39	0.50	0.79	0.82	0.30	0.19	-0.09	-0.12
		(0.49)	(0.55)	(0.56)	(0.45)	(0.43)	[0.00]	[0.00]	[0.00]	[0.00]
	Annual trend	0.01	0.02	0.01	0.01	0.02	-0.01	-0.00	0.00	-0.00
		(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	[0.57]	[0.65]	[0.81]	[0.70]
Household size	2001-2005 avg	5.63	6.37	6.10	5.71	5.63	-0.75	-0.47	-0.11	-0.03
		(2.08)	(2.85)	(2.60)	(3.08)	(2.88)	[0.00]	[0.00]	[0.02]	[0.77]
	Annual trend	-0.07	-0.02	-0.00	0.03	0.18	-0.05	-0.07	-0.10	-0.25
		(0.02)	(0.04)	(0.02)	(0.03)	(0.06)	[0.27]	[0.02]	[0.01]	[0.00]
Health insurance, household	2001-2005 avg	0.57	0.01	0.01	0.02	0.07	0.55	0.55	0.57	0.52
		(0.52)	(0.13)	(0.12)	(0.16)	(0.29)	[0.00]	[0.00]	[0.00]	[0.00]
	Annual trend	0.03	0.01	0.00	0.01	-0.00	0.02	0.03	0.04	0.06
		(0.01)	(0.00)	(0.00)	(0.00)	(0.01)	[0.00]	[0.00]	[0.00]	[0.00]
Urban	2001-2005 avg	0.08	0.16	0.09	0.21	0.16	-0.08	-0.02	-0.14	-0.09
		(0.28)	(0.41)	(0.32)	(0.46)	(0.41)	[0.00]	[0.00]	[0.00]	[0.00]
	Annual trend	0.00	0.01	0.01	0.00	0.00	-0.01	-0.01	-0.00	-0.00
		(0.00)	(0.01)	(0.00)	(0.00)	(0.01)	[0.29]	[0.03]	[0.50]	[0.55]
Has comm. health worker	2001-2005 avg	0.58	0.76	0.82	0.89	0.86	-0.17	-0.24	-0.28	-0.26
		(0.52)	(0.48)	(0.43)	(0.36)	(0.38)	[0.00]	[0.00]	[0.00]	[0.00]
	Annual trend	0.01	-0.01	0.01	0.01	0.00	0.02	-0.00	0.03	0.03
		(0.01)	(0.01)	(0.00)	(0.00)	(0.01)	[0.05]	[0.55]	[0.00]	[0.00]
Num. of unsatisfied basic needs†	2001-2005 avg	2.44	2.51	2.39	2.21	2.15	-0.07	0.05	0.17	0.22
		(0.89)	(1.17)	(1.00)	(1.11)	(1.26)	[0.02]	[0.00]	[0.00]	[0.00]
	Annual trend	-0.07	-0.13	-0.04	-0.02	-0.01	0.06	-0.03	-0.10	-0.10
		(0.01)	(0.02)	(0.01)	(0.01)	(0.03)	[0.00]	[0.02]	[0.00]	[0.00]
Num. of durable assets†	2001-2005 avg	0.07	0.22	0.20	0.39	0.39	-0.16	-0.14	-0.30	-0.30
		(0.36)	(0.70)	(0.67)	(1.01)	(1.03)	[0.00]	[0.00]	[0.00]	[0.00]
	Annual trend	0.01	0.11	0.06	0.10	0.10	-0.09	-0.04	-0.06	-0.07
		(0.00)	(0.01)	(0.01)	(0.01)	(0.02)	[0.00]	[0.00]	[0.00]	[0.00]

Summary statistics: Continued from previous page

Outcome/covariate	Year	Group averages Mean / (s.e.)					Difference from Rwanda Diff / [p-val]			
		Rwanda	Lagged four antenatal	Covars four antenatal	Lagged antenatal tetanus	Covars antenatal tetanus	Lagged four antenatal	Covars four antenatal	Lagged antenatal tetanus	Covars antenatal tetanus
Institutional deliveries	2001-2005 avg	0.26	0.09	0.12	0.42	0.49	0.19	0.16	-0.15	-0.21
		(0.46)	(0.32)	(0.36)	(0.55)	(0.56)	[0.00]	[0.00]	[0.00]	[0.00]
	Annual trend	0.02	0.01	0.00	0.02	0.01	0.03	0.03	0.02	0.02
		(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	[0.00]	[0.00]	[0.01]	[0.07]
Any antenatal visits	2001-2005 avg	0.95	0.26	0.28	0.71	0.66	0.69	0.67	0.24	0.30
		(0.23)	(0.49)	(0.50)	(0.51)	(0.53)	[0.00]	[0.00]	[0.00]	[0.00]
	Annual trend	0.00	0.04	0.02	0.02	0.00	-0.03	-0.02	-0.01	0.00
		(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	[0.00]	[0.06]	[0.15]	[0.79]
Four antenatal visits	2001-2005 avg	0.12	0.14	0.14	0.46	0.47	-0.00	-0.00	-0.33	-0.33
		(0.36)	(0.38)	(0.38)	(0.56)	(0.56)	[0.89]	[0.83]	[0.00]	[0.00]
	Annual trend	0.00	0.02	0.01	-0.00	-0.00	-0.00	0.01	0.02	0.02
		(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	[0.97]	[0.23]	[0.00]	[0.10]
Antenatal tetanus shot	2001-2005 avg	0.64	0.27	0.34	0.65	0.65	0.38	0.31	0.01	0.01
		(0.53)	(0.50)	(0.53)	(0.53)	(0.53)	[0.00]	[0.00]	[0.62]	[0.72]
	Annual trend	0.01	0.03	0.00	0.01	0.01	-0.02	0.01	0.00	0.00
		(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	[0.14]	[0.37]	[0.87]	[0.84]
Birth order	2001-2005 avg	3.93	4.48	4.15	3.80	3.41	-0.57	-0.25	0.11	0.49
		(2.60)	(3.04)	(3.00)	(2.83)	(2.67)	[0.00]	[0.00]	[0.03]	[0.00]
	Annual trend	-0.05	-0.17	-0.17	-0.11	-0.14	0.10	0.10	0.04	0.07
		(0.03)	(0.07)	(0.06)	(0.03)	(0.05)	[0.18]	[0.11]	[0.31]	[0.20]
Age under 20	2001-2005 avg	0.05	0.11	0.13	0.12	0.10	-0.07	-0.09	-0.08	-0.05
		(0.23)	(0.36)	(0.38)	(0.37)	(0.33)	[0.00]	[0.00]	[0.00]	[0.00]
	Annual trend	-0.01	-0.02	-0.01	-0.01	-0.01	0.01	0.01	0.01	0.01
		(0.00)	(0.01)	(0.01)	(0.00)	(0.01)	[0.08]	[0.48]	[0.19]	[0.13]
Age over 35	2001-2005 avg	0.18	0.15	0.17	0.14	0.12	0.03	0.01	0.04	0.06
		(0.41)	(0.40)	(0.42)	(0.39)	(0.36)	[0.04]	[0.30]	[0.00]	[0.00]
	Annual trend	-0.01	-0.02	-0.01	-0.01	-0.01	0.01	0.01	0.01	0.00
		(0.00)	(0.01)	(0.01)	(0.00)	(0.01)	[0.20]	[0.41]	[0.28]	[0.84]
Primary education	2001-2005 avg	0.69	0.12	0.20	0.56	0.63	0.57	0.49	0.13	0.06
		(0.49)	(0.37)	(0.45)	(0.55)	(0.54)	[0.00]	[0.00]	[0.00]	[0.00]
	Annual trend	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.00	0.01
		(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	[0.48]	[0.30]	[0.57]	[0.49]
Household size	2001-2005 avg	5.63	5.97	5.83	5.97	5.49	-0.37	-0.23	-0.37	0.11
		(2.08)	(2.53)	(2.42)	(3.02)	(2.82)	[0.00]	[0.00]	[0.00]	[0.23]
	Annual trend	-0.07	-0.07	-0.03	-0.03	-0.11	0.00	-0.04	-0.04	0.04
		(0.02)	(0.06)	(0.05)	(0.02)	(0.06)	[0.95]	[0.44]	[0.22]	[0.48]
Health insurance, household	2001-2005 avg	0.57	0.01	0.00	0.02	0.08	0.58	0.59	0.57	0.51
		(0.52)	(0.09)	(0.06)	(0.16)	(0.30)	[0.00]	[0.00]	[0.00]	[0.00]
	Annual trend	0.03	0.01	0.00	0.01	0.03	0.05	0.05	0.05	0.02
		(0.01)	(0.00)	(0.00)	(0.00)	(0.01)	[0.00]	[0.00]	[0.00]	[0.05]
Urban	2001-2005 avg	0.08	0.16	0.11	0.23	0.43	-0.08	-0.03	-0.16	-0.36
		(0.28)	(0.41)	(0.34)	(0.47)	(0.55)	[0.00]	[0.00]	[0.00]	[0.00]
	Annual trend	0.00	0.01	0.02	0.01	0.00	-0.01	-0.02	-0.01	0.00
		(0.00)	(0.01)	(0.01)	(0.00)	(0.01)	[0.38]	[0.02]	[0.28]	[0.99]
Has comm. health worker	2001-2005 avg	0.58	0.45	0.59	0.74	0.61	0.16	0.01	-0.13	-0.01
		(0.52)	(0.56)	(0.55)	(0.49)	(0.54)	[0.00]	[0.39]	[0.00]	[0.73]
	Annual trend	0.01	0.00	0.02	0.00	-0.01	0.04	0.02	0.03	0.04
		(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	[0.01]	[0.17]	[0.00]	[0.00]
Num. of unsatisfied basic needs†	2001-2005 avg	2.44	2.71	2.47	2.19	1.74	-0.33	-0.10	0.19	0.64
		(0.89)	(1.04)	(0.96)	(1.25)	(1.49)	[0.00]	[0.00]	[0.00]	[0.00]
	Annual trend	-0.07	-0.07	-0.00	-0.03	-0.05	-0.04	-0.11	-0.09	-0.07
		(0.01)	(0.03)	(0.02)	(0.01)	(0.03)	[0.10]	[0.00]	[0.00]	[0.05]
Num. of durable assets†	2001-2005 avg	0.07	0.14	0.13	0.44	0.75	-0.04	-0.04	-0.35	-0.65
		(0.36)	(0.62)	(0.58)	(1.09)	(1.35)	[0.03]	[0.03]	[0.00]	[0.00]
	Annual trend	0.01	0.03	0.02	0.07	0.14	0.01	0.01	-0.04	-0.10
		(0.00)	(0.02)	(0.02)	(0.01)	(0.03)	[0.66]	[0.46]	[0.00]	[0.00]

Note: Annual trends are estimated using linear regressions; standard errors are shown in parenthesis. Statistics are weighted. Within each year, weights for Rwanda's arm 1 and arm 2 areas sum to one, separately. For the full sample control group, weights within each control region sum to one within each year. For the synthetic control group, weights for all synthetic control observations sum to one within each year.

†Total unsatisfied basic needs and total durable assets are shown for conciseness for the covariate predictors. Instead, the analysis includes indicators for one, two, three, and four unsatisfied basic needs, and indicators for ownership of televisions, landline telephones, cars/trucks, and refrigerators.

B.3 Additional results, robustness with alternate control groups

Table B.5: Effects of incentives versus unconditional finance, alternate controls, comparison with RCT (percentage points)

	(1)	(2)	(3)	(4)	(5)
	2005, 2007			2001-2008	
	Basinga RCT	Lagged outcome	Covars outcome	Lagged outcome	Covars outcome
Institutional deliveries					
Arm 1 x post 2006 (incentives vs. uncond. finance)	8.1**	8.3**	8.6**	8.0***	8.1***
	[$p = 0.02$]	(3.5)	(3.7)	(2.4)	(2.5)
Observations	2,108	42,834	6,083	137,247	21,830
2005 mean (Rwanda arm 1)	35	32.2	32.2	32.2	32.2
Any antenatal visits					
Arm 1 x post 2006 (incentives vs. uncond. finance)	0.2	-2.9*	-2.9*	-0.1	-0.1
	[$p = 0.88$]	(1.5)	(1.5)	(0.9)	(0.9)
Observations	2,309	28,654	2,971	87,662	10,751
2005 mean (Rwanda arm 1)	95	98.2	98.2	98.2	98.2
Four antenatal visits					
Arm 1 x post 2006 (incentives vs. uncond. finance)	0.8	-3.2	-4.0	1.0	1.0
	[$p = 0.83$]	(3.2)	(3.1)	(3.4)	(3.4)
Observations	2,223	2,809	3,277	9,757	11,428
2005 mean (Rwanda arm 1)	18	22.9	22.9	22.9	22.9
Antenatal tetanus shot					
Arm 1 x post 2006 (incentives vs. uncond. finance)	5.1*	-2.3	-3.2	4.7*	4.3*
	[$p = 0.06$]	(4.7)	(5.0)	(2.5)	(2.5)
Observations	2,856	29,322	3,143	89,262	11,172
2005 mean (Rwanda arm 1)	71	70.7	70.7	70.7	70.7

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, region fixed effects, and month-year fixed effects. Standard errors are clustered by region and are shown in parenthesis. Weights are applied as described in the text and Table B.3. The maternal synthetic control is matched according to the maternal service rate, while the outcome synthetic control is matched by each outcome, e.g., matching deliveries for the deliveries regression. The lagged specification uses outcomes as predictors, while the covariate specification uses outcomes and covariates as predictors.

Table B.6: Institutional deliveries: Differential effects by arm and phase, excluding each country

	(1) All controls	(2) Without Benin	(3) Without Ethiopia	(4) Without Kenya	(5) Without Lesotho	(6) Without Malawi	(7) Without Nigeria	(8) Without Tanzania	(9) Without Uganda	(10) Without Zimbabwe
β_1: Arm 1 x post 2006 (initial incentive effect)	23.9***	24.4***	23.3***	23.9***	23.6***	25.2***	23.5***	23.7***	24.5***	22.8***
	(1.4)	(1.4)	(1.5)	(1.5)	(1.5)	(1.4)	(1.4)	(1.4)	(1.4)	(1.4)
β_2: Arm 2 x post 2006 (uncond. finance effect)	15.7***	16.1***	15.0***	15.7***	15.4***	17.0***	15.2***	15.5***	16.3***	14.6***
	(2.4)	(2.4)	(2.5)	(2.4)	(2.5)	(2.4)	(2.4)	(2.5)	(2.4)	(2.4)
β_3: Arm 1 x post 2008 (med. run incentive effect)	12.2***	12.2***	12.0***	12.3***	12.7***	13.1***	11.9***	12.2***	12.4***	11.4***
	(1.6)	(1.6)	(1.6)	(1.6)	(1.6)	(1.6)	(1.6)	(1.6)	(1.6)	(1.6)
β_4: Arm 2 x post 2008 (scale-up effect)	16.8***	16.9***	16.6***	16.9***	17.1***	17.8***	16.5***	16.7***	17.0***	16.0***
	(2.0)	(2.0)	(2.0)	(2.0)	(2.0)	(2.0)	(2.0)	(2.0)	(2.0)	(2.0)
Observations	195,266	168,820	174,760	181,578	187,915	166,760	150,070	180,472	180,764	184,956
2005 mean (full sample)	52.4	50.4	57.4	53.0	52.3	49.2	53.2	52.5	52.9	50.3
Hypothesis testing										
β_1 - $\beta_2 = 0$ (incentives vs. uncond. financing)	8.2	8.2	8.3	8.2	8.2	8.2	8.2	8.3	8.2	8.3
p-value	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]
$\beta_1 + \beta_3 = 0$ (med. run incentives, total)	36.1	36.5	35.3	36.2	36.3	38.3	35.4	35.9	36.9	34.3
p-value	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]
$\beta_2 + \beta_4 = 0$ (med. run arm 2, total)	32.5	33.0	31.6	32.6	32.6	34.7	31.7	32.2	33.3	30.6
p-value	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]
$(\beta_1 + \beta_3)$ - $(\beta_2 + \beta_4) = 0$ (arm 1 vs. 2, post 2008)	3.6	3.5	3.7	3.6	3.7	3.6	3.7	3.7	3.6	3.7
p-value	[0.19]	[0.20]	[0.19]	[0.19]	[0.18]	[0.20]	[0.18]	[0.19]	[0.19]	[0.18]

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, region fixed effects, and month-year fixed effects. Standard errors are clustered by region and are shown in parenthesis. Weights are applied as described in the text.

Table B.7: Any antenatal visits: Differential effects by arm and phase, excluding each country

	(1) All controls	(2) Without Benin	(3) Without Ethiopia	(4) Without Kenya	(5) Without Lesotho	(6) Without Malawi	(7) Without Nigeria	(8) Without Tanzania	(9) Without Uganda	(10) Without Zimbabwe
β_1: Arm 1 x post 2006 (initial incentive effect)	-1.4	-1.5	0.6	-1.5	-1.7	-1.3	-1.5	-1.7	-1.8*	-2.0*
	(1.0)	(1.0)	(0.7)	(1.1)	(1.0)	(1.1)	(1.0)	(1.0)	(1.0)	(1.1)
β_2: Arm 2 x post 2006 (uncond. finance effect)	-1.2	-1.3	0.8	-1.2	-1.5	-1.0	-1.3	-1.5	-1.5	-1.8*
	(1.0)	(1.0)	(0.8)	(1.0)	(1.0)	(1.0)	(1.0)	(1.0)	(1.0)	(1.0)
β_3: Arm 1 x post 2008 (med. run incentive effect)	0.7	0.7	0.8	0.6	1.0	0.9	1.1	0.7	0.7	0.1
	(0.7)	(0.7)	(0.7)	(0.8)	(0.8)	(0.8)	(0.7)	(0.8)	(0.8)	(0.7)
β_4: Arm 2 x post 2008 (scale-up effect)	-1.9**	-1.9**	-1.7**	-2.1**	-1.7*	-1.8*	-1.6*	-1.9**	-2.0**	-2.6***
	(0.9)	(0.9)	(0.7)	(0.9)	(0.9)	(0.9)	(0.8)	(0.9)	(0.9)	(0.9)
Observations	127,945	110,830	113,992	118,760	122,152	107,917	102,712	117,589	119,141	119,830
2005 mean (full sample)	85.9	85.8	92.6	85.4	84.8	83.5	87.2	84.5	84.6	84.7
Hypothesis testing										
β_1 - $\beta_2 = 0$ (incentives vs. uncond. financing)	-0.2	-0.2	-0.2	-0.3	-0.2	-0.3	-0.3	-0.3	-0.3	-0.3
p-value	[0.79]	[0.80]	[0.85]	[0.76]	[0.83]	[0.77]	[0.78]	[0.78]	[0.76]	[0.79]
$\beta_1 + \beta_3 = 0$ (med. run incentives, total)	-0.7	-0.8	1.5	-0.9	-0.7	-0.4	-0.5	-1.0	-1.2	-1.9
p-value	[0.56]	[0.52]	[0.09]	[0.50]	[0.57]	[0.74]	[0.70]	[0.41]	[0.35]	[0.12]
$\beta_2 + \beta_4 = 0$ (med. run arm 2, total)	-3.1	-3.2	-0.9	-3.3	-3.2	-2.9	-2.9	-3.4	-3.5	-4.4
p-value	[0.04]	[0.04]	[0.47]	[0.05]	[0.05]	[0.08]	[0.06]	[0.03]	[0.03]	[0.01]
$(\beta_1 + \beta_3)$ - $(\beta_2 + \beta_4) = 0$ (arm 1 vs. 2, post 2008)	2.4	2.4	2.4	2.4	2.5	2.4	2.5	2.4	2.4	2.5
p-value	[0.09]	[0.08]	[0.05]	[0.10]	[0.09]	[0.10]	[0.07]	[0.10]	[0.10]	[0.09]

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, region fixed effects, and month-year fixed effects. Standard errors are clustered by region and are shown in parenthesis. Weights are applied as described in the text.

Table B.8: Four antenatal visits: Differential effects by arm and phase, excluding each country

	(1) All controls	(2) Without Benin	(3) Without Ethiopia	(4) Without Kenya	(5) Without Lesotho	(6) Without Malawi	(7) Without Nigeria	(8) Without Tanzania	(9) Without Uganda	(10) Without Zimbabwe
β_1: Arm 1 x post 2006 (initial incentive effect)	15.3***	15.3***	16.8***	15.6***	15.3***	14.6***	15.4***	13.9***	15.4***	15.6***
	(2.3)	(2.3)	(2.3)	(2.3)	(2.3)	(2.3)	(2.3)	(2.2)	(2.3)	(2.3)
β_2: Arm 2 x post 2006 (uncond. finance effect)	14.8***	14.8***	16.3***	15.2***	14.8***	14.1***	14.9***	13.3***	14.9***	15.2***
	(2.6)	(2.7)	(2.7)	(2.7)	(2.7)	(2.7)	(2.7)	(2.6)	(2.7)	(2.7)
β_3: Arm 1 x post 2008 (med. run incentive effect)	4.5*	4.9*	4.0	5.1*	4.8*	3.5	4.9*	4.5*	5.1*	3.3
	(2.6)	(2.6)	(2.6)	(2.6)	(2.6)	(2.7)	(2.6)	(2.6)	(2.6)	(2.6)
β_4: Arm 2 x post 2008 (scale-up effect)	6.8***	7.2***	6.3***	7.2***	7.2***	5.9***	7.1***	6.9***	7.3***	5.5***
	(1.6)	(1.6)	(1.7)	(1.6)	(1.7)	(1.7)	(1.7)	(1.7)	(1.7)	(1.6)
Observations	127,945	110,830	113,992	118,760	122,152	107,917	102,712	117,589	119,141	119,830
2005 mean (full sample)	53.7	53.4	58.1	53.8	50.6	53.8	53.4	53.6	54.2	52.1
Hypothesis testing										
β_1 - β_2 = 0 (incentives vs. uncond. financing)	0.5	0.5	0.5	0.4	0.5	0.5	0.5	0.6	0.5	0.4
p-value	[0.88]	[0.88]	[0.88]	[0.90]	[0.86]	[0.87]	[0.89]	[0.86]	[0.88]	[0.90]
$\beta_1 + \beta_3 = 0$ (med. run incentives, total)	19.8	20.2	20.8	20.7	20.2	18.1	20.2	18.4	20.5	18.8
p-value	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]
$\beta_2 + \beta_4 = 0$ (med. run arm 2, total)	21.6	22.0	22.6	22.4	22.0	19.9	22.0	20.2	22.2	20.7
p-value	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]
$(\beta_1 + \beta_3)$ - $(\beta_2 + \beta_4) = 0$ (arm 1 vs. 2, post 2008)	-1.8	-1.8	-1.8	-1.7	-1.8	-1.8	-1.8	-1.8	-1.8	-1.8
p-value	[0.55]	[0.56]	[0.55]	[0.57]	[0.55]	[0.55]	[0.56]	[0.54]	[0.56]	[0.54]

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, region fixed effects, and month-year fixed effects. Standard errors are clustered by region and are shown in parenthesis. Weights are applied as described in the text.

Table B.9: Antenatal tetanus prophylaxis: Differential effects by arm and phase, excluding each country

	(1) All controls	(2) Without Benin	(3) Without Ethiopia	(4) Without Kenya	(5) Without Lesotho	(6) Without Malawi	(7) Without Nigeria	(8) Without Tanzania	(9) Without Uganda	(10) Without Zimbabwe
β_1: Arm 1 x post 2006 (initial incentive effect)	6.3***	6.2***	8.3***	6.2***	5.7***	6.4***	5.9***	5.0***	6.8***	6.2***
	(1.9)	(1.9)	(1.7)	(1.9)	(1.9)	(1.9)	(1.9)	(1.8)	(1.9)	(1.9)
β_2: Arm 2 x post 2006 (uncond. finance effect)	1.4	1.3	3.3*	1.3	0.7	1.5	1.1	0.0	1.9	1.3
	(2.0)	(2.0)	(2.0)	(2.0)	(2.0)	(2.0)	(2.0)	(2.0)	(2.0)	(2.0)
β_3: Arm 1 x post 2008 (med. run incentive effect)	2.4	2.5	2.2	2.4	3.2**	1.8	3.0*	2.6	2.2	1.9
	(1.6)	(1.6)	(1.6)	(1.6)	(1.6)	(1.7)	(1.6)	(1.6)	(1.6)	(1.6)
β_4: Arm 2 x post 2008 (scale-up effect)	4.2*	4.3*	4.3*	4.2*	4.9**	3.4	4.8**	4.4*	4.0	3.7
	(2.4)	(2.5)	(2.3)	(2.5)	(2.4)	(2.5)	(2.4)	(2.4)	(2.5)	(2.4)
Observations	127,480	110,453	113,675	119,741	121,712	107,406	101,023	117,125	118,586	119,419
2005 mean (full sample)	75.2	75.3	80.0	73.7	75.0	72.6	76.1	74.9	75.2	74.2
Hypothesis testing										
β_1 - $\beta_2 = 0$ (incentives vs. uncond. financing)	4.9	4.9	5.0	4.8	5.0	4.9	4.9	5.0	4.9	4.9
p-value	[0.04]	[0.04]	[0.04]	[0.05]	[0.04]	[0.04]	[0.04]	[0.04]	[0.04]	[0.04]
$\beta_1 + \beta_3 = 0$ (med. run incentives, total)	8.7	8.7	10.6	8.6	8.9	8.1	9.0	7.6	9.0	8.1
p-value	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]
$\beta_2 + \beta_4 = 0$ (med. run arm 2, total)	5.6	5.6	7.6	5.6	5.7	4.8	5.9	4.5	5.9	5.0
p-value	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.01]	[0.00]	[0.00]	[0.00]	[0.00]
$(\beta_1 + \beta_3)$ - $(\beta_2 + \beta_4) = 0$ (arm 1 vs. 2, post 2008)	3.1	3.1	3.0	3.0	3.2	3.3	3.1	3.2	3.1	3.2
p-value	[0.17]	[0.18]	[0.19]	[0.19]	[0.16]	[0.17]	[0.17]	[0.17]	[0.17]	[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, region fixed effects, and month-year fixed effects. Standard errors are clustered by region and are shown in parenthesis. Weights are applied as described in the text.