WPS8162

Policy Research Working Paper 8162

Looking into the Performance-Based Financing Black Box

Evidence from an Impact Evaluation in the Health Sector in Cameroon

Damien de Walque
Paul Jacob Robyn
Hamadou Saidou
Gaston Sorgho
Maria Steenland

WORLD BANK GROUP
Development Research Group
Human Development and Public Services Team

Health Nutrition and Population Global Practice Group
August 2017

Policy Research Working Paper 8162

Abstract

Performance-based financing is a complex health system intervention aimed at improving coverage and quality of care. This paper presents the results of an impact evaluation in Cameroon that seeks to isolate the role of specific components of the performance-based financing approach on outcomes of interest, such as explicit financial incentives linked to results, additional resources available at the point of service delivery (not linked to performance), and enhanced supervision, coaching, and monitoring. Four evaluation groups were established to measure the effects of each component that was studied. In general, the results indicate that performance-based financing in Cameroon is an efficient mechanism to channel payments and funding to the provider level, leading to significant increases in utilization in the performance-based financing arm for several services (child and maternal vaccinations and use of modern family planning), but not for others, such as antenatal care visits and facility-based deliveries. However, for many of those outcomes, the differences between the performance-based financing group and the additional financing group are not significant. In terms of quality, performance-based financing was found to have a significant impact on the availability of essential inputs and equipment, qualified health workers, reduction in formal and informal user fees, and increased satisfaction among patients and providers. However, there was a clear effect of additional financing, irrespective of whether it was linked to incentives, in combination with reinforced supervision through performance-based financing. This result suggests that enhanced supervision and monitoring on their own are not sufficient to improve maternal and child health outcomes.

This paper is a product of the Human Development and Public Services Team, Development Research Group and the Health Nutrition and Population Global Practice Group. It is part of a larger effort by the World Bank to provide open access to its research and make a contribution to development policy discussions around the world. Policy Research Working Papers are also posted on the Web at http://econ.worldbank.org. The authors may be contacted at ddewalque@worldbank.org, probyn@worldbank.org, saidoutheo@yahoo.fr, gsorgho@worldbank.org, and mws475@mail.harvard.edu.

Looking into the Performance-Based Financing Black Box: Evidence from an Impact Evaluation in the Health Sector in Cameroon*

Damien de Walque, Development Research Group, The World Bank

Paul Jacob Robyn, Health Nutrition and Population, The World Bank

Hamadou Saidou, Health Nutrition and Population, The World Bank and University of Paris Dauphine

Gaston Sorgho, Health Nutrition and Population, The World Bank

Maria Steenland, Department of Global Health and Population
Harvard School of Public Health

Keywords: Performance-based Financing; Health; Africa

JEL classification: I15 J13; O15.

* We are extremely grateful for a fruitful collaboration with the Ministry of Health of Cameroon and in particular with his Excellency Andre Mama Fouda, Minister, Enandjoum Bwanga, Emannuel Maina Djoulde, Victor Ndiforchu, and all the Regional Health Delegates, Performance Purchasing Agency Managers and staff and from the North-West, South-West and East Regions. The baseline and endline data were collected by the Institut de Formation et de Recherche Démographiques (IFORD) with special thanks to Gervais Beninguisse, Didier Nganawara, Evina Akam and all staff implementing the baseline and endline surveys. This impact evaluation is funded by the Health Results Innovation Trust Funds (HRITF) at the World Bank. We are grateful to Emanuela Di Gropello, Guenter Fink, Gyuri Fritsche, Elisabeth Huybens, Gil Shapira, Jean Claude Taptue Fotso, Robert Soeters and Omer Zhang for useful comments and discussions. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the World Bank, its Executive Directors, or the countries they represent.

1. Introduction

Confronted with slow progress on the health-related Millennium Development Goals (United Nations Statistical Division, 2015) and the Sustainable Development Goals, various countries have introduced and experimented with Results-Based Financing (RBF) in the health sector. RBF is an approach that aims to improve health systems and prioritize health outcomes by using financial incentives paid after predefined results have been attained and verified. Among the RBF approaches, Performance-Based Financing (PBF) is a specific supply-side intervention which comprises a set of health system reforms meant to increase the coverage and quality of essential health services, as well as efficiency and equity, often with a special focus on maternal and child health. PBF program models differ but all involve the purchasing of health services using a pre-defined list of services and prices (Fritsche, Soeters, & Meessen, 2014). Performance-based financing also includes a strong verification system that relies on systematic and detailed review of health facility records as well as community-level client tracing whereby reported patients are asked a series of questions to confirm health care receipt (Fritsche et al., 2014). Many PBF programs also involve increasing health facility autonomy.

Over time, PBF has been implemented in a growing number of countries. Several studies have shown a positive impact of PBF on health service coverage, often coupled with improvements in quality. An early impact evaluation in Rwanda where districts were randomly assigned to treatment (PBF) and comparison (input financing with matched financial resources) found large and statistically significant positive impacts on institutional deliveries and preventive care visits from young children and also on quality of prenatal care (Basinga et al., 2010). After this promising start, many other quasi-experimental studies have shown similarly positive results (Gertler & Giovagnoli, 2014; Ir et al., 2015; Zeng, Cros, Wright, & Shepard, 2013) and several others have shown favorable results for many – though not all – outcomes assessed (Basinga et al., 2010; Binyaruka et al., 2015; Bonfrer, Soeters, et al., 2014; Bonfrer, Van de Poel, & Van Doorslaer, 2014; Falisse, Ndayishimiye, Kamenyero, & Bossuyt, 2015). Despite this, other studies have found only limited positive results and the research community has not reached a consensus about the effectiveness of PBF at increasing health service coverage (Huillery & Seban, 2014).

While the evidence about the impacts of PBF accumulates, few studies have examined the factors and mechanisms that influence its impact, an area of substantial theoretical and practical significance since PBF often involves a package of interventions: linking payment and results, independent verification of results, managerial autonomy to facilities and enhanced systematic supervision and coaching of facilities. We designed the impact evaluation of the PBF package in Cameroon to try to understand better the role of some of these mechanisms. In particular, we tried to isolate the role of explicit financial incentives as opposed to additional funding not linked to performance, as well as separating the impact of enhanced supervision and monitoring. Specifically, the evaluation compared four arms: (1) the standard PBF package (T1), (2) the same level of financing as T1 but not linked to performance, and with the same levels of supervision, monitoring, and autonomy as PBF (C1), (3) no additional resources or autonomy, but the same levels of supervision and monitoring as PBF (C2), and (4) pure comparison (C3).

In general, the results indicate that PBF in Cameroon is an efficient mechanism to bring payments and funding at the provider level, leading to significant increases in utilization (child and maternal immunization, family planning, HIV testing) and improvements in structural quality of care. However, for many of those outcomes, the differences between the PBF group (T1) and the additional financing group (C1) are not significant. It should be noted that the C1 group offered all the elements of PBF except the direct link between individual facility performance and additional financing. There was, however, a clear effect of additional financing, irrespective of incentives, plus reinforced supervision through PBF instruments (comparing groups T1 and C1 with the C2 group and then C3, the control group), suggesting that enhanced supervision and monitoring are not sufficient to improve maternal and child health (MCH) outcomes.

We note, however, that we did not identify an impact for some MCH indicators such as skilled deliveries and ANC visits. It is possible that the supply-side incentives for providers were not sufficient given existing user fees which might act as a barrier on the demand-side. In addition, given that these types of services are primarily provided at the health center, outreach and community-based service delivery was not an option for providers to implement to increase coverage. In terms of quality of care, most of the positive impacts were observed on structural quality, presence of qualified staff, and provider and patient satisfaction. However, despite an increase in providers and supplies available at health facilities, PBF did not increase the completeness of service provision (content of care) during antenatal care and child health consultations. But PBF led to a decrease in out-of-pocket payments, in particular unofficial payments.

Section 2 of this paper presents the context and program design of PBF in Cameroon, sections 3 and 4 respectively describe the methods and results while section 5 discusses study limitations and the policy implications of the findings.

2. Context and Design

HEALTH BACKGROUND

Despite being one of the more wealthy countries in the Central Africa region, and the country's relatively high health spending of $59 per capita in 2014 (World Health Organization Global Health Expenditure database, 2016), Cameroon's health indicators resemble countries that spend much less on health care (The World Bank, 2016). Cameroon did not achieve Millennium Development Goals 4 & 5 which called for large reductions in maternal and child mortality. Moreover, Cameroon is one of the few countries with high maternal mortality where maternal mortality did not decrease at all between 1990 and 2015 (Alkema et al., 2015). One in every 25 women of reproductive age in Cameroon continues to die from pregnancy-related causes (Institut National de la Statistique (INS) et ICF International, 2012). Though many factors affect maternal mortality, one potential explanation for the country's slow progress in achieving declines in maternal mortality is that the proportion of women delivering in health facilities has remained unchanged over much of the last decade. During the last ten years (between 2004 and 2014), the percentage of deliveries that were assisted by a skilled health professional increased from 61.7 percent to only 64.7 percent (ICF International, 2012; Institut National de la Statistique, 2015).

Like skilled delivery, the percentage of women receiving four or more antenatal care visits, another health service with the potential to decrease maternal and child mortality, remained unchanged over the last decade. In 2004, 59.1% of women had received four antenatal visits, the number recommended by the World Health Organization, compared to 58.8% in 2014 (Institut National de la Statistique, 2015). Even when women obtain antenatal services, they still may not receive important service components due to low quality of care. Previous studies have shown that fewer than half of women are informed about the danger signs to look for during pregnancy, over 40% did not have the recommended number of tetanus vaccinations during their last pregnancy (ICF International, 2012), and 25% did not provide a urine sample to test for protein, a required procedure to test for preeclampsia (The World Bank, 2013).

Child mortality declined in Cameroon by approximately 21% between 1991 and 2014 (ICF International, 2012; Institut National de la Statistique, 2015; Institut National de la Statistique (INS) et ICF International, 2012); nonetheless, according to the most recent data approximately 1 in 10 children still dies before their fifth birthday (Alkema et al., 2015; The United Nations Children's Fund, 2015).

Health financing challenges in Cameroon

Several aspects of the health care financing landscape in Cameroon contribute to the low quality of primary health care service provision, and sub-optimal coverage of essential maternal and child health care services. Cameroon spends $10 dollars more than the average for Sub-Saharan Africa (excluding South Africa) and has similarly high health spending as Senegal and Nigeria (Bove, Basile., Robyn, & Singh, 2013). However, despite this relatively high level of overall spending, the share of government spending on health is low and has not reached above 9 percent of the total budget in the previous 10 years (Bove et al., 2013). Due to these low levels, in 2012 the share of per capita total health spending paid for by the government was only US $14 (i.e. 21.7 percent). Much of the remaining 70.4% of health spending is paid for through out-of-pocket users fees (Ministere de la Sante Publique, 2016). Another important part of the problem is that the operational level receives a small fraction of the health budget. Although the health sector budget has more than doubled in recent years, the majority of these resources have been allocated for administration and infrastructure. The public budget favors central-level administration and tertiary care, with less than 10% of the budget being allocated to service providers and deconcentrated levels of the Ministry of Public Health nationwide at the regional, district and health facility level (The World Bank, 2017). This has resulted in a scarcity of funds to meet operating expenses incurred in the day-to-day business of a district health system (e.g., consumables, drugs, regular maintenance, community outreach, etc.) (Ministere de la Sante Publique, 2016).

PBF timeline in Cameroon

Cameroon’s first experience with PBF began with the Redynamisation des Soins de Sante à l’Est du Cameroun (REDSSEC) project. In 2006, REDSSEC implemented a pilot PBF program in Faith Based Organization (FBO) facilities in the East region with support from Cordaid and Catholic Relief Services. The project began with four FBO facilities in Batouri district, and then expanded to FBO facilities in Bertoua, Doume and Yokadouma districts through 2011 (Appendix Figure A1).

In 2008, the World Bank approved a US$25 million loan to the Government of Cameroon through the Bank's Health Sector Support Investment Project (HSSIP). In 2011, through support from the HSSIP, a PBF pilot began in the Littoral region covering four health districts. The program began in July 2012 in the North West and South West regions, with four districts included in each region. In October 2012, the program expanded to the East region, covering all 14 health districts in the region.

Of the 26 health districts throughout Cameroon implementing PBF, 14 districts were included in the impact evaluation (see Figure 1 and Appendix Figures A2 – A4). The other 12 (four in Littoral and eight in the East) had already begun implementing some form of PBF before the impact evaluation baseline survey was conducted, or were added after the baseline survey was conducted.

Description of the PBF program design in Cameroon

Program overview

The administrative and technical aspects of the PBF program in Cameroon were managed at the regional level by Performance Purchasing Agencies (PPAs). PPAs are autonomous entities that have a contractual relationship with the Government of Cameroon who entrusts the agencies with contracting the health facilities, verification of the data declared by the health facility and management of funding intended for health care providers through PBF. All PBF health facilities signed a Performance Contract issued by the PPA which described conditions required to obtain PBF subsidies. These requirements included efforts to improve management, minimum quality levels, governance and financial inclusion, and clauses for termination of the contract. Additionally, PBF facilities prepared quarterly business plans and used frameworks for the health administration linked to performance payments. All facilities were trained to use the indice tool; however, use of the indice tool varied among Cameroon PBF facilities.

PBF contracts were signed for a period of three months. Health facilities with performance contracts were responsible for completing registers and a monthly activity report/declaration form. This report and registers were used to document reported health service provision and were used as the primary basis for service verification. A list of the health services subsidized by the program, and the subsidy amounts is located in Appendix Table A1. A copy of this report was sent to the PPA each month. After the verification of the quantity of services provided and declared in the monthly report was completed by reviewing health facility records, the bill of the health facility was established and paid monthly. The declarations form verified and validated was used to justify the payment subsidies provided to the health facility.

Service verification

Health service verification was completed on a monthly basis by the PPA verification agents. The verification agents used the facility register and tally sheets to verify that the number of services reported by the health facility in the payment request form was consistent with the facility documentation. If the supervisor encountered any errors, these problems were corrected in the presence of the facility staff, and any fraudulent cases were tracked and documented. As an added means of quantity verification, a sample of patients for a set of health services targeted by the program was contacted either by phone or in person by local PPA staff to confirm that they received the health service reported by the health facility and to assess patient satisfaction. If error rates for a certain indicator surpassed 10-15% (varying slightly by region), the service was not

paid and 25% of the PBF payment to the health facility was retained. Additionally, the District Medical Team in collaboration with the Fund Holder Agency assessed the quality of the health services provided by PBF health facilities. This assessment used a standardized checklist to verify that a minimum quality level is met, and to calculate a quality score for the health facility ranging from 0 – 100%. The quality score was used to calculate a quality bonus that is received by the health facility.[1]

CALCULATION OF PAYMENTS

The validated quantitative data, and the quality assessment were used to calculate performance payments for PBF health facilities. The quality bonus provided an increase of up to 30% of the total payment based on health service quantity. This percentage depended on the health facility quality score. Quality assessments were conducted quarterly, and focused on facility management, hygiene and sanitation, as well as specific attributes of service delivery. The services delivery items included, among many other categories, listing user charges, privacy, the condition of the waiting area and consultation room, and the correct management of cases. For example, if the health facility received a 65% quality score, and their total payment amount based on the services they provided was 597,240 CFA, the quality bonus provided to the health facility would be calculated as follows: 597,240 x 0.30 x 0.65 = 116,461 CFA, and the total payment to the health facility would be 597,240 + 116,461 = 713,701 CFA.

Additionally, an equity bonus was included in the calculation of performance payments. The equity bonus was paid to health facilities that faced serious structural problems making service provision more challenging. Equity bonuses ranged from 0% to 80% of the basic subsidy. Each region applied the criteria listed below differently. North West and South West applied the same scoring; however, few facilities in South West received the equity bonus since almost all facilities in the region were located in urban areas. The East region had a slightly different scoring approach but also used the same criteria.

The following issues were considered in the calculation of this bonus:

- Geographical inaccessibility (hard-to-reach) that makes it difficult to retain staff;
- The size of the health area and low population densities that create viability issues (high running costs)
- Extreme poverty

The facility management committee had the authority to decide on the allocation of PBF revenue. These decisions must have been clearly documented in facility business plans.

RESEARCH QUESTIONS OF THE IMPACT EVALUATION

[1] An example of the quality checklist can be found here: Performance Based Financing Implementation Procedures Manual North-West Region of Cameroon North West Region, Bamenda-Cameroon: Performance Purchasing Agency; 2012. *www.fbrcameroun.org/cside/contents/docs/Procedure_Manual.pdf.*

As PBF had never been implemented in Cameroon on any meaningful scale and had never been systematically evaluated, our larger policy objectives for the impact evaluation are to (a) Identify the impact of PBF on maternal and child health (MCH) service coverage and quality, and to (b) Identify key factors responsible for this impact. In doing so, we expect that the results from the impact evaluation will be useful to designing national PBF policy in Cameroon and will also contribute to the larger body of knowledge on PBF. Though we are interested in a wide range of outcomes in this report, we consider the main outcomes in terms of coverage to be ANC including anti-tetanus vaccination, skilled deliveries, vaccinations and family planning. We used skilled deliveries for the power calculations to determine the study sample size.

The impact evaluation will focus on the following research questions:

1. Does the PBF program increase the coverage of MCH services?

2. Does the PBF program increase the quality of MCH services delivered?

3. Is it the enhanced monitoring & evaluation and supervision or the link between payments and results that leads to improvements observed in quality or coverage?

4. What is the contribution of enhanced supervision and monitoring to improving MCH service coverage and quality in the absence of increased autonomy or additional financial resources?

In addition, the impact evaluation will also examine the following research questions that relate to intermediate outcomes in the hypothesized causal pathway:

1. Does the PBF program lower informal charges for health services?

2. Does the PBF program lower formal user charges?

3. Does the PBF program improve physical and social accessibility of health services? Accessibility of health services will be examined in terms of the convenience of facility opening hours, availability of services through outreach, client perceptions of convenience of accessing health services and client perceptions of health providers' attitudes towards clients.

4. Does the PBF program lower staff absenteeism?

3. Methods

TREATMENT GROUPS

Table 1 describes the four study groups formed by randomizing Medicalized Health Centers (CMAs), or primary health centers with a medical doctor on staff, and Integrated Health Centers (CSIs) (primary health care centers without a doctor). The randomization for this study was at the health facility level. From an operational and public health perspective, randomizing at the district level would have made more sense given the proximity of some facilities. Indeed, the risk with

facility-level randomization is that neighboring facilities from different groups might learn from each other and apply principles outside their treatment group. However, this was not feasible given that the Government of Cameroon had already decided and announced which districts would be included in the PBF pilot. Randomization at the district level was therefore precluded.

Public randomization ceremonies were held in each region between February and June 2012, just prior to the launching of the PBF program in each region (De Walque, Robyn, & Sorgho, 2013). All health facility management staff from health facilities in the districts covered by the evaluation attended the randomization ceremony.

For the purposes of our study, the 'full' PBF package of interventions included the following elements:

- Linking payment and results, including performance bonuses for health workers.
- Independent monitoring of results.
- Systematic supervision of health facilities defined as regular supervision by an external supervisor from the district hospital team using a structured checklist and providing immediate feedback to facility staff on problems identified and potential solutions to improve service delivery. Systematic supervision included monitoring whether the facility is complying with national user fee guidelines.
- Limited managerial autonomy to facilities defined as autonomy over use of resources combined the ability to hire *additional* staff using health facility income and managerial discretion.

Facilities in group T1 implemented this full PBF package. Facilities assigned to group C1 received a fixed per capita budgetary supplement that matches the per capita budgetary allocation for T1 facilities. However, this supplement was not linked to performance. C1 facilities received the same supervision and monitoring and managerial autonomy over the budgetary supplement received. Both T1 and C1 facility managers had the autonomy to hire staff with their PBF revenues or budgetary supplement received, and also to fire these staff if necessary. T1 and C1 facility managers also had the autonomy over how to use these revenues. C2 facilities received no additional resources but the same supervision and monitoring as T1 and C1 facilities. District-level supervisors responsible for supervising T1, C1 and C2 facilities used the same tools and received the same supplementary payments for visits to facilities in these three groups. However, quality scores were linked to facility payments only in the case of T1 facilities. C3 facilities were the 'business as usual' facilities and did not receive any additional resources or inputs. C2 and C3 facility managers did not have the autonomy to hire/ fire staff or financial autonomy. National user fee caps, and facility user fee rates, were published on a signboard placed in <u>all</u> study group health facilities. The IE team also included monitoring of adherence to national guidelines as part of the monitoring and supervision intervention in T1, C1 and C2 facilities. As the status quo group, the C3 facilities did not receive this additional monitoring & supervision. A summary table describing the intervention groups is provided in Appendix Table A2.

The number, type, and percent private of study health facilities in each study district are shown in Table 2. All public and private health facilities in the 14 study districts that were officially registered with the Ministry of Public Health were eligible for inclusion in the study. All district hospitals in 14 health districts were included in the full PBF (i.e., treatment) arm. This is because

district hospitals play a critical role in supervising and acting as source of referral services for all facilities in the district. District hospitals did supervise and support treatment and comparison group CMAs and CSIs differently based on the group they are assigned to. Household and facility-based surveys were implemented in district hospitals and households associated with their catchment areas[2] in the 14 pilot districts to gain insights into the role that district hospitals are playing in the 4 study groups. However, these data will not be used for making inferences about the impact of PBF, and are not included in the analyses presented in this report.

DATA SOURCES

The evaluation relied on two main sources of data to answer the impact evaluation research questions identified:

1. Household surveys: A household survey implemented at baseline (i.e., before implementation of PBF began), and at endline (i.e., after PBF was implemented for two years).
2. Facility-based surveys: A facility-based survey was also implemented at baseline and at endline.

Both surveys are described in detail in the Appendix.

STATISTICAL METHODS

FACILITY AND CATCHMENT AREA EXCLUSIONS

Most of the analysis included in this report includes only those health facilities that were surveyed at both baseline and endline. Similarly, for all of the results coming from the household surveys, these results include only those villages that were surveyed at baseline and endline. The only exceptions are the analyses that use data from the direct observation and exit interviews from ANC and child health consultations. These analyses include all of the women and children who received these services on the day of the facility survey. The reason that we deviated from the above stated exclusion criteria is that, especially in the ANC sample, the exclusion results in large sample size loss. Specifically, the full sample for ANC included data from 733 visits and 118 health facilities while the restricted sample included 561 visits at 46 health facilities.

SPECIFICATIONS

The main difference-in-differences specification that we used in the household data analysis is displayed below:

$$Y_{ijt} = \alpha_j + \gamma_{2015} + \beta_1 T1_j I_{2015} + \beta_2 C1_j I_{2015} + \beta_3 C2 I_{2015} + \beta' X_{it} + \epsilon_{ijt}$$

Where Y_{ijt} is receipt of the health service for woman/pregnancy *i* in enumeration area *j* in survey year *t*. α_j is an enumeration area fixed effect, γ_{2015} is a dummy variable that is equal to 0 in 2012 (baseline) and 1 in 2015 (endline). *T1*, *C1*, and *C2* are dummy variables that are equal to 1 when the enumeration area was assigned to each treatment group respectively and zero otherwise. The

[2] Some villages will not fall within the catchment areas of other CSIs and CMAs in the district. Households in these villages will be excluded from the sample for the impact evaluation.

treatment variable is based on the assigned catchment area where the household is located; however, this may not have been the health facility where the household sought health care. $\beta_1 T1_j I_{2015}$, $\beta_2 C1_j I_{2015}$, $and\ \beta_3 C2 I_{2015}$ are interaction terms between each of the *T1*, *C1*, and *C2* groups and the post indicator. These interaction terms measure the treatment effect in each group and can be interpreted as the difference in the change in health service use over the study period between the control group, and each treatment group respectively. $\beta' X_{it}$ is a vector of control variables at the individual level (age, marital status, education level, religion, ethnicity, working status and type of work), and at the household level (number of individuals in the household, housing type, house ownership, water source, and type of sanitation). Control variables were included in all analysis of household level data.

For the household survey, a random sample of 16 to 20 households was selected in each health facility catchment area. The analysis of the household survey in this report starts from the assumption that household members seek care in the health facility closest to where they live, or in other words that people living in the catchment area of a facility obtain health care in that facility. However, it is apparent from the baseline survey data that households do not always seek care from the closest health facility in their health zone (see Appendix Table A3). We address this issue in the Appendix section titled Analysis of Household Care Seeking Behavior.

The main specification used in the analysis of facility-level data is presented below:

$$Y_{it} = \alpha_i + \beta_1 I_{2015} + \beta_2 T1_i I_{2015} + \beta_3 C1_i I_{2015} + \beta_4 C2_i I_{2015} + \beta' X_i + \beta' X_{it} + \varepsilon_{it}$$

A similar specification was used in most facility level analysis; however, in all analyses involving direct observation or exit interview data we have used an alternative specification without facility fixed effects. The alternative specification was chosen after finding that the variable measuring the duration of ANC visits contained many outliers, and the use of fixed effects produced results that diverged widely from changes observed descriptively by comparing means. Therefore, this analysis instead included treatment group dummy variables to control for baseline differences between groups. For consistency, all analysis of women sampled from antenatal care, and of caregivers sampled from child health consultations used this alterative specification. Additionally, sampling for these services was limited in many health facilities because antenatal care is only provided on certain days of the week, and due to low patient flows in smaller facilities. For this reason, analysis of patients sampled from health facilities was not restricted to health facilities represented in both the baseline and endline survey data.

Analysis at the facility level included the following time invariant controls: type of health facility (public/religious/private) and location of the health facility (urban/rural). Additionally, when the analysis was at the individual level (i.e. women sampled from ANC visits, care givers sampled from child health consultations) the following individual level controls were also included: age, sex, marital status, and education level.

4. Results

Appendix Tables A8 and A9 display the baseline levels of individual and household level characteristics from the household survey. Group level means are compared individually using two-sample statistical testing, and F-tests were conducted to test for overall differences in the four

study groups. All statistical testing adjusted standard errors for clustering. The groups appear balanced on all of the individual characteristics assessed including age, religion, ethnicity, educational attainment, literacy, employment and marital status. The study treatment groups were also generally well balanced on household level characteristics including household composition, type of household, household ownership, and sanitation type. However, the study groups were not balanced at baseline on the type of water source used at the household.

Appendix Tables A10 and A11 display facility characteristics and health service coverage at baseline. The facility sample appears well balanced at baseline on most characteristics assessed; however, there was a difference between the study groups in the likelihood that the facility had an incinerator. The sample was well balanced for most services assessed, though we found statistical differences at baseline for growth monitoring and for documented childhood vaccination coverage.

COMPARISON OF OPERATIONAL FUNDS AVAILABLE AND SUBSIDY PAYMENTS TO C1 AND T1 AT ENDLINE

One important feature of the impact evaluation design was that the subsidies provided to control group 1 (C1) – which were not linked to performance – should be equal to the payment amounts provided to facilities in the full PBF treatment group. At endline, the health facility survey collected data on the amount of funding available at the health facilities for operation expenses, including the amount of revenue collected from cost recovery, funding provided directly by the Ministry of Health, and subsidies from the PBF program. To verify the equal receipt of financing, we compared data on total financing available, and subsidy amount received between T1 and C1 health facilities. The first panel of table 3 displays the average amount of health facility revenue from all of these sources during each trimester of 2014 in T1, the full PBF group, and C1, the increased financing group. The second panel compares the amount of subsidies received in each of the two groups. Neither the subsidy amount, nor the total amount of financing available at the health facility differed statistically between the two groups during any trimester of 2014. We also sought to assess whether per capita payments were equal between groups, as this was the intention of the study. However, exact catchment area population numbers are not known. Therefore, we standardized payments by the number of health workers in each facility as a proxy measure for catchment area population. Using this proxy measure, we found no differences in payment amounts between groups (Table 3).

We also confirmed the subsidy data collected in the facility questionnaire by requesting subsidy data from the regional funds. These data, which represent the payments made by the regional funds to the health facilities in treatment groups T1 (full PBF) and C1 (additional financing), are presented in Appendix Figures A5 – A7. These figures also show that the total payments provided to the health facilities in each treatment group were equal during the entire study period.

While the payment results show that overall the two groups received equivalent financing volumes for the same number of health facilities (approximately 50 each) during the study period, when looking at per capita financing we find that the C1 group in fact received higher levels of per capita financing than the T1 group. This is largely due to the fact that in the South-West region, while the T1 and C1 groups had the same number of facilities (21 for T1 and 20 for C1), the total population covered by these health facilities varied substantially. In the T1 group several health

facilities had very large catchment areas, resulting in a target population approximately three times higher than in the C1 group. As such the per capita payments in the South-West region were three times higher for the C1 group than the T1 group, which also affects the overall annual per capita payments (Appendix Figure A8). That being said, the overall payments for each group, when combining the three regions, shows equal payments across the two groups (Appendix Figure A9). Populations of catchment areas in Cameroon should also be interpreted with precaution given the lack of an up-to-date and comprehensive national health map.

FACILITY SURVEY UTILIZATION RESULTS

This section describes the results of PBF on health services provision as recorded in facility registers. To assess the reliability of these data, we examined the health service counter-verification data that were collected routinely as part of the PBF program design through community client satisfaction surveys. Health service verification took place in all PBF health facilities, as well as in health facilities in control groups C1 (additional financing) and C2 (additional supervision). Thirty-five patients were sampled for 7 health service categories each quarter. Figure 2 shows the percentage of patients who were reported by health facilities to the PBF verification terms, who were later confirmed to have received health services at the health facility. During most quarters of the three-year study period in all three study regions, over 80% of reported patients were confirmed. The trend in confirmed patients increased slightly over time in North-West and East, with confirmation rates above 85% in all three regions during the final year of the study. Though we find reassuringly high quality data among the treatment group health facilities, it is possible that facilities in the full control group had less incentive to keep records of all services provided. Given that we did not verify the health service data collected from the full control group, we cannot investigate this possibility. Therefore, the possibility of incomplete reporting in the full control remains a limitation of this analysis.

Table 4 displays health service utilization results as assessed in the facility survey. Facility level provision of health services in the six months before the survey took place was assessed using patient registers from study health facilities. All data were collected at the monthly level; therefore, the interaction term coefficients represent differences between groups in the change in monthly services provided. Provision of skilled delivery did not increase in the control group during the study period. Relative to the control there were no statistical differences in mean monthly provision of skilled delivery in the full PBF and the additional supervision group; however, there was a relative increase of approximately 2 monthly deliveries in the additional financing group (additional financing 1.855, p-value = 0.071). The overall change between 2012 and 2015 in antenatal care provision in study health facilities was positive, but was not statistically significant. Relative to the control group there was an increase in antenatal care in all three treatment groups; however, none of these differences were statistically significant (Table 4).

Comparing the six-months before the baseline, and the six months before the endline survey, provision of tetanus toxoid vaccine declined by a monthly average of almost 17 vaccinations in control facilities each month (Table 4, column 3). Compared to the change observed in the control group, there was a positive and statistically significant difference in the PBF and additional financing groups (PBF 21.521, p-value = 0.001; additional financing 15.989, p-value = 0.014). Compared to control facilities, facilities with additional supervision provided on average approximately nine more tetanus vaccines monthly to pregnant women, but this difference was

not statistically significant. Like tetanus vaccine, there was a statistically significant decline in postnatal care provision in the control group over the study period. On average, approximately four fewer monthly postnatal care visits were provided in control health facilities at endline compared to baseline. Though the interaction term was positive, there was no statistical difference between the additional supervision group and the control group. However, the change in monthly provision of postnatal care in facilities in the PBF group and the additional financing group was greater than the change in the control group (PBF 4.309, p-value = 0.059, additional financing 5.513, p-value = 0.016) (Table 4, column 4).

Table 4 column 5 presents the results of the effect of PBF on modern contraception. Modern contraception included women provided contraceptive implants, injectables, oral contraceptive pills, and the intrauterine device. There was little change in the control facilities in monthly modern contraceptive provision. In the PBF group, there was an increase of just over nine women each month on top of the small change in the control group who were provided modern family planning (9.240, p-value < 0.001). The change in modern contraceptive delivery in the additional financing group is greater than the small change in the control group with approximately six more women per month provided family planning (5.794, p-value = 0.001) in each health facility. The change in the additional supervision group was not statistically different from the change in the control group. This strong effect of PBF and additional financing on family planning in the facility register data differs from the null finding we found in the household data. The difference in the findings between data sources could be explained by several factors. One possibility is that the household survey oversampled recently pregnant women (because having a least one recently pregnant womanin the household was an inclusion criteria). It is possible that the need for modern family planning is less important in this population of women among whom many have recently shown a willingness to have children. In addition, there is some evidence that women might have been uncomfortable disclosing their use of family planning in the household survey. For example, we find that 47% of women report that their husbands are against use of FP to avoid pregnancy, and these women may not have been comfortable discussing family planning in their household.

Table 4, columns 6 – 8 display the results of the impact of PBF on provision of childhood vaccines. In general vaccine provision in the six months before the baseline declined as compared to provision in the six months before the endline, as demonstrated by the negative and statistically significant coefficients on the post indicators for all three vaccine outcomes in the table below. This decline has been explained by Ministry of Health officials as resulting from country level stock-outs during the months before the endline. Despite these large declines, the facility results for vaccine provision may not be inconsistent with the household vaccine coverage results, which showed large increases over time, and a particularly large change in the PBF group. According to the recommended vaccine schedule, most childhood vaccines should be received between birth and 12 months of age. To avoid including infants who, due to their age, should not yet have finished their vaccine schedule, household level coverage only included children between 12 – 23 months old. Consequently, the children included in the household data would have received vaccinations prior to the months covered in the facility level data at endline (i.e. before the vaccine stock-out). Both sets of results show that the PBF group, and to a lesser degree, the additional financing group, performed much better than the control group. Therefore, the findings from the household and facility data showing large and statistically significant differences between groups

are consistent, and we believe that the level changes can be explained by a country level shock affecting vaccine availability that affected the facility, but not the household data.

Facility-level provision of the third dose of polio vaccine decreased by approximately five vaccines per month in the control group (Table 4, column 6). There was an increase of 4.583 final polio vaccinations relative to the control group in the PBF group, and this difference was statistically significant (p=0.035). The coefficient on the interaction term was also positive in the additional financing group but the difference was not statistically significant, and there was no difference between the control and the additional supervision groups. There was a large and statistically significant decline in meningitis vaccination provision over the study period (-45.970, p-value < 0.001). Of the three treatment groups, only the change in the additional financing group was statistically different from the control group (21.931, p-value=0.050). Finally, in the control group measles vaccine provision declined by an average of approximately four children per month over the study period. There was no difference between the change observed in the control group, and the changes in the treatment groups (Table 4, column 8).

Facility register data also contained data documenting facility provision of HIV-related services (Table 4, columns 9 – 11). We found a large and statistically significant effect of both PBF and additional financing on HIV testing. An average of 61 more patients were tested for HIV in PBF facilities than control facilities, and 51 more patients were tested monthly in the additional financing arm compared to the control. There was very little change in HIV testing in the additional supervision group, and the effect of PBF was greater than the effect of additional supervision. Though there was a small increase in PMTCT in all of the treatment groups relative to the full control, none of these differences were statistically significant, and there was no difference between the effect of PBF and the other treatment groups. Similarly, there was no difference in the change in ART provision between the treatment groups and the full control.

HOUSEHOLD SURVEY UTILIZATION RESULTS

Table 5 displays the difference-in-differences regression results for the study outcomes related to health care received during pregnancy as assessed in the household survey interviewing women with recent pregnancies or birth experiences. In this table, the post indicator can be interpreted as the change in the outcome over the study period in the control group. Each of the interaction terms can be interpreted as the difference between the change observed in each treatment group respectively compared to the change in the control group. Table 5 shows that overall few treatment effects were observed for study outcomes related to care during pregnancy in the household survey data. Among women who had been pregnant in the 24 months before the survey, there was an overall increase of approximately five percentage points in delivery with a skilled birth attendant. There was no difference between the control group and the PBF group, or the group that received additional financing in the change in skilled delivery over the study period. However, skilled delivery declined (-0.050*, p-value = 0.087) in the group receiving only improved supervision relative to the control group.

The change between 2012 and 2015 in the percentage of women who received at least two antenatal care visits was not statistically significant in the control group. Similar to delivery care, there was no difference between the change in the control group and the PBF and additional financing groups but the change among women in the improved supervision group was less than

the change observed in the full control group (-0.044**, p-value = 0.022) (Table 5, column 2). There was very little change over the study period in receipt of tetanus vaccine during pregnancy, and there was no difference in the change in vaccine receipt between the three treatment groups, and the control group (Table 5, column 3). Postnatal care receipt increased by over ten percentage points over the study period in control group (0.105, p-value = 0.001). Compared to the control group, there was a smaller increase over time in postnatal care in the three treatment groups; however, this difference was not statistically significant in the PBF group and the additional financing group. Postnatal care receipt increased by seven percentage points less in the improved supervision group (C2) than in the control group, and this difference was statistically significant (-0.070, p-value = 0.075) (Table 5, column 4). Results from testing the equality of coefficients show that for skilled delivery the additional financing intervention outperformed the PBF group. For antenatal care, facilities with the PBF intervention performed better than those facilities with only additional supervision.

To further investigate the influence of health care bypassing behavior, we conducted additional analyses for key indicators dividing the sample into high and low bypass strata. We specifically tested whether the effect of the treatment groups differed in catchment areas where a larger proportion of women sought care outside of their treatment group health facility at baseline. To do this, we generated a binary variable = 1 if the women went to her assigned health facility at baseline and = 0 if she did not, and collapsed the data at the health facility level taking the mean of the bypass indicator variable. We then merged this variable onto the original file so that for each woman in the data set we know the proportion of women in her catchment area who went to their assigned facility at baseline. Then we divided the sample into high bypassing (above the median) and low bypassing (below the median), conducted the analysis in the separate groups, and compared the coefficients on the three interaction terms between groups. We included antenatal care, skilled delivery and postnatal care in this analysis.

The hypothesis behind these additional analyses is that if there are spillovers, we should see larger treatment effects in areas with low bypass rates. The only difference in the results between high and low bypassing areas was that the effect of additional financing was negative (-0.108) in the high bypassing group and positive in the low bypassing group (0.045), and this difference was statistically significant (p=0.058) (results shown in Appendix Table A7). Otherwise the impacts of the different interventions tested did not vary according to whether health care bypassing behavior was high or moderate at baseline. This analysis further suggests that the bypassing behavior observed in Cameroon, while substantial, did not significantly bias our impact measures.

Non-pregnant fertile women of reproductive age (15 – 49 years) with a current sexual partner were asked whether they were using any form on contraception in the household survey. Modern contraception included the intrauterine device, injectables, implants, oral pills, diaphragm, foam/jelly & lactational amenorrhea. The percentage of women of reproductive age who used any form of modern contraception, excluding condoms, did not increase between 2012 and 2015 in the control group (Table 5, column 5). The changes observed in the treatment groups did not differ statistically from the change in the control group, and there was no difference between the treatment groups.

Mothers or primary care givers of all children under five years of age were asked about their child's vaccination history. For all children with a vaccine card, study enumerators recorded all documented vaccinations and their respective receipt dates. Mothers/primary caretakers were also asked to report any vaccinations that were not recorded in the vaccine card. For these questions, enumerators asked a separate question for each vaccine type that referenced the vaccine name and also gave an indication of its method of administration (i.e. for polio "that is drops in the mouth") as a guide for respondents. Only children between 12 – 23 months of age were included in these analyses. Both outcomes include the following vaccines: oral polio vaccine, yellow fever, diphtheria and whooping cough (DTC), measles, and Bacillus Calmette–Guérin (BCG). Table 6 column 1 shows that among children with a vaccine card, there was an almost 13 percentage point increase in full vaccination over the study period (0.127, p-value = 0.080). In the PBF group, there was a further 17 percentage point increase in full vaccination (0.170, p-value = 0.076). There was no difference between the control group, and the additional funding and additional supervision groups in full vaccination as documented in vaccination records. The second vaccination outcome included both documented vaccine receipt as well as any self-reported vaccines. This outcome displayed a similarly large increase in the control group (0.108, p-value = 0.039). Additionally, there was a further 16.4 percentage point increase in full vaccination over the study period in the PBF group (p-value = 0.019). Finally, neither the additional funding, nor the additional supervision group showed an increase in full vaccination beyond the increase in the control group. The p-values directly comparing the PBF group with the other groups (C1 and C2) further indicate that the PBF group outperformed the three other study groups (Table 6, column 2).

The next two outcomes assessed from the household survey were growth monitoring in the month before the survey, and having slept under a bednet the night before the survey. Both outcomes were assessed among children under five years of age; however, children who were less than 12 months old were not included in the growth monitoring analysis. There was no change in growth monitoring during the study period in the control group. Similarly, growth monitoring did not increase in the PBF and the additional supervision groups. However, growth monitoring increased by approximately 3 percentage points more in the group that received additional funding (0.031*, p-value = 0.071) than the control group (Table 6, column 3). The proportion of children who slept under a bednet the night before the survey declined by approximately 18 percentage points during the study period (-0.186, p-value = 0.000) (Table 6, column 4). A similar decline was observed in all the treatment groups as shown by the small and non-statistically significant coefficients on the interaction terms. It should be noted that neither growth monitoring nor bednet distribution were included in the package of services incentivized in the PBF program.

ANTHROPOMETRICS

The height and weight of all children under 5 years of age were recorded during both baseline and endline surveys. Though nutrition indicators were not incentivized in the Cameroon PBF program until 2015, we used the height and weight data collected in the household survey to assess the effect of PBF on child nutrition outcomes. Using the WHO child growth standard's Stata package, we calculated a z-score for child height-for-age, weight-for-age, and weight-for-height. These z-scores represent the location where each child falls according to international standards defining healthy child development. Using these scores, we calculated the prevalence of stunting (height-for-age of less than -2 standard deviations from the mean), underweight (weight-for-age of less

than -2 standard deviations from the mean) and wasting (weight-for-age of less than -3 standard deviations from the mean). As shown in table 6, columns 5 – 7, there was no difference between the changes in prevalence of stunting, underweight, and wasting in the treatment groups and the control groups. While there was no change over time in stunting and underweight, the prevalence of wasting increased at endline compared to baseline.

HEALTH CARE SPENDING

All household members were asked if they had been sick in the four weeks before the survey. Mothers or caregivers responded on behalf of household children. All respondents who had been sick were asked if they had gone to any health facility, health personnel or traditional healer to seek care for this illness. Respondents were then asked how much the household spent out-of-pocket for the treatment of this recent illness. Table 7 presents the results for several different types of reported spending. Neither the change in the control group on spending for official provider fees, nor the differences between the treatment groups and the control group were statistically significant (Table 7, column 1). There was an increase of approximately 2,052 CFA ($3.31) in the amount of unofficial provider fees paid in the control group between 2012 and 2015. The difference between the PBF group and the control group in the change over the study period was -2,254 CFA ($3.64), and this difference was statistically significant. Relative to the control group, unofficial payments also declined in the additional financing group and the additional supervision group; however, these differences were not statistically significant (Table 7, column 2). There was a non-significant increase of approximately 1,048.64 CFA ($1.69) in laboratory and x-ray fees over time in the control group. Compared to the control group, laboratory and x-rays fees declined by 1,473.44 CFA ($2.38), and this difference was statistically significant (p-value = 0.060) (Table 7, column 3). Finally, transportation fees did not change between 2012 and 2015 in the control group. There was a statistically significant difference between the control group, and the additional financing group with a decrease of 495.14 or $0.80 (Table 7, column 4).

Health spending was also assessed in the exit interview surveys conducted among women who had received prenatal care, and the caregivers of children under five years of age who visited the health facility for a child health consultation. Table 8 displays these results, even though reports by patients about health spending might be more reliable when given in the privacy of the household (household survey) than during exit interviews which take place in or close to the health facility compound, especially for reports about unofficial payments to providers. There were no changes over time, and no statistically significant differences between groups in unofficial provider fees, medicine fees, and total fees for antenatal care. The change in official provider fees for antenatal care was -1,025.34 CFA ($1.68) lower in the PBF group compared to the control group, and this difference was statistically significant (p-value = 0.083). Relative to the control group, the change in the additional financing group in official provider fees was 1,824.81 CFA ($2.98) higher (p-value = 0.038). Spending on official provider fees was significantly lower in the PBF groups than in the two treatment groups. Additionally, there was a statistically significant increase in the additional financing group and the additional supervision group in total fees for antenatal care. There were no changes over time, or differences between groups in any of the spending outcomes (official provider fee, unofficial provider fee, medicine fees, and total fees) for child health consultations.

PATIENT SATISFACTION

SATISFACTION WITH ANTENATAL CARE

Table 9 provides an overview of participants in the ANC exit interviews at baseline. The average age of respondents was just over 25 years, 79 percent of respondents were married at the time of the interview, and 65 percent were literate. The intervention arms were well balanced across all of these dimensions. There is, however, evidence of differences in education. In particular, women attending facilities in the additional financing arm were significantly more likely to have attended higher education than were women attending other facilities (p=0.07).

Women were asked a series of 12 questions related to their satisfaction with individual elements of their visits including, for example, their satisfaction with costs, wait times, and health worker communication. For each, a statement was read, and women were asked if they agreed, were neutral, or they disagreed. Binary variables were created by coding responses as "1" if a woman agreed, and "0" otherwise. Overall satisfaction scores were calculated by averaging over these 12 components. An overall score of "1" indicates that a woman agreed with all 12 questions, while a score of "0" indicates that she either disagreed or was neutral on all 12 questions.

The impact of the interventions on overall satisfaction is shown in Table 10. There is no indication that satisfaction changed over time in the control group ($\beta = 0.006$, $p = 0.847$). Relative to the pure control, the PBF group was associated with an 8.6 percentage point increase in satisfaction ($p = 0.077$). The results suggest a stronger effect in the full PBF than in the additional supervision group (10.5-percentage point increase) (Table 10, column 1).

Table 10 also shows the breakdown of the 12 individual components of satisfaction. Regression results indicate the percentage point increase or decrease in probability that a respondent in a given intervention group will agree with a statement, relative to the pure control, after adjusting for individual-level indicators (age, literacy, marital status, and education level) and facility-level indicators (type of health facility public/private/religious, urban/rural).

For the first three dependent variables reporting whether women found the fees reasonable, while few of the results meet the 10 percent cutoff for statistical significance, women in the PBF group are consistently more likely to say that fees are reasonable than either the full control or partial treatment groups. The only statistically significant difference among these fee-related variables was that women in the PBF group were more likely to agree that medicine fees were reasonable than women in the additional financing and improved supervision groups (Table 10, column 4). There were no differences between groups in the likelihood that women agreed with the statement that health workers did not ask for additional presents or payments (Table 10, column 5).

Focusing on reported facility cleanliness, women in the PBF and the additional financing group both reported significantly higher levels of agreement than in the pure control group, although these scores were not significantly different from one another (Table 10, column 6). Although not meeting the 10 percent cut-off, this pattern was also seen in response to wait times, where the PBF group was associated with 16.1-percentage point increase over the pure control, compared to a decrease of 5.5-percentage points in the supervision only group. The pattern is slightly different in response to the question on privacy – while the additional supervision group continued to underperform relative to the pure control, the point estimate of the difference between the PBF

and pure control is almost 0, while the additional financing group reported the highest rate of satisfaction. However, none of these estimates achieve statistical significance (Table 10, columns 7 & 8). The results for the adequacy of hours are consistent with the pattern observed earlier. Compared to the control group, PBF results in a large and statistically significant 15.4-percentage point increase satisfaction with the facility's hours, while the additional supervision is associated with a non-significant reduction in satisfaction relative to the pure control. As indicated by the p-value on PBF versus the improved supervision, PBF performed consistently and significantly better than the improved supervision group on facility cleanliness, and adequacy of hours (Table 10, column 6 & 9).

Women attending facilities receiving the full PBF intervention reported significantly higher levels of satisfaction with health worker communication than did women attending control clinics (β=0.106, $p<0.05$), but there was no evidence of an impact of PBF on the courteousness of health staff, time with health workers, or the ease of getting prescribed medicines (Table 10, columns 10 – 13). Women attending facilities receiving additional financing reported significantly higher levels of satisfaction with the amount of time they spent with health workers than women in control clinics (β=0.139, $p<0.1$). This increase in the additional financing group was greater than the change in the PBF group (which was negative relative to the control group).

SATISFACTION WITH CHILD HEALTH CONSULTATIONS (< 5 YEARS OLD)

Table 11 provides an overview of participants in the child health consultation exit interviews at baseline. The mean age of children attending facilities in the PBF arm was 23.7 months, which is older, on average, than those attending facilities in other arms ($p<0.05$). The youngest group was those attending facilities in the additional financing arm (mean age: 14.2 months). The intervention arms are well balanced on gender, with females accounting for approximately 51 percent of children attending the facilities.

Looking at the caretaker characteristics, those in the control group were less likely to be single (16 percent, compared to a sample mean of 20 percent), while those in the PBF group were least likely to be married (60 percent, compared to a sample mean of 73 percent). Few (3 percent, overall) were divorced or widowed, and the treatment arms were well balanced on literacy and education. Approximately 74 percent of the caretakers were literate (low: 69 percent in the improved supervision arm; high: 77 percent in the additional financing and pure control arms). The majority of women had some education, with most stopping during primary school (39 percent) or secondary level 1 (30 percent).

Caretakers were asked a series of twelve questions related to their satisfaction with individual elements of visits. These questions were the same as those asked following antenatal visits and included, for example, their satisfaction with costs, wait times, and health worker communication. For each, a statement was read and women were asked if they agreed, were neutral, or they disagreed. Binary variables were created by coding responses as "1" if the caretaker agreed, and "0" otherwise. Overall satisfaction scores were calculated by averaging over these 12 components. An overall score of "1" indicates that a woman agreed with all twelve questions, while a score of "0" indicates that she either disagreed or was neutral on all 12 questions.

We find evidence that PBF had a positive impact on overall satisfaction with child health services (Table 12, column 1). Relative to the pure control, the PBF was associated with a statistically

significant 9.9-percentage point increase in satisfaction ($p<0.05$). As was the case with the ANC exit interviews, we find a stronger effect in the full PBF than in the additional financing group (5.4 percentage point increase relative to full control) or supervision group (2.2-percentage point increase), and neither of the partial treatments achieves statistical significance at the 10 percent cutoff. While PBF and the additional financing group are not statistically significantly different from one another, the 7.7-percentage point difference in reported satisfaction between the PBF and the enhanced supervision (C2) group is statistically significant ($p<0.10$).

The remaining columns in 12 show the breakdown of the twelve individual components of satisfaction. Point estimates indicate the percentage point increase or decrease in probability that a respondent in a given intervention group will agree with a given statement, relative to the pure control, after adjusting for individual-level indicators (age, literacy, marital status, and education level) and facility-level indicators (availability of electricity, availability of piped water, availability of latrine, facility open 24 hours, type of health facility, urban/rural status, and number of health workers employed at the facility).

Focusing on the costs associated with care, all three groups are associated with statistically significantly higher satisfaction with laboratory fees relative to the pure control (Table 12, column 3). However, they are not significantly different from one another. The additional financing control had the highest (non-significant 11.1-percentage point) estimated impact on satisfaction with medicine costs; however, none of the differences between the control and treatment groups were statistically significant for this outcome (Table 12, column 4). There was no difference between any of the intervention arms, and the control group in satisfaction with registration fees, and informal payments (Table 12, columns 2 & 5).

PBF is associated with a large and statistically significant impact on satisfaction with the health facility cleanliness ($\beta=0.227$, $p=0.090$). Neither the additional financing nor the additional supervision intervention groups (C1 and C2) performed better on cleanliness at endline than they did at baseline (Table 12, column 6). Focusing on waiting times, all of the intervention arms appear to result in improvements over the control group, although none of the differences were statistically significant (Table 12, column 7). All of the arms also resulted in increased satisfaction with the privacy at health facilities, and the very large point estimate on the PBF (33.6 percentage points) is significant at $p<0.01$ (Table 12, column 8). Satisfaction with the opening hours did not change over time in any of the treatment groups, but the change in the PBF group was greater than the change in the additional supervision group (Table 12, column 9).

Moving to health worker characteristics, PBF does not appear to have had any impact on health staff courteousness ($\beta = -0.012$), and the point estimates on the two partial treatments indicate a negative impact on health worker courteousness, relative to the pure control (Table 12, column 10). All three intervention groups have negative, but not statistically significant, point estimates on satisfaction with the time spent with health workers, compared to the pure control (Table 12, column 12). By contrast, all three were associated with positive but non-statistically significant impacts on the ease of getting prescribed medications (Table 12, column 13). Overall, standard errors are large relative to point estimates and, across these variables, no statistically significant differences could be seen between groups, either between the pure control and the three intervention groups or between the different intervention arms themselves.

HEALTH WORKER SATISFACTION AND MOTIVATION

In all, 434 health workers were interviewed at baseline. Key characteristics are described in Table 13. The mean age of health workers was approximately 39 years and, on average, workers had been employed at the facility for between 4.4 (improved supervision) and 6.2 years (control; statistically different with p=0.05). Approximately two-thirds of health workers were female, and 75 percent had received either basic or level 2 secondary education. Just under half of workers were employed by the Ministry of health, with the remaining employed by religiously affiliated health facilities (19 percent), the facility (16 percent), or other employers (20 percent).

Health workers were read a series of statements relating to their wellbeing over the two weeks prior to the survey. These statements were taken from the World Health Organization's (WHO) Well-being Index (Appendix Section WHO Being Index) and included, for example, "In the last two weeks, I have felt active and vigorous." For each question, health workers were asked to indicate whether the statement described their state most of the time, more than half the time, less than half the time, only rarely, or never. For the purposes of analysis, these data were recoded into binary indicators. Responses were coded as "0" if the health worker replied half the time or less, and "1" otherwise. Thus, point estimates on the postXintervention arms indicate the percentage point increase or decrease in probability that a respondent in a given intervention group reported that a given statement was true at least half the time, relative to the pure control, after adjusting for individual-level indicators (age, sex, marital status, and education level) and facility-level indicators (type of health facility, urban/rural status).

Overall, the data do not provide strong evidence that PBF affected attributes included in the WHO's wellbeing index (Table 14). Point estimates are generally small relative to standard errors, and there are no statistical differences between the pure control and the other intervention groups. However, the change in index score for three items – Active and energetic in the last 2 weeks, Refreshed and rested in the morning in the last 2 weeks, and Days filled with interesting things in the last 2 weeks was lower in the PBF group than the additional financing group.

Health workers were also asked a series of 26 questions related to their satisfaction with working conditions. These questions covered a range of topics, including relationships with individuals within and outside of the health facility, facility infrastructure and readiness to deliver services, salary and benefits, and their overall capacity to provide high-quality health services. For each question, a score of "1" indicates satisfaction, while a "0" indicates that a health worker reported either indifference or dissatisfaction. Point estimates, thus, indicate the percentage point increase or decrease in probability that a respondent in a given intervention group is satisfied with a particular issue, relative to the pure control, after adjusting for individual-level indicators (age, sex, marital status, and education level) and facility-level indicators (availability of electricity, availability of piped water, availability of latrine, facility open 24 hours, type of health facility, urban/rural status, and number of health workers employed at the facility).

There is little evidence of impact on working relationships (Table 15). The additional financing arm had the largest estimated impact on the relationship between the facility and District or Ministry of Health staff, but at 12.7 percentage points, this did not meet the 10-percent cutoff for significance. The PBF point estimate for an impact on the relationship with District or Ministry of Health staff is also positive and fairly large (β=0.103), but not statistically significant. The

impact on intra-facility working relationships was estimated to be negative, though not statistically significant, in all treatment groups. Satisfaction with working relationships between management and staff within the facilities declined in both the PBF and the supervision arms. Additionally, working relationships with management staff improved in C1 compared to the PBF group. There was no difference between the change in the control group and the change in the intervention groups in satisfaction with collaboration with the regional health delegation, or in the quality of the management of the health facility.

Similarly, there is no strong evidence of an impact either on the relationships between the facility and local leaders or on health workers' perceptions of their own status within the community (Table 15, columns 6 & 7). In both cases, we report a negative but not statistically significant effect of both the PBF and the improved supervision interventions. The point estimates within the financing arms are positive and, in the case of relationships with local leaders, relatively large at 9.3 percentage points, but they are not statistically significant. There was also no evidence of a difference in impact between the intervention arms.

By contrast, we see a large and consistent impact on health workers' satisfaction with the quantity and quality of equipment and other supplies at health facilities, shown in Table 16. Both the PBF and the additional financing arms result in similarly large and highly significant improvements in these measures: an approximately 19 percentage point increase in reported satisfaction with the quantity of equipment ($p<0.05$), approximately 26 percentage point increase in reported satisfaction with the quality of equipment ($p<0.05$), and a 33 to 40 percentage point increase in satisfaction with the availability of other supplies at the health facilities within these two arms ($p<0.01$). By contrast, there was less evidence of an impact in the improved supervision arm. While point estimates are positive, they are not statistically significant at the 10 percent cut-off level. There is also less evidence for an effect on the quality and quantity of medicines. Although the two arms that include additional revenue – the PBF and the financing only arms – both result in positive point estimates, they do not achieve statistical significance.

Table 16 also provides an overview of reported satisfaction with the physical condition of health facilities. Health workers in the PBF arm were 31 percentage points more likely to be satisfied with the physical condition of the health facility building, relative to the pure control ($p<0.01$). While both the financing and the improved supervision arms had positive point values (10.6 percentage points and 9.6 percentage points, respectively), neither was statistically significant at the 10 percent cut-off, and workers in the full PBF arm were more likely to express satisfaction than were workers in either the additional financing ($p<0.05$) or the improved supervision ($p<0.10$) arms.

Despite increased satisfaction with both the physical infrastructure and the quantity and quality of equipment, PBF did not have a statistically significant impact on health workers' perceptions of their ability to provide high-quality care. The effects associated with the two partial treatment arms also failed to achieve statistical significance (Table 16, column 7).

The interventions including financial support also appear to have positively impacted satisfaction with salary and benefits (Table 17). Health workers in the PBF arm were 9.1 percentage points more likely to express satisfaction with their salary and 18.3 percentage points more likely to express satisfaction with their benefits (although only the latter was statistically significant

($p<0.05$)). Effects were even stronger in the additional financing group. Health workers at facilities receiving additional financing were 13.4 percentage points more likely to be satisfied with their salary ($p< 0.10$) and 28.7 percentage points more likely to be satisfied with benefits ($p<0.01$). These effects are not seen in the improved supervision arm. While the point estimates associated with supervision are positive, they are relatively small and are not statistically significant. Satisfaction with living accommodations improved somewhat in all three groups, with point estimates suggesting that health workers were 10 to 16 percentage points more likely to express satisfaction after the pilot. However, these impacts are not statistically significant in any of the three groups.

The point estimates of PBF's impact on opportunities to discuss issues with supervisors, supervisors' recognition of good work, opportunities to be rewarded for hard work, and opportunities for promotion were all negative (β=-0.08, -0.04, -0.02 and -0.11, respectively), although none were statistically significant (Table 17, columns 4 – 7). The estimated impact of the improved supervision arm was also generally negative and, with a point estimate of -25.1 percentage points, the impact of the improved supervision on opportunities for promotion was statistically significant ($p<0.05$). The estimated impact of the financing arm was positive, but was only statistically significant for the opportunity to be rewarded for hard work. The improvement in the financing arm was greater than the change in the in the PBF group for this outcome, and for opportunities for promotion.

As shown in Table 17, the PBF intervention had a small but negative estimated impact on reported opportunities to upgrade skills through training and a small but positive estimated impact on reported opportunities to use skills on the job (neither measure met statistical significance). Point estimates for both measures were positive in the additional financing arm, and the financing appears to have positively impacted satisfaction with opportunities to use skills on the job (β =0.194, $p< 0.10$). The point estimates associated with the improved supervision arm to either use or upgrade skills did not meet the 10 percent cut-off for statistical significance.

Satisfaction with safety and security in the community and with available schooling for children were unchanged by the interventions. The point estimate in the PBF group was large at 13.9 percentage points, but did not reach statistical significance. The point estimates for the other two treatment groups were small and non-significant (Table 17, columns 10 & 11).

Health workers were also asked about their satisfaction with their jobs, overall. While all three interventions had positive point estimates, none met the 10 percent cut-off for statistical significance. Relative to the pure control, the largest impact was seen in the PBF arm, with health workers 10.5 percent more likely to express satisfaction, followed by the supervision only group (5.3 percent more likely to express satisfaction), followed by the financing only group (4.8 percent more likely) (Table 17, column 12).

HEALTH WORKER AVAILABILITY IN THE HEALTH FACILITY

The facility survey asked the head of the health facility, or the most informed staff member, to list the names of the all of the health workers employed at the health facility. This information was recorded onto the staff roster. The staff roster also collected information on the post occupied by each health worker and on whether they were present on the day of data collection. Table 18,

column 1 presents results from analysis of the number of nurses present at the health facility on the day of data collection. There was a small and non-significant increase in the number of nurses present over the study period. The increase in the number of nurses in the PBF group was greater than in the full control group (p-value = 0.01). Adding the coefficient on the interaction term of PBF and post to the coefficient on the post indicator (0.191+1.222=1.413) indicates that there was an average increase of almost 1.5 nurses present in PBF facilities over the study period. The coefficients on the two other treatment groups – additional financing and additional supervision – were not statistically significant; however, there was a larger increase in the full PBF group compared to the change in the additional supervision group.

DRUGS AND EQUIPMENT IN THE HEALTH FACILITY

A composite indicator was created to assess any impact on the availability of basic clinical equipment. The indicator included information on the presence of a clock, a child weighing scale, height measure, tape measure, adult weighing scale, blood pressure instrument, thermometer, stethoscope, fetoscope, otoscope, flashlight, stretcher, and wheelchair. Scores indicate the proportion of these thirteen pieces of equipment that was available at a given facility and range from 0 to 1. Point estimates indicate the estimated impact on this score.

Both the PBF and the additional financing intervention arms resulted in large and statistically significant improvements in the availability of equipment. Facilities in the PBF arm had a 10.0 percentage point increase over that seen in the control ($p<0.05$), while those in the additional financing arm had an increase of 12.5 percentage points over the control ($p < 0.01$). This increase was not seen in the improved supervision arm; while the point estimate was positive, it was small and not statistically significant. There was no measurable difference in the impacts of the PBF and financing only arms, but there was statistically significant difference between the PBF intervention and improved supervision ($p<0.05$) (Table 18, column 2).

Table 18, column 3 shows the interventions' impact on the availability of vaccination equipment, which includes a thermometer for the vaccine fridge, a cold box or vaccine carrier, a deep freezer, a refrigerator and ice packs. There is no evidence of a differential impact in any of the three arms. The point estimates are all very small, with large standard errors.

Table 18, column 4 shows the impact on delivery equipment, which includes the following: delivery bed, partograph, delivery light, aspirator, newborn resuscitation bag, newborn eye drops or ointment, scissors, umbilical cord clamp or sterile tape/tie, suturing material, examination gloves, sterile cotton gauze, hand soap or detergent, hand scrubbing brush, sterile tray, plastic container with plastic liner for the placenta, plastic container with a plastic liner for medical waste, adult stethoscope, Pinard or fetal stethoscope, blood pressure instrument, kidney basin, protective apron and plastic draw sheet, baby scale, needle holder, syringes and disposable needles, 16- or 18-guage needles, speculum, clamps, hand or foot operated suction pump, vacuum extractor, and a uterine curette. Both the PBF and the additional financing interventions had large and positive impacts on the availability of delivery equipment. Scores in the PBF arm improved by 21 percentage points more than did those in the control and those in the additional financing arm increased by 18.9 percentage points relative to the control. While there was a positive point estimate on the impact of the improved supervision intervention group, the difference (estimated at 8.2 percentage points) does not meet the cut-off for statistical significance. The impacts of the

PBF and financing arms are not statistically significantly different from one another. However, the impact of PBF is significantly larger than that seen in the improved supervision arm ($p<0.10$).

Table 18, column 5 shows the impacts on an index of general medicines, including paracetamol, amoxicillin tabs or syrup, ORS, iron tabs, and cotrimoxazole. While all the point estimates for the intervention arms are positive, they are small relative to their standard errors and no there is no evidence of an impact, relative to either the pure control or one of the other intervention groups.

The same is true for family planning methods, shown in Table 18, column 6. The index is comprised of condoms, oral contraceptive tablets, Depot Medroxyprogesterone Acetate (DMPA), and implants. The estimated impact for PBF, in particular, is large at 16.8 percentage points and statistically significant ($p < 0.10$). The point estimates indicating the effect of the other two different treatment arms are also positive, though they are not large enough to pass the test of statistical significance. Also, the effect in the PBF group was not statistically different from the effects observed in the other treatment groups.

Table 18, column 7 assesses the impact of the interventions on the availability of malaria medicines, including Coartem, ACT lumefantrine artesunate tablet, ACT lumefantrine artesunate syrup, and sulphadoxine-pyrimethamine. There is no evidence of any effect on malaria treatment, as indicated by the very small point estimates and large standard errors. Table 18, column 8 provides an overview of the estimated impact on the availability of vaccines, including Bacille Calmette-Guerin (BCG), Oral Polio Vaccine (OPV), tetanus toxoid, Diptheria Tetanus and Pertussis (DTP), Hepatitis B (HBV), measles, Hemophilus influenza B (Hib), DPT, Hepatitis and Hemophilus influenza (Pentavalent). While the point estimates for both the PBF and the financing arms were positive, indicating an approximately 13 and 11-percentage point increase respectively over the control, neither of these met the 10 percent cut-off for statistical significance. While still positive, the point value for the improved supervision arm was smaller, at 5.3 percentage points, and also failed to meet the cut-off for statistical significance.

THE QUALITY OF CONSULTATIONS FOR CHILDREN UNDER 5 YEARS OF AGE

Enumerators observed a total of 575 child health consultations. For each, enumerators compared the exchange against a standardized checklist and noted whether the health worker performed the following nine routine activities: greeted the patient, washed hands, asked age, duration of the complaint, if the child is able to drink or breastfeed, if the child vomits everything, if the child is lethargic, if the child took any medicine and if the child had diarrhea. An overall quality score was calculated for each visit by calculating the proportion of these activities that was conducted.

As shown in Table 19 column 1, there is no evidence of impact in any of the three intervention groups. Both the PBF and the improved supervision groups have small but negative point estimates with large standard errors and, while the additional financing group has a positive estimated impact (4 percentage points), this too fails to meet the cut-off for statistical significance.

THE QUALITY OF ANTENATAL CARE

Enumerators observed 729 ANC consultations. They compared each exchange against a standardized checklist and noted whether the health worker performed the following eleven

routine activities: took a background,[3] asked about past issues,[4] asked about current issues,[5] provided iron supplementation, gave advice about warning signs,[6] helped to prepare for the birth,[7] checked HIV status, tested for syphilis, provided malaria prophylaxis, discussed appropriate nutrition, and checked the following vital signs: blood pressure, weight, conjunctiva, hemoglobin, rhesus, urine glucose, uterine size, fetal heartbeat and fetal presentation. As with the child health consultations, these data were used to calculate aggregate quality scores. National protocols provide guidelines for care that are specific to the number of prior visits and gestational age of the pregnant women. As not all activities are appropriate for all consultation, the quality indices were adjusted to reflect variations by gestational age (<32 weeks, 32 to 35 weeks, and >35 weeks) and whether or not the patient was experiencing her first pregnancy. Results are presented in Table 19 column 2.

We see a strong positive trend in the quality of ANC over time. On average, consultation scores improved by improved by 12.9 percentage points between baseline and endline ($p<0.05$). As shown in Table 19 column 2, there were no differences in any of the treatment groups in the change in ANC quality relative to the full control.

5. Discussion and Conclusions

In order to distinguish the influence of the different components of the PBF reform, this evaluation compared four arms: (1) the standard PBF package, (2) the same level of financing but not linked to performance, and with the same levels of supervision, monitoring, and autonomy as PBF, (3) no additional resources or autonomy, but the same levels of supervision and monitoring as PBF, and (4) pure comparison.

Overall, the impact evaluation results reveal significant increases in the PBF arm for several indicators (child and mother vaccinations, use of modern family planning), but not for others, such as antenatal care visits and in-facility deliveries. Structural quality as measured by equipment availability, staff presence and staff satisfaction, improved in the PBF group. This finding is consistent with evidence from Malawi showing increases in functional equipment and essential drug stocks for maternal and newborn health services as a result of PBF (Brenner et al., 2017;

[3] A composite score ranging from 0 to 1 indicating whether the worker asked about the patient's age, medicines, and date of last menstruation.

[4] A composite score ranging from 0 to 1 indicating whether the worker asked whether the patient had any prior deliveries, stillbirths, neonatal deaths, abortions, heavy bleeding during or after delivery, or assisted delivery.

[5] A composite score ranging from 0 to 1 indicating whether the worker asked whether the patient had any bleeding, fever, headache or blurred vision, swollen face or hands, tiredness or breathlessness, felt the baby move, or if the client noticed any other symptoms or problems related to the pregnancy.

[6] A composite score ranging from 0 to 1 indicating whether the worker warned the patient to watch for vaginal bleeding, fever, excessive tiredness or breathlessness, swollen hands and face, and severe headache or swollen vision.

[7] A composite score ranging from 0 to 1 indicating whether the worker advised the client to prepare for the birth, including arranging money and transportation, advised to have skilled assistance at delivery, discussed what items to have on hand, emphasized the importance of immunization and the importance of exclusive breastfeeding.

Kambala et al., 2017). However, despite an increase in providers and supplies available at health facilities, PBF did not increase the completeness of service provision during antenatal care and child health consultations. Importantly, out-of-pocket health expenditures decreased for households in the PBF arm, including unofficial payments and this decrease in revenue did not come at the cost of process quality: there were no negative spillover effects on completeness of services and advice provided during antenatal visits and consultations for children under 5. Perhaps not surprisingly, then, given decreased out of pocket costs, and improvements in structural and process quality, client satisfaction also increased for medical consultations for children younger than 5.

While some – but not all - of the improvements measured for PBF were also observed in the additional financing arm C1, few improvements were observed in the group C2 offering enhanced supervision without additional financing or financial incentives. The comparison between the PBF and the C1 group is delicate because the two interventions share many similarities: same supervision and monitoring mechanisms, same level of managerial autonomy and increased financing. The only difference was that in the PBF group (T1) the additional financing was linked to the performance of the individual facility while in the C1 group it was linked to the average performance of the PBF facilities in the same district. It is possible that this distinction might not have been salient enough among the health facility management and staff for them to act upon it and modify their practice, explaining overall similar results.

The lower impacts obtained in the C2 group however suggest that reinforced supervision is not sufficient to change behaviors and improve outcomes. Additional financing appears to be required and its impact seems in some instances stronger when linked to results as in PBF.

This study has several limitations that we have tried to acknowledge in this report. The randomization for this study was at the health facility level. This is beneficial from the point of view of statistical power. From an operational and public health perspective, however, randomizing at the district level would have made more sense given the proximity of some facilities. Indeed, the risk with facility-level randomization is that neighboring facilities from different groups might learn from each other or from supervising staff and apply principles outside their treatment group. However, this was not feasible given that the Government of Cameroon had already decided and announced which districts would be included in the PBF pilot. Randomization at the district level was therefore not an option.

We have analyzed in detail the phenomenon of health care bypassing behavior whereby households look for health care beyond the closest health facility. We found that health care shopping behavior by households was widespread in Cameroon at baseline in 2012 and continues to be widespread at endline in 2015, but does not appear to be a consequence of the introduction of PBF. Overall, the results do not suggest that the health care seeking behavior is driven or even significantly influenced by the introduction of PBF or the other interventions in C1 and C2 limiting the concerns for systematic bias. However, this bypassing behavior likely leads to estimates which are below the true causal effect of the intervention. This is a substantial limitation of the household survey analysis that needs to be kept in mind.

While overall we found that the results from the household survey and the health facility survey analyses were consistent, for example on the absence of impacts for the ANC and skilled delivery

indicators and the presence of positive impact on immunizations, we also noted some discrepancies. In some cases, because of the nature of the information collected, one of the two data collection methods might be superior. For example, patients might be more open about reporting unofficial payments in a household survey than during an exit interview conducted within or in the vicinity of the facility. In contrast, women might be reluctant to report family planning utilization at home and therefore facility-level data for this indicator might be more reliable.

Another potential limitation is that the differences between the three intervention study groups were sometimes subtle. This was certainly the case, as discussed above, between the PBF group and the C1 group that offered all the elements of PBF except the direct link between individual facility performance and additional financing.[8] It is not obvious that all these differences in intervention design have been fully grasped by staff and management. We should also acknowledge that the monitoring of adherence to national guidelines done as part of the monitoring and supervision intervention in T1, C1 and C2 facilities was not ideal from an evaluation point of view because it means that all three treatment groups receive a separate intervention which the control group does not receive. This is obviously not something that could have easily been avoided from an implementation perspective, and it seems likely that the impact of these protocols is small.

Finally, this report relies on quantitative household and health facility surveys. A companion qualitative study has been conducted and its analysis is ongoing. The qualitative analysis will help understand and interpret some of the impacts measured – or their absence – and will shed light on possible mechanisms.

From a policy point of view, these impact evaluation results suggest the following lessons. In general, PBF is an effective mechanism to bring payments and funding at the provider level, leading to significant increases in coverage (child and maternal immunization, family planning, HIV testing) and improvements in structural quality of care. It also leads to a decrease in out-of-pocket payments, in particular unofficial payments. For many of those outcomes, the differences between the PBF group (T1) and the additional financing group (C1) are not significant. It should be noted that the C1 group offered all the elements of PBF except the direct link between individual facility performance and additional financing. It is not obvious these differences in intervention design have been salient enough for staff and management. There was, however, a clear effect for the importance of additional financing plus reinforced supervision through PBF instruments (comparing groups T1 and C1 vs groups C2 and C3). Enhanced supervision and monitoring are not sufficient to improve MCH outcomes.

Given the way the public budget for health in Cameroon is currently organized and allocated, the results from the evaluation suggest that several modifications in the structure and prioritization of public financing would lead to improved health service delivery outcomes in the country. First,

[8] In appendix tables A12-A15, we present results from a specification where we pooled the T1 and C1 group into one group, to see whether this specification – which is not as per the initial evaluation design – would yield different results. Compared to the main tables 4-7, this new specification does not yield qualitatively different results.

the public budget should be distributed more equitably among the different levels of institutions, especially in favor of levels of care that are closest to the user and where cost-effective care is provided, such as primary, preventive, and community health services, enabling them to operate more efficiently and provide more attractive and higher quality services. Second, the use of global budgets for health facilities (by creating a single line for each entity instead of multiple lines for different activities, as is currently the case) should also be considered. To accompany these global budgets for each entity, these health facilities should be empowered for the proper management of their structures and must have the management autonomy to use these resources in order to solve their problems with strategies that are developed locally, for the specific context of each health facility. In order to avoid leakages (the 2009 PETS found that less than 50% of the resources destined to primary care facilities (CSI and CMA) actually arrived at the facility), fund transfers to peripheral-level providers should be completed through a direct transfer to the health facilities' bank accounts, which will avoid the loss of resources along the way.

To provide more equitable allocations of the public budget, Cameroon could consider implementing an intra-regional budget allocation system to deploy resources where the need is greatest, as PBF does with the existing Equity Bonus. According to key principles of the Performance Based Financing (PBF) program (see below), health facilities in the PBF areas retain all their income at the level of their structures and do not transfer a percentage of their revenues to the central level. A method to improve the effectiveness of the allocation and to strengthen the capacity of health service providers to provide high quality care would be to eliminate any transfer of providers' revenues to the central government level, as is currently the case outside PBF zones.

The absence of resources targeted towards high-impact, cost-effective services, may also explain the persistence of poor health outcomes in Cameroon. Using strategic purchasing, as is done in PBF where cost-effective services (prevention, promotion) receive higher per-service subsidies than curative services, could be scaled-up through the public budget to prioritize the implementation of high-impact interventions at health facilities. This approach would require a shift from the funding of these health facilities to a service or performance-based payments, replacing the current system that focuses exclusively on infrastructure-related operating costs (or the appropriations allocated in the previous year).

The absence of impacts for some MCH indicators such as skilled deliveries and ANC visits was surprising. It is possible that the supply-side incentives for providers were not sufficient given existing user fees which might act as a barrier on the demand-side. A policy discussion about combining demand-side and supply-side incentives would be useful. In terms of quality of care, most of the positive impacts were observed on structural quality. However, despite an increase in providers and supplies available at health facilities, PBF did not increase the completeness of service provision (content of care) during antenatal care and child health consultations. Further reflection and efforts should be devoted to identify mechanisms to incentivize or otherwise improve the content of care beyond equipment, supplies and staff availability.

References

Alkema, L., Broaddus, E., Chou, D., Hogan, D., Mathers, C., Moller, A.-B., . . . Zhang, S. (2015). Trends in Maternal Mortality: 1990 to 2015 Estimates by WHO, UNICEF, UNFPA, World Bank Group and the United Nations Population Division In W. H. Organization (Ed.): World Health Organization

Basinga, P., Gertler, P. J., Binagwaho, A., Soucat, A. L. B., Sturdy, J. R., & Vermeersch, C. M. J. (2010). Paying Primary Health Care Centers for Performance in Rwanda. *Policy Research Working Papers*.

Binyaruka, P., Patouillard, E., Powell-Jackson, T., Greco, G., Maestad, O., & Borghi, J. (2015). Effect of Paying for Performance on Utilisation, Quality, and User Costs of Health Services in Tanzania: A Controlled Before and After Study. *PLoS One, 10*(8), e0135013. doi: 10.1371/journal.pone.0135013

Bonfrer, I., Soeters, R., Van de Poel, E., Basenya, O., Longin, G., van de Looij, F., & van Doorslaer, E. (2014). Introduction of performance-based financing in Burundi was associated with improvements in care and quality. *Health Affairs, 33*(12), 2179-2187.

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

Bove, A., Basile., P., Robyn, P. J., & Singh, R. (2013). Towards Greater Equity: A Special Focus on Health. Cameroon Economic Update In T. W. Bank (Ed.). Washington DC: The World Bank.

Brenner, S., Wilhelm, D., Lohmann, J., Kambala, C., Chinkhumba, J., Muula, A. S., & De, M. (2017). Implementation research to improve quality of maternal and newborn health care, Malawi. *Bulletin of the World Health Organization*.

De Walque, D., Robyn, P. J., & Sorgho, G. (2013). Public randomization ceremonies. *The World Bank Development Impact Blog.* Retrieved December 18, 2016, from http://blogs.worldbank.org/impactevaluations/public-randomization-ceremonies

Falisse, J. B., Ndayishimiye, J., Kamenyero, V., & Bossuyt, M. (2015). Performance-based financing in the context of selective free health-care: an evaluation of its effects on the use of primary health-care services in Burundi using routine data. *Health Policy Plan, 30*(10), 1251-1260. doi: 10.1093/heapol/czu132

Fritsche, G. B., Soeters, R., & Meessen, B. (2014). *Performance-based financing toolkit*: World Bank Publications.

Gertler, P. J., & Giovagnoli, P. I. (2014). Rewarding provider performance to enable a healthy start to life: evidence from Argentina's Plan Nacer. *World Bank Policy Research Working Paper*(6884).

Huillery, E., & Seban, J. (2014). *Performance-Based Financing, Motivation and Final Output in the Health Sector: Experimental Evidence from the Democratic Republic of Congo*. Retrieved from http://econ.sciences-po.fr/sites/default/files/file/elise/paper_DRC_July2014_AER.pdf

ICF International. (2012). The DHS Program STATcompiler. Retrieved June 10, 2015, from http://www.statcompiler.com

Institut National de la Statistique. (2015). Enquête par grappes à indicateurs multiples (MICS5), 2014, Rapport de résultats clés. Yaoundé, Cameroun: Institut National de la Statistique.

Institut National de la Statistique (INS) et ICF International. (2012). Enquête Démographique et de Santé et à Indicateurs Multiples du Cameroun 2011. In I. International (Ed.). Calverton, Maryland, USA: ICF International

Ir, P., Korachais, C., Chheng, K., Horemans, D., Van Damme, W., & Meessen, B. (2015). Boosting facility deliveries with results-based financing: a mixed-methods evaluation of

the government midwifery incentive scheme in Cambodia. *BMC Pregnancy Childbirth, 15*, 170. doi: 10.1186/s12884-015-0589-x

Kambala, C., Lohmann, J., Mazalale, J., Brenner, S., Sarker, M., Muula, A. S., & De Allegri, M. (2017). Perceptions of quality across the maternal care continuum in the context of a health financing intervention: Evidence from a mixed methods study in rural Malawi. *BMC health services research, 17*(1), 392.

Ministere de la Sante Publique. (2016). Analyse Situationnelle du financement de la santé au Cameroun: Stratégie de Financement de la Santé 2017-2021. In M. d. l. S. Publique (Ed.). Yaounde, Cameroon.

The United Nations Children's Fund. (2015). Enquête par grappes à indicateurs multiples (MICS5) 2014 Rapport de résultats clés. In T. U. N. C. s. Fund (Ed.), *Multiple Indicator Cluster Survey (MICS)*.

The World Bank. (2013). Cameroon Performance-based Financing: Results from the health facility baseline survey: Results from the health facility baseline survey. In T. W. Bank (Ed.). Washington DC: The World Bank.

The World Bank. (2016). World Development Indicators. from The World Bank

The World Bank. (2017). Cameroon Public Expenditure Review.

United Nations Statistical Division. (2015). Millennium Development Indicators: Country and Regional Progress Snapshots. from http://mdgs.un.org/unsd/mdg/Resources/Static/Products/Progress2015/Snapshots/CMR.pdf

World Health Organization Global Health Expenditure database. (2016). Health expenditure per capita (current US$). (December 8). http://data.worldbank.org/indicator/SH.XPD.PCAP

Zeng, W., Cros, M., Wright, K. D., & Shepard, D. S. (2013). Impact of performance-based financing on primary health care services in Haiti. *Health Policy Plan, 28*(6), 596-605. doi: 10.1093/heapol/czs099

Figure 1: Cameroon PBF project and Impact Evaluation map

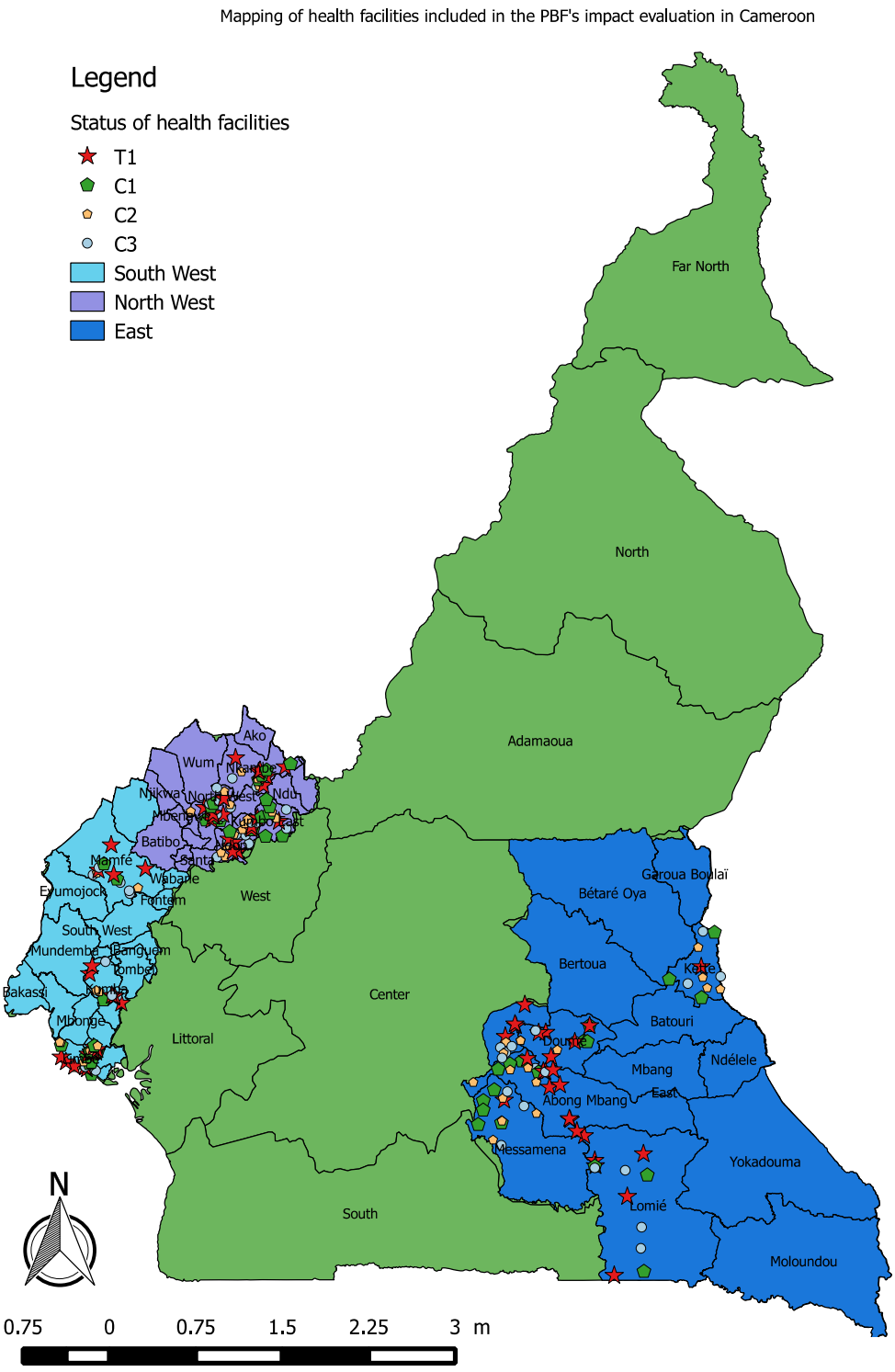


FIGURE 2: PERCENT OF REPORTED PATIENTS CONFIRMED DURING VERIFICATION

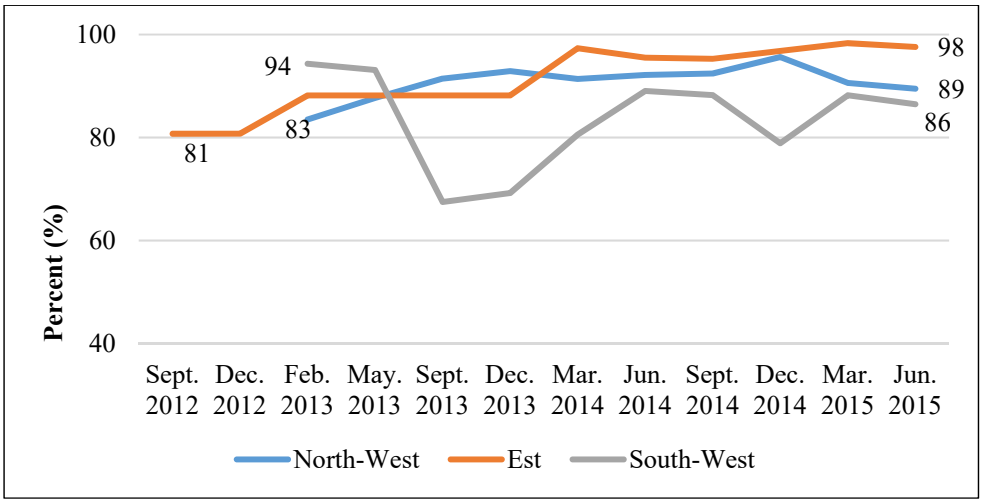


TABLE 1: IMPACT EVALUATION STUDY GROUPS

T1: PBF with health worker performance bonuses	**C1:** Same per capita financial resources as PBF but not linked to performance; Same supervision and monitoring and managerial autonomy as T1
C2: No additional resources but same supervision and monitoring as PBF arms and T1 and C1	**C3:** Status quo

*See Appendix Table 1 for detailed description

TABLE 2: SUMMARY OF HEALTH FACILITIES INCLUDED IN THE IMPACT EVALUATION

District	Number of health facilities						
	CSI Public	CMA Public	District Hospital	Confessional CSI/ CMA/ Hospital	For-profit/ Para-public	Total	Private (%)
Abong-Mbang	14	2	1	4	2	23	27%
Doume	9	1	1	2	1	14	23%
Lomie	7	2	1	2	0	12	18%
Messamena	9	1	1	2	0	13	17%
Nguelemendouka	5	0	1	1	0	7	17%
Kette	9	0	1	0	0	10	0%
Total in East	53	6	6	11	3	79	19%
Kumbo East	17	2	1	6	4	30	34%
Nkambe	11	2	1	4	2	20	32%
Ndop	12	2	1	8	4	27	46%
Fundong	9	3	1	12	3	28	56%
Total in North West	49	9	4	30	13	105	43%
Mamfe	11	1	1	1	0	14	8%
Kumba	10	1	1	5	1	18	35%
Buea	10	3	1	0	9	23	41%
Limbe	10	1	1	1	7	20	42%
Total in South West	41	6	4	7	17	75	34%
Pilot Zone total	**143**	**21**	**14**	**48**	**33**	**259**	**33%**

TABLE 3: OPERATIONAL FINANCING AVAILABLE AT THE HEALTH FACILITY AND PBF PAYMENTS ASSESSED AT ENDLINE

Total (cost recovery, Ministry of health, subsidies)			
	T1 (full PBF)	C1 (increased financing)	p-value
Q1-2014 received	3420226	4163127	0.5281
Q2-2014 received	3339472	4044135	0.5143
Q3-2014 received	3799585	4395996	0.748
Q4-2014 received	3876873	3873040	0.9979
n	53	48	101
Subsidies	T1	C1	p-value
Q1-2014 received	1322834	1725858	0.2009
Q2-2014 received	1241536	1751481	0.1805
Q3-2014 received	1757934	1401405	0.5654
Q4-2014 received	1428642	1264029	0.7897
n	53	48	101
Standardized by number of health workers			
	T1 (full PBF)	C1 (increased financing)	p-value
Q1-2014 received	587480	559007	0.7962
Q2-2014 received	442595	550209	0.2212
Q3-2014 received	436623	471578	0.6895
Q4-2014 received	467895	528527	0.5869
n	49	43	
Subsidies			
Q1-2014 received	295233	391902	0.2906
Q2-2014 received	242738	361116	0.1223
Q3-2014 received	238236	260399	0.7039
Q4-2014 received	212285	312874	0.2159
n	49	43	

TABLE 4: PROVISION OF REPRODUCTIVE AND CHILD HEALTH SERVICES†

Panel A	(1) Skilled delivery	(2) ANC	(3) Tetanus vaccine during pregnancy	(4) Postnatal care	(5) Modern contraception	(6) Third dose of polio vaccine
Post indicator	0.514	3.213	-16.881***	-3.802*	0.679	-5.280***
	[0.798]	[3.932]	[5.563]	[2.175]	[1.002]	[1.983]
PBF/Post interact	1.374	3.007	21.521***	4.309*	9.240***	4.583**
	[1.011]	[7.641]	[6.145]	[2.269]	[2.529]	[2.162]
Control 1/Post interact	1.855*	1.545	15.989**	5.513**	5.794***	2.765
	[1.021]	[5.407]	[6.450]	[2.262]	[1.746]	[2.389]
Control 2/Post interact	0.047	3.993	8.707	3.515	3.321	1.081
	[1.358]	[5.441]	[7.692]	[2.499]	[2.061]	[3.953]
p-value PBF vs. C1	0.581	0.841	0.183	0.190	0.205	0.252
p-value PBF vs. C2	0.289	0.894	0.031	0.570	0.046	0.322
p-value PBF vs. C3	0.176	0.694	0.001	0.059	<0.001	0.035
Baseline mean C3	7.76	20.57	32.84	10.22	3.02	23.90
N	2182	2220	2220	2220	2220	2220

Panel B	(7) Meningitis	(8) Measles	(9) HIV testing	(10) PMTCT	(11) ART
Post indicator	-45.970***	-3.736*	4.239	-3.552	1.021*
	[9.769]	[2.249]	[3.031]	[3.323]	[0.609]
PBF/Post interact	19.041	3.758	61.115***	2.084	-1.455
	[13.471]	[2.552]	[17.817]	[4.011]	[0.888]
Control 1/Post interact	21.931*	1.892	51.466***	2.372	-0.671
	[11.131]	[2.700]	[13.668]	[3.189]	[0.573]
Control 2/Post interact	8.47	-0.740	6.596	1.648	-0.681
	[13.547]	[3.546]	[5.757]	[3.156]	[0.595]
p-value PBF vs. C1	0.753	0.337	0.656	0.905	0.235
p-value PBF vs. C2	0.387	0.135	0.003	0.851	0.252
p-value PBF vs. C3	0.159	0.143	0.001	0.604	0.103
Baseline mean C3	46.65	20.90	9.98	9.86	0.012
N	2220	2220	2220	2220	2220

* = p < 0.10, ** p < 0.05, *** p< 0.01

† Results from difference-in-differences regression models examining the effect of PBF on reproductive health service provision reported in facility registers. Monthly number of services provided during the six months before the baseline and endline surveys used as the dependent variable. Regression models adjusted for facility controls (type of health facility public/private/religious, urban/rural). Standard errors were clustered at the health facility level.

TABLE 5: COVERAGE OF REPRODUCTIVE HEALTH SERVICES† AND PROVISION OF MODERN FAMILY PLANNING‡					
	(1)	(2)	(3)	(4)	(5)
	Skilled delivery	At least two ANC visits	Tetanus vaccine during pregnancy	Postnatal care	Modern contraception
Post indicator	0.053***	0.022	0.001	0.105***	0.002
	[0.019]	[0.014]	[0.019]	[0.031]	[0.044]
PBF/Post interact	-0.043	0.010	0.024	-0.029	-0.037
	[0.028]	[0.020]	[0.023]	[0.041]	[0.054]
Control 1/Post interact	0.020	-0.024	0.003	-0.019	-0.054
	[0.032]	[0.019]	[0.025]	[0.041]	[0.055]
Control 2/Post interact	-0.050*	-0.044**	0.01	-0.070*	0.000
	[0.029]	[0.019]	[0.023]	[0.039]	[0.053]
p-value PBF vs. C1	0.055	0.111	0.369	0.798	0.731
p-value PBF vs. C2	0.828	0.010	0.520	0.277	0.429
p-value PBF vs. C3	0.117	0.617	0.306	0.484	0.486
Baseline mean C3	0.784	0.894	0.878	0.323	0.180
N	5858	5974	5975	5966	4498

* = $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

† Results from difference-in-differences regression models examining the effect of PBF on reproductive health service use among female respondents included in the household survey who had been pregnant in the previous 24 months. ‡ Results from difference-in-differences regression models examining the effect of PBF on modern contraceptive use among female respondents of reproductive age (15 – 49) included in the household survey. Regression models adjusted for individual (age, marital status, education level, religion, ethnicity, working status and type of work) and household control variables (number of individuals in the household, housing type, house ownership, water source, and type of sanitation). Standard errors were clustered at the health facility level.

TABLE 6: FULL VACCINATION COVERAGE, GROWTH MONITORING, BEDNET USE, STUNTING, UNDERWEIGHT AND WASTING AMONG CHILDREN†

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Fully vaccinated documented by vaccine card	Fully vaccinated by vaccine card or self-report	Growth monitoring in the last month	Slept under a bednet	Stunting	Under-weight	Wasted
Post indicator	0.127*	0.108**	-0.014	-0.181***	-0.008	-0.010	0.047**
	[0.072]	[0.052]	[0.013]	[0.025]	[0.025]	[0.022]	[0.021]
PBF/Post interact	0.170*	0.164**	-0.002	0.001	0.008	0.046	-0.004
	[0.095]	[0.069]	[0.017]	[0.042]	[0.033]	[0.028]	[0.028]
Control 1/Post interact	-0.054	-0.015	0.031*	-0.005	0.010	0.043	-0.029
	[0.092]	[0.065]	[0.017]	[0.038]	[0.037]	[0.032]	[0.029]
Control 2/Post interact	0.018	0.029	0.022	0.003	0.037	0.018	-0.028
	[0.092]	[0.073]	[0.019]	[0.036]	[0.033]	[0.028]	[0.027]
p-value PBF vs. C1	0.009	0.003	0.047	0.893	0.957	0.907	0.381
p-value PBF vs. C2	0.075	0.052	0.215	0.967	0.338	0.272	0.328
p-value PBF vs. C3	0.076	0.019	0.930	0.979	0.810	0.104	0.876
Baseline mean C3	0.599	0.645	0.048	0.809	0.444	0.147	0.067
N	1569	2448	7055	10107	8711	8672	8480

* = p < 0.10, ** p < 0.05, *** p< 0.01

† Results from difference-in-differences regression models examining the effect of PBF on child vaccination among children (12 – 23 months), growth monitoring in the last month among children (12 – 59 months), having slept under a bednet the night before the survey and on child anthropometric outcomes (stunting, underweight and wasting) among children under 5 years of age included in the household survey. Regression model adjusted for individual (age, father in the household, religion, ethnicity) and household control variables (number of individuals in the household, housing type, house ownership, water source, and type of sanitation). Standard errors were clustered at the health facility level.

Table 7: Health care spending as reported in household data†				
	(1) Official provider fee	(2) Unofficial provider fee	(3) Lab and x-ray fees	(4) Transportation fees
Post indicator	1811.58	2052.12*	1048.64	123.03
	[1475.25]	[1057.18]	[711.54]	[201.09]
PBF/Post interact	-1495.83	-2254.12*	-1473.44*	-455.41
	[1538.26]	[1305.64]	[779.60]	[288.36]
Control 1/Post interact	-334.73	-2736.04	-521.02	-495.14**
	[1506.73]	[1778.02]	[868.01]	[241.38]
Control 2/Post interact	-1378.05	-1422.67	-639.27	-368.79
	[3969.75]	[1244.33]	[885.00]	[236.41]
p-value PBF vs. C1	0.191	0.750	0.051	0.880
p-value PBF vs. C2	0.974	0.392	0.128	0.732
p-value PBF vs. C3	0.332	0.086	0.060	0.116
Baseline mean C3	1689.22	2183.33	1603.09	910.30
N	2374	2261	2292	2365

* = p < 0.10, ** p < 0.05, *** p< 0.01

† Results from difference-in-differences regression models examining the effect of PBF on health care spending in the last 4 weeks among respondents in the household survey. Regression model adjusted for individual (age, sex) and household control variables (number of individuals in the household, housing type, house ownership, water source, and type of sanitation). Standard errors were clustered at the health facility level.

TABLE 8: HEALTH SPENDING FOR ANC AND CHILD HEALTH CONSULTATIONS								
	ANC†				Child health consultations‡			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Official provider fee	Unofficial provider fee	Medicines fees	Total fees	Official provider fee	Unofficial provider fee	Medicines fees	Total fees
Post indicator	472.17	-217.86	-695.08	319.34	-14.37	-34.34	227.67	403.6
	[438.97]	[199.56]	[647.31]	[1608.44]	[57.05]	[20.87]	[559.02]	[887.20]
PBF/Post interact	-1025.34*	136.57	701.7	2501.29	79.17	36.22	679.38	1545.01
	[585.71]	[231.57]	[708.71]	[2637.04]	[133.41]	[26.31]	[729.58]	[1282.14]
Control 1/Post interact	1824.81**	312.76	1260.72	4445.44*	43.76	12.05	442.13	636
	[867.46]	[278.58]	[825.5]	[2460.9]	[113.03]	[32.24]	[819.08]	[1283.09]
Control 2/Post interact	-67.28	203.69	2374.42	5178.78**	53.25	-22.76	14.41	731.23
	[483.70]	[191.43]	[1813.13]	[2560.66]	[89.29]	[60.83]	[786.80]	[1105.73]
p-value PBF vs. C1	0.001	0.373	0.392	0.490	0.814	0.400	0.758	0.485
p-value PBF vs. C2	0.015	0.555	0.337	0.351	0.862	0.307	0.408	0.499
p-value PBF vs. C3	0.058	0.556	0.324	0.354	0.554	0.171	0.353	0.230
Baseline mean C3	604.91	232.79	1881.96	5239.51	286.79	85.57	2105.00	2921.51
N	725	730	652	724	613	612	556	609

* = p < 0.10, ** p < 0.05, *** p< 0.01

† Results from difference-in-differences regression models examining the effect of PBF on spending for antenatal care among respondents to antenatal care exit interviews. Regression model adjusted for individual (age, literacy, education level, and marital status) and facility variables (type of health facility, urban/rural). Standard errors were clustered at the health facility level. ‡ Results from difference-in-differences regression models examining the effect of PBF on spending for child health consultations among respondents to child health care exit interviews. Regression model adjusted for individual (age, literacy, education level, and marital status) and facility variables (type of health facility, urban/rural). Standard errors were clustered at the health facility level in all regressions.

TABLE 9: SAMPLE CHARACTERISTICS OF WOMEN INCLUDED IN THE ANTENATAL CARE EXIT INTERVIEWS AS BASELINE*

	Mean T1	Mean C1	Mean C2	Mean C3	Mean total	p-value T1/C3	p-value C1/C3	p-value C2/C3	p-value F-test	N
Age	25.82	24.95	25.26	24.93	25.25	0.33	0.99	0.78	0.71	258
Currently married	0.74	0.81	0.83	0.79	0.79	0.57	0.73	0.53	0.77	258
Literate	0.82	0.72	0.70	0.65	0.72	0.05	0.51	0.64	0.21	256
No education	0.08	0.13	0.10	0.10	0.10	0.68	0.63	0.96	0.84	258
Primary education	0.55	0.39	0.51	0.43	0.47	0.25	0.73	0.42	0.49	258
Secondary education	0.33	0.27	0.33	0.38	0.33	0.62	0.32	0.63	0.80	258
Secondary education level 2	0.05	0.06	0.04	0.08	0.06	0.36	0.67	0.37	0.76	258
Higher education	0.00	0.15	0.01	0.02	0.04	0.30	0.05	0.93	0.07	258

* Standard errors adjusted for facility-level clustering of observations

TABLE 10: SATISFACTION WITH ANTENATAL CARE CONSULTATIONS REPORTED DURING FACILITY EXIT INTERVIEWS†

Panel A	(1) Overall satisfaction score	(2) Reasonable registration fees	(3) Reasonable lab fees	(4) Reasonable medicine fees	(5) No additional payment	(6) Clean health facility	(7) Reasonable wait time
Post indicator	0.006	0.055	0.074	0.088	0.010	-0.045	-0.009
	[0.034]	[0.097]	[0.075]	[0.090]	[0.060]	[0.075]	[0.079]
PBF/Post interact	0.086*	0.037	0.154	0.190	-0.043	0.241**	0.161
	[0.048]	[0.128]	[0.113]	[0.134]	[0.078]	[0.111]	[0.115]
Control 1/Post interact	0.051	-0.051	-0.051	-0.069	0.067	0.228**	0.014
	[0.044]	[0.142]	[0.129]	[0.121]	[0.092]	[0.106]	[0.129]
Control 2/Post interact	-0.019	-0.085	-0.027	-0.127	-0.020	0.002	-0.055
	[0.049]	[0.127]	[0.127]	[0.118]	[0.087]	[0.111]	[0.134]
p-value PBF vs. C1	0.419	0.523	0.129	0.044	0.230	0.903	0.269
p-value PBF vs. C2	0.036	0.309	0.185	0.015	0.776	0.040	0.127
p-value PBF vs. C3	0.077	0.774	0.176	0.158	0.586	0.032	0.163
Baseline mean C3	0.853	0.804	0.782	0.754	0.885	0 .787	0 .738
N	730	669	665	689	723	730	727

Panel B	(8) Enough privacy during visit	(9) Adequate hours	(10) Courteous health staff	(11) Good health worker communication	(12) Sufficient visit time with health worker	(13) Easy to get prescribed medicines
Post indicator	-0.025	-0.026	-0.041	-0.031	0.055	0.018
	[0.061]	[0.041]	[0.036]	[0.024]	[0.052]	[0.068]
PBF/Post interact	0.042	0.154**	0.037	0.106**	-0.045	0.000
	[0.086]	[0.071]	[0.063]	[0.050]	[0.074]	[0.079]
Control 1/Post interact	0.149	0.032	0.070	0.039	0.139*	0.041
	[0.093]	[0.055]	[0.053]	[0.052]	[0.080]	[0.085]
Control 2/Post interact	-0.010	-0.079	0.019	0.101	0.030	-0.040
	[0.098]	[0.062]	[0.051]	[0.077]	[0.101]	[0.075]
p-value PBF vs. C1	0.254	0.073	0.599	0.308	0.026	0.488
p-value PBF vs. C2	0.605	0.002	0.783	0.961	0.443	0.402
p-value PBF vs. C3	0.629	0.033	0.556	0.038	0.544	0.997
Baseline mean C3	0 .902	0.900	0 .967	0 .951	0 .852	0 .883
N	728	724	730	725	728	716

* = p < 0.10, ** p < 0.05, *** p< 0.01

† Results from difference-in-differences regression models examining the effect of PBF on satisfaction with antenatal care components reported by patients during facility exit interviews. Regression models adjusted for individual (age, literacy, marital status, education level) and facility-level control variables (type of health facility public/private/religious, urban/rural). Standard errors were clustered at the health facility level.

TABLE 11: SAMPLE CHARACTERISTICS OF CHILDREN AND MOTHERS INCLUDED IN CHILD HEALTH CONSULTATION EXIT INTERVIEWS AT BASELINE*

	Mean T1	Mean C1	Mean C2	Mean C3	Mean total	p-value T1/C3	p-value C1/C3	p-value C2/C3	p-value F-test	N
Child characteristics										
Age in months	23.66	14.15	22.45	18.00	20.01	0.08	0.27	0.17	0.04	185
Female	0.58	0.40	0.54	0.49	0.51	0.33	0.43	0.65	0.48	188
Caretaker characteristics										
Single	0.35	0.20	0.26	0.16	0.25	0.04	0.69	0.27	0.18	187
Currently married	0.60	0.77	0.74	0.80	0.73	0.03	0.79	0.51	0.16	187
Divorced or widowed	0.04	0.03	0.00	0.04	0.03	0.97	0.78	0.16	0.16	187
Literate	0.75	0.77	0.69	0.77	0.74	0.84	0.98	0.33	0.77	190
No education	0.06	0.11	0.17	0.08	0.11	0.74	0.66	0.16	0.34	187
Primary education	0.38	0.37	0.43	0.38	0.39	0.96	0.94	0.59	0.93	187
Secondary education level 1	0.27	0.26	0.30	0.38	0.30	0.26	0.25	0.36	0.59	187
Secondary education level 2	0.21	0.17	0.07	0.12	0.14	0.20	0.52	0.43	0.24	187
Higher education	0.08	0.09	0.04	0.04	0.06	0.40	0.46	0.94	0.71	187

* Standard errors adjusted for facility-level clustering of observations

TABLE 12: SATISFACTION WITH CHILD HEALTH CONSULTATIONS REPORTED DURING FACILITY EXIT INTERVIEWS†							
Panel A	(1) Overall satisfaction score	(2) Reasonable registration fees	(3) Reasonable lab fees	(4) Reasonable medicine fees	(5) No additional payment	(6) Clean health facility	(7) Reasonable wait time
Post indicator	-0.036	-0.054	-0.143	-0.045	-0.004	-0.141	-0.111**
	[0.026]	[0.061]	[0.118]	[0.067]	[0.055]	[0.099]	[0.055]
PBF/Post interact	0.099***	0.112	0.347**	0.043	-0.007	0.227*	0.110
	[0.037]	[0.101]	[0.175]	[0.124]	[0.117]	[0.133]	[0.097]
Control 1/Post interact	0.054	0.019	0.331*	0.111	-0.020	0.136	0.143
	[0.040]	[0.076]	[0.167]	[0.103]	[0.077]	[0.118]	[0.097]
Control 2/Post interact	0.022	0.074	0.420**	0.033	-0.055	-0.049	0.021
	[0.045]	[0.080]	[0.166]	[0.120]	[0.090]	[0.131]	[0.100]
p-value PBF vs. C1	0.280	0.325	0.925	0.595	0.908	0.403	0.777
p-value PBF vs. C2	0.092	0.709	0.685	0.945	0.696	0.019	0.442
p-value PBF vs. C3	0.009	0.268	0.050	0.731	0.953	0.090	0.259
Baseline mean C3	0.881	0.957	0.846	0.854	0.904	0.868	0.943
N	614	488	369	544	605	612	608
Panel B	(8) Enough privacy during visit	(9) Adequate hours	(10) Courteous health staff	(11) Good health worker communication	(12) Sufficient visit time with health worker	(13) Easy to get prescribed medicines	
Post indicator	-0.098	-0.046	0.062	0.017	0.091	-0.002	
	[0.098]	[0.052]	[0.062]	[0.073]	[0.066]	[0.059]	
PBF/Post interact	0.336***	0.085	-0.012	0.053	-0.094	0.055	
	[0.124]	[0.068]	[0.079]	[0.092]	[0.087]	[0.082]	
Control 1/Post interact	0.202	0.036	-0.080	-0.080	-0.018	0.068	
	[0.131]	[0.068]	[0.077]	[0.105]	[0.111]	[0.087]	
Control 2/Post interact	0.093	-0.116	-0.101	0.055	-0.031	0.112	
	[0.115]	[0.081]	[0.082]	[0.103]	[0.104]	[0.106]	
p-value PBF vs. C1	0.256	0.427	0.323	0.170	0.461	0.890	
p-value PBF vs. C2	0.021	0.007	0.214	0.988	0.507	0.594	
p-value PBF vs. C3	0.007	0.210	0.876	0.566	0.279	0.498	
Baseline mean C3	0.774	0.942	0.887	0.830	0.830	0.925	
N	612	608	613	606	609	610	

* = p < 0.10, ** p < 0.05, *** p< 0.01

† Results from difference-in-differences regression models examining the effect of PBF on satisfaction with child health consultations reported by mothers during facility exit interviews. Regression models adjusted for individual (child age, child sex, maternal literacy, marital status, education level) and facility-level control variables (type of health facility public/private/religious, urban/rural). Standard errors were clustered at the health facility level.

TABLE 13: SAMPLE CHARACTERISTICS OF HEALTH WORKERS IN STUDY HEALTH FACILITIES AT BASELINE*

	Mean T1	Mean C1	Mean C2	Mean C3	Mean total	p-value T1/C3	p-value C1/C3	p-value C2/C3	p-value F-test	N
Provider age	38.72	41.31	37.18	38.67	38.96	0.98	0.11	0.36	0.14	434
Years employed at facility	5.43	4.50	4.35	6.24	5.13	0.39	0.06	0.05	0.14	428
Provider sex	0.66	0.68	0.66	0.69	0.67	0.75	0.94	0.68	0.97	434
Primary education	0.23	0.15	0.13	0.22	0.18	0.86	0.19	0.08	0.15	434
Secondary education	0.31	0.36	0.44	0.44	0.39	0.04	0.25	0.92	0.12	434
Secondary education level 2	0.40	0.41	0.34	0.30	0.36	0.15	0.12	0.53	0.33	434
Higher education	0.06	0.07	0.06	0.04	0.06	0.46	0.28	0.41	0.69	434
Employed by MOH	0.45	0.44	0.49	0.45	0.46	0.98	0.91	0.67	0.95	433
Religious employer	0.13	0.25	0.20	0.17	0.19	0.61	0.38	0.71	0.49	433
Employed by facility	0.19	0.09	0.15	0.20	0.16	0.88	0.12	0.47	0.39	433
Other employer	0.23	0.21	0.16	0.18	0.20	0.48	0.63	0.73	0.68	433

* Standard errors adjusted for facility-level clustering of observations

TABLE 14: WORLD HEALTH ORGANIZATION WELL-BEING INDEX†					
	(1) Happy and in a good mood in the last 2 weeks	(2) Calm and relaxed in the last 2 weeks	(3) Active and energetic in the last 2 weeks	(4) Refreshed and rested in the morning in the last 2 weeks	(5) Days filled with interesting things in the last 2 weeks
Post indicator	-0.025	-0.128	0.055	-0.069	0.081
	[0.057]	[0.084]	[0.063]	[0.090]	[0.086]
PBF/Post interact	0.044	0.016	-0.117	-0.053	-0.157
	[0.082]	[0.108]	[0.074]	[0.108]	[0.112]
Control 1/Post interact	-0.009	0.094	0.022	0.134	0.062
	[0.087]	[0.114]	[0.079]	[0.113]	[0.116]
Control 2/Post interact	-0.039	0.037	-0.123	0.058	-0.096
	[0.080]	[0.113]	[0.076]	[0.120]	[0.106]
p-value PBF vs. C1	0.569	0.451	0.025	0.037	0.047
p-value PBF vs. C2	0.332	0.835	0.906	0.254	0.528
p-value PBF vs. C3	0.592	0.885	0.117	0.624	0.163
Baseline mean C3	0.816	0.684	0.776	0.643	0.582
N	991	991	990	991	991

* = p < 0.10, ** p < 0.05, *** p< 0.01

† Results from difference-in-differences regression models examining the effect of PBF on WHO well-being index items reported by health workers. Regression models adjusted for individual (age, sex, marital status, education level) and facility-level control variables (type of health facility public/private/religious, urban/rural). Standard errors were clustered at the health facility level.

TABLE 15: INTERNAL AND EXTERNAL WORKING RELATIONSHIPS †							
	(1) Working relationships with District/ Ministry of Health staff	(2) Working relationships with other facility staff	(3) Working relationships with Management staff within the health facility	(4) Collaboration with the Regional Health Delegation	(5) Quality of the management of the health facility by the management staff within the health facility	(6) The relationship between the health facility and local traditional leaders	(7) Your level of respect in the community
Post indicator	-0.038	0.074	0.034	0.172	0.057	0.030	0.001
	[0.085]	[0.087]	[0.072]	[0.148]	[0.084]	[0.089]	[0.053]
PBF/Post interact	0.103	-0.029	-0.172*	-0.006	-0.089	-0.002	-0.034
	[0.106]	[0.104]	[0.093]	[0.178]	[0.123]	[0.104]	[0.065]
Control 1/Post	0.127	-0.049	0.003	0.054	0.063	0.093	0.002
interact	[0.109]	[0.105]	[0.096]	[0.182]	[0.120]	[0.118]	[0.074]
Control 2/Post	-0.003	-0.129	-0.184*	-0.162	-0.057	-0.074	-0.027
interact	[0.125]	[0.106]	[0.095]	[0.198]	[0.129]	[0.112]	[0.069]
p-value PBF vs. C1	0.800	0.818	0.045	0.677	0.199	0.320	0.568
p-value PBF vs. C2	0.342	0.259	0.895	0.332	0.808	0.392	0.899
p-value PBF vs. C3	0.333	0.780	0.067	0.973	0.471	0.981	0.604
Baseline mean C3	0.793	0.763	0 .758	0.475	0.591	0.648	0.847
N	840	946	938	655	961	908	987

* = p < 0.10, ** p < 0.05, *** p< 0.01

† Results from difference-in-differences regression models examining the effect of PBF on internal and external relationships. Regression models adjusted for individual (age, sex, marital status, education level) and facility-level control variables (type of health facility public/private/religious, urban/rural). Standard errors were clustered at the health facility level.

TABLE 16: HEALTH WORKER SATISFACTION WITH SUPPLY AVAILABILITY AND PHYSICAL CONDITION OF HEALTH FACILITIES†

	(1) Quantity of medicine available in the health facility	(2) Quality of medicine available in the health facility	(3) Quantity of equipment in the health facility	(4) Quality and physical condition of equipment in the health facility	(5) Availability of other supplies in the health facility	(6) The physical condition of the health facility building	(7) Your ability to provide high quality of care given the current working conditions in the facility
Post indicator	0.092	0.070	0.032	0.022	-0.032	-0.084	0.069
	[0.081]	[0.067]	[0.060]	[0.071]	[0.100]	[0.078]	[0.074]
PBF/Post interact	0.071	0.001	0.190**	0.256**	0.404***	0.306***	-0.009
	[0.114]	[0.096]	[0.095]	[0.109]	[0.120]	[0.111]	[0.097]
Control 1/Post	0.176	0.050	0.210**	0.247**	0.332***	0.106	0.123
interact	[0.111]	[0.096]	[0.090]	[0.101]	[0.121]	[0.099]	[0.103]
Control 2/Post	0.025	-0.058	0.122	0.080	0.170	0.096	-0.129
interact	[0.119]	[0.104]	[0.094]	[0.107]	[0.131]	[0.118]	[0.116]
p-value PBF vs.	0.340	0.602	0.845	0.931	0.455		
C1						0.036	0.184
p-value PBF vs.	0.701	0.568	0.512	0.124	0.034		
C2						0.074	0.287
p-value PBF vs.	0.536	0.990	0.048	0.020	0.001		
C3						0.006	0.926
Baseline mean C3	0.505	0.763	0.196	0.278	0.531	0 .449	0.526
N	960	984	988	987	982	990	988

* = p < 0.10, ** p < 0.05, *** p< 0.01

† Results from difference-in-differences regression models examining the effect of PBF on the quantity and quality of health supplies, medicine and equipment in the health facility, the physical condition of the health facility and ability to provide high quality care given health facility conditions reported by health workers. Regression models adjusted for individual (age, sex, marital status, education level) and facility-level control variables (type of health facility public/private/religious, urban/rural). Standard errors were clustered at the health facility level.

TABLE 17: HEALTH WORKER SATISFACTION WITH FINANCIAL AND NON-FINANCIAL BENEFITS†						
Panel A	(1) Your salary	(2) Your benefits (such as housing, travel allowance, bonus including performance bonus, etc.)	(3) Living accommodations	(4) Your opportunity to discuss work issues with your immediate supervisor	(5) Your immediate supervisor's recognition of your good work	(6) Your opportunity to be rewarded for hard work, financially or otherwise
Post indicator	0.036	-0.048	0.128	0.054	-0.008	0.066
	[0.036]	[0.059]	[0.084]	[0.067]	[0.077]	[0.063]
PBF/Post interact	0.091	0.183**	0.138	-0.081	-0.034	-0.019
	[0.061]	[0.075]	[0.109]	[0.095]	[0.098]	[0.107]
Control 1/Post interact	0.134*	0.287***	0.096	0.004	0.066	0.185*
	[0.069]	[0.083]	[0.109]	[0.097]	[0.101]	[0.102]
Control 2/Post interact	0.053	0.037	0.159	-0.061	-0.016	-0.113
	[0.061]	[0.098]	[0.117]	[0.103]	[0.107]	[0.115]
p-value PBF vs. C1	0.572	0.170	0.675	0.386	0.249	0.096
p-value PBF vs. C2	0.587	0.113	0.845	0.848	0.841	0.477
p-value PBF vs. C3	0.138	0.016	0.207	0.396	0.729	0.861
Baseline mean C3	0.055	0.133	0.299	0.663	0.765	0.302
N	943	862	972	980	975	971
Panel B	(7) Your opportunities for promotion	(8) Your opportunities to upgrade your skills and knowledge through training	(9) The opportunities to use your skills in your job	(10) Safety and security in the community	(11) Available schooling for your children	(12) Overall, how satisfied are you with your job?
Post indicator	0.137**	0.116	-0.013	0.093	-0.026	0.183**
	[0.062]	[0.082]	[0.072]	[0.083]	[0.115]	[0.085]
PBF/Post interact	-0.113	-0.025	0.007	0.139	0.027	0.105
	[0.092]	[0.108]	[0.094]	[0.106]	[0.160]	[0.113]
Control 1/Post interact	0.085	0.050	0.194*	-0.036	-0.006	0.048
	[0.087]	[0.115]	[0.103]	[0.102]	[0.171]	[0.113]
Control 2/Post interact	-0.251**	-0.075	0.007	-0.003	-0.044	0.053
	[0.100]	[0.121]	[0.118]	[0.107]	[0.172]	[0.114]
p-value PBF vs. C1	0.033	0.497	0.047	0.045	0.841	0.608
p-value PBF vs. C2	0.189	0.666	0.999	0.124	0.672	0.639
p-value PBF vs. C3	0.222	0.819	0.937	0.189	0.866	0.355
Baseline mean C3	0.152	0.309	0.694	0.619	0.347	0.337
N	918	967	989	984	726	986

* = p < 0.10, ** p < 0.05, *** p< 0.01

† Results from difference-in-differences regression models examining the effect of PBF on health worker satisfaction with financial and non-financial benefits. Regression models adjusted for individual (age, sex, marital status, education level) and facility-level control variables (type of health facility public/private/religious, urban/rural). Standard errors were clustered at the health facility level.

TABLE 18: NURSES, BASIC CLINICAL EQUIPMENT AND MEDICINES AVAILABLE AT HEALTH FACILITIES†				
	(1) Number of nurses present	(2) Basic clinical equipment	(3) Vaccination equipment	(4) Delivery equipment
Post indicator	0.191	0.030	0.102**	0.016
	[0.344]	[0.035]	[0.046]	[0.047]
PBF/Post interact	1.222***	0.100**	-0.013	0.209***
	[0.468]	[0.043]	[0.063]	[0.064]
Control 1/Post interact	0.738	0.125***	0.021	0.189***
	[0.475]	[0.043]	[0.060]	[0.060]
Control 2/Post interact	-0.172	0.024	-0.037	0.082
	[0.475]	[0.044]	[0.070]	[0.070]
p-value PBF vs. C1	0.291	0.488	0.563	0.729
p-value PBF vs. C2	0.003	0.043	0.724	0.061
p-value PBF vs. C3	0.010	0.021	0.842	0.001
Baseline mean C3	2.725	0.679	0.702	0.535
N	369	370	370	370
	(5) General medicines	(6) Family planning methods	(7) Malaria treatment medicines	(8) Vaccines available
Post indicator	0.045	-0.105*	-0.048	-0.116**
	[0.039]	[0.063]	[0.063]	[0.057]
PBF/Post interact	0.064	0.168*	-0.014	0.131
	[0.058]	[0.091]	[0.080]	[0.088]
Control 1/Post interact	0.089	0.078	-0.028	0.113
	[0.063]	[0.097]	[0.083]	[0.091]
Control 2/Post interact	0.050	0.100	0.021	0.053
	[0.064]	[0.097]	[0.093]	[0.093]
p-value PBF vs. C1	0.701	0.361	0.844	0.850
p-value PBF vs. C2	0.830	0.491	0.672	0.432
p-value PBF vs. C3	0.270	0.067	0.864	0.136
Baseline mean C3	0.768	0.482	0 .646	0.530
N	370	370	370	370

* = p < 0.10, ** p < 0.05, *** p< 0.01

† Results from difference-in-differences regression models examining the effect of PBF on the number of nurses present on the day of the survey, basic clinical equipment and medicines available at the health facility. Regression model adjusted for facility-level control variables (type of health facility public/private/religious, urban/rural). Standard errors were clustered at the health facility level.

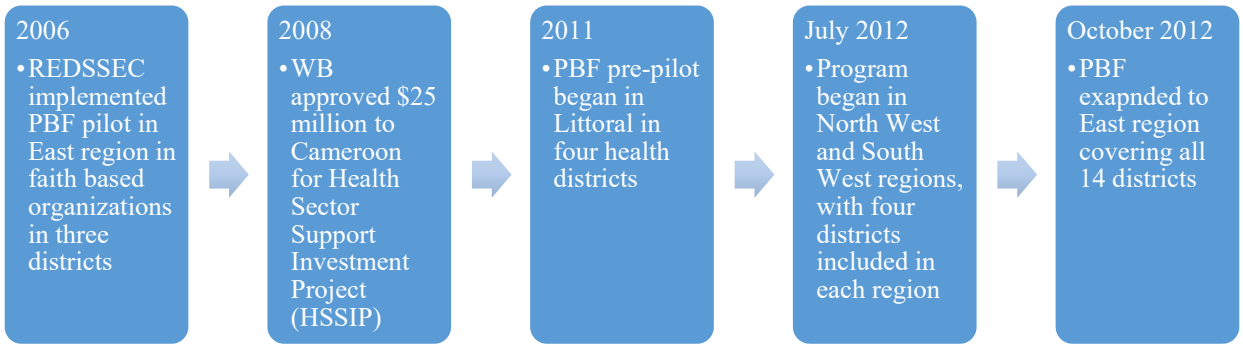
TABLE 19: OVERALL QUALITY SCORE OF CHILD HEALTH† AND ANC CONSULTATIONS‡		
	(1)	(2)
	Child health consultations	ANC consultations
Post indicator	0.030	0.129***
	[0.041]	[0.029]
PBF/Post interact	-0.021	-0.056
	[0.055]	[0.042]
Control 1/Post interact	0.044	0.015
	[0.056]	[0.045]
Control 2/Post interact	-0.029	-0.042
	[0.058]	[0.042]
p-value PBF vs. C1	0.230	0.131
p-value PBF vs. C2	0.888	0.742
p-value PBF vs. C3	0.705	0.191
Baseline mean in C3	0.511	0.592
N	575	729

* = p < 0.10, ** p < 0.05, *** p< 0.01

† Results from difference-in-differences regression models examining the effect of PBF on the overall quality score from child health consultations from direct observation of child health consultations. Regression model adjusted for individual child-level variables (age, sex), maternal variables (marital status, education level) and facility-level control variables (type of health facility public/private/religious, urban/rural). ‡Results from difference-in-differences regression models examining the effect of PBF on the overall quality score from antenatal care consultations from direct observation. Regression model adjusted for individual variables (age, literacy, education, marital status), and facility-level control variables (type of health facility public/private/religious, urban/rural). In addition to the standard controls, also controlled for whether it is the first pregnancy (Y/N) and where in the pregnancy the woman is (<32 weeks, 32-35 weeks, >35 weeks). Standard errors were clustered at the health facility level.

APPENDIX

FIGURE A1: TIMELINE OF PBF IMPLEMENTATION IN CAMEROON



REGIONAL HEALTH FACILITY MAPS

FIGURE A2

Mapping of health facilities included in the PBF's impact evaluation in the East

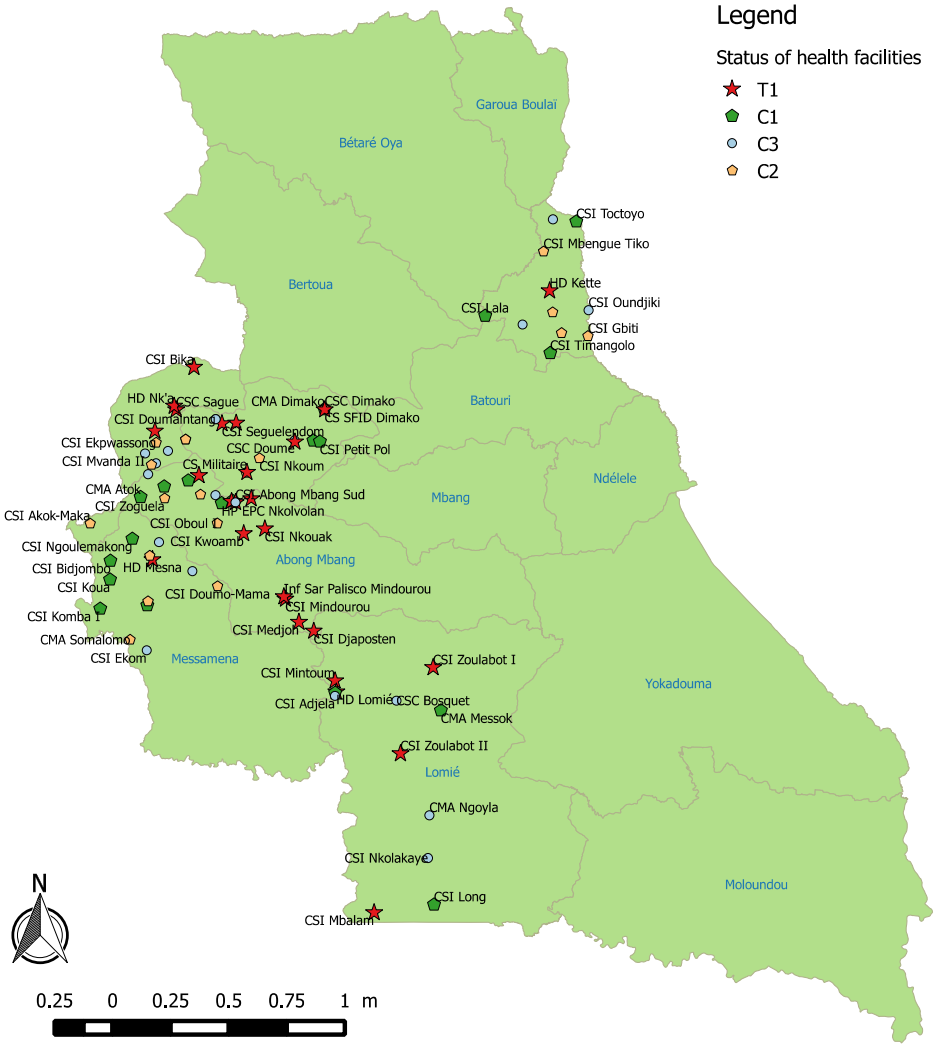


FIGURE A3

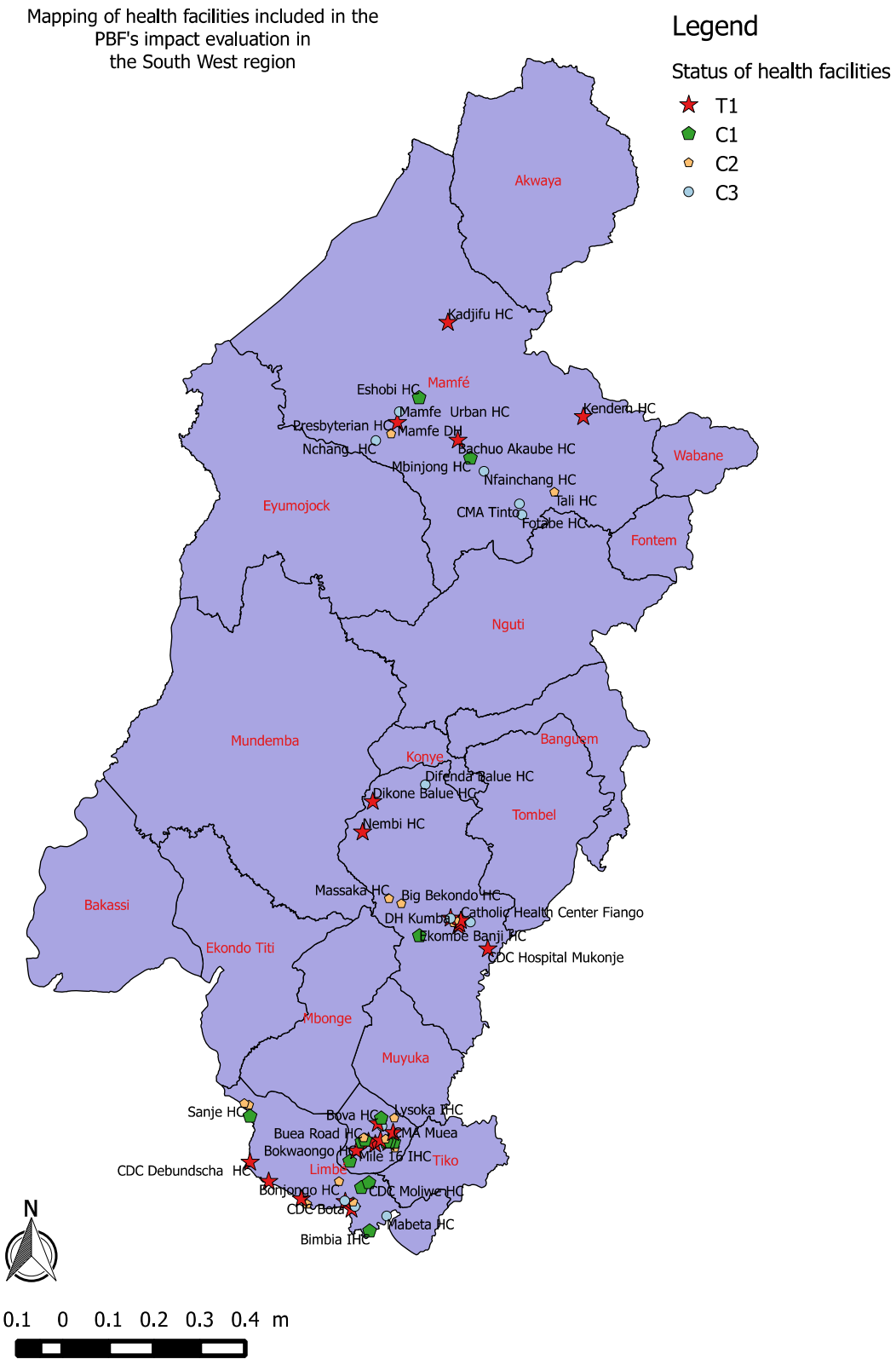
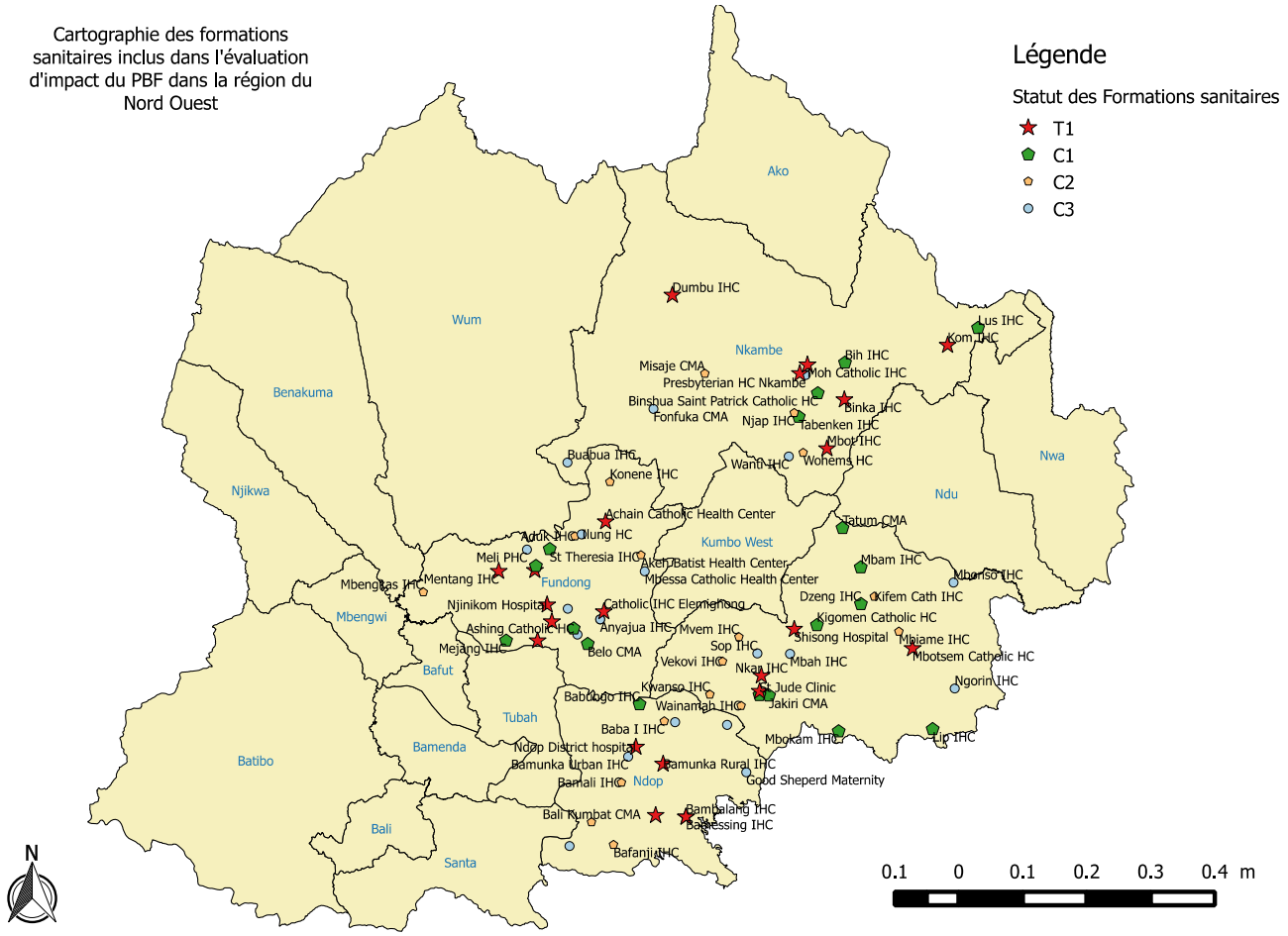

Mapping of health facilities included in the PBF's impact evaluation in the South West region
Legend
Status of health facilities
T1
C1
C2
C3
Akwaya
Kadjifu HC
Mamfé
Eshobi HC
Mamfe Urban HC
Kendem HC
Presbyterian HC
Mamfe DH
Nchang HC
Bachuo Akaube HC
Wabane
Mbinjong HC
Nfainchang HC
Eyumojock
Tali HC
CMA Tinto
Fotabe HC
Fontem
Nguti
Mundemba
Banguem
Konye
Difenda Balue HC
Dikone Balue HC
Tombel
Nembi HC
Bakassi
Massaka HC
Big Bekondo HC
DH Kumba
Catholic Health Center Fiango
Ekombe Banji HC
Ekondo Titi
CDC Hospital Mukonje
Mbonge
Muyuka
Sanje HC
Bova HC
Lysoka IHC
Buea Road HC
CMA Muea
Bokwaongo HC
Mile 16 IHC
Tiko
CDC Debundscha HC
Limbe
Bonjongo HC
CDC Moliwe HC
CDC Bota
Mabeta HC
Bimbia IHC
N
0.1 0 0.1 0.2 0.3 0.4 m

FIGURE A4


Cartographie des formations sanitaires inclus dans l'évaluation d'impact du PBF dans la région du Nord Ouest
Légende
Statut des Formations sanitaires
T1
C1
C2
C3
Ako
Wum
Nkambe
Benakuma
Njikwa
Ndu
Nwa
Kumbo West
Fundong
Mbengwi
Bafut
Tubah
Bamenda
Batibo
Ndop
Bali
Santa
Dumbu IHC
Lus IHC
Kom IHC
Bih IHC
Misaje CMA
Moh Catholic IHC
Presbyterian HC Nkambe
Binshua Saint Patrick Catholic HC
Binka IHC
Fonfuka CMA
Njap IHC
Tabenken IHC
Mbot IHC
Buabua IHC
Konene IHC
Wanu IHC
Wohems HC
Achain Catholic Health Center
Tatum CMA
Adulk IHC
Ulung HC
Meli PHC
St Theresia IHC
Akeh Batist Health Center
Mbam IHC
Mbessa Catholic Health Center
Mbengkas IHC
Mentang IHC
Mbonso IHC
Njinikom Hospital
Catholic IHC Elemenong
Dzeng IHC
Kifem Cath IHC
Ashing Catholic HC
Anyajua IHC
Mvem IHC
Kigomen Catholic HC
Shisong Hospital
Mbiame IHC
Mejang IHC
Belo CMA
Sop IHC
Mbotsem Catholic HC
Vekovi IHC
Nka IHC
Mbah IHC
Ngorin IHC
Kwanso IHC
St Jude Clinic
Babungo IHC
Wainamah IHC
Jakiri CMA
Baba I IHC
Mbokam IHC
Up IHC
Ndop District hospital
Bamunka Urban IHC
Bamunka Rural IHC
Bamali IHC
Good Sheperd Maternity
Bali Kumbat CMA
Bambalang IHC
Batessing IHC
Bafanji IHC
N
0.1 0 0.1 0.2 0.3 0.4 m

OUTPUT INDICATORS FOR THE MINIMUM PACKAGE OF HEALTH

TABLE A1: PBF SUBSIDY TABLE

N°	Curative Care	Definition	Support documents for data collection	Unit cost in FCFA
1	Out Patient Consultations (new cases): Nurse	Number of persons consulting the health center with a new episode of illness (consulted by nurses)	Outpatient consultation register or register used for curative care consultations	200
2	Out Patient Consultations (new cases): Doctor	Number of persons consulting the health center with a new episode of illness (consulted by Medical Doctors)	Outpatient consultation register or register used for curative care consultations	650
3	Out Patient Consultations of epidemics (new cases): Doctor or nurse (free)	Number of persons consulting the health center with a new epidemic case (consulted by Medical Doctors or nurses)	Outpatient consultation register or register used for curative care consultations or special epidemics registers	1000
4	Hospital bed days (observation/Hospitalization)	Total Number of days spent by all the inpatients in the health center (for observation or awaiting referral) period limited to a maximum of 48 hours	Inpatient (hospitalization register of the health facility	400
5	Hospital bed days (observation/Hospitalization) for epidemic cases (free)	Total Number of days spent by all the inpatients epidemic cases in the health center (for observation or awaiting referral) period limited to a maximum of 48 hours	Inpatient (hospitalization register of the health facility	1500
6	Minor surgery cases	Total number of New cases of minor surgery treated in the health facility (incision of abscesses, wound sutures, circumcisions etc.)	Minor surgery register	1500
7	Referral received in the hospital	Total number of referred patients who are received at the referral hospital	Referral register of the health center, referral forms at the level of the Hospital, consultation registers of the hospital, Hospitalization registers	1500
	Preventive Services/Care			
8	Children Completely Vaccinated	Children 0-11 months who received all of the following vaccines (BCG, Pentavalent 1, Pentavalent 2, Pentavalent 3 yellow fever and measles)	Vaccination register of the health facility	2500
9	VAT2 or VAT3 or VAT4 or VAT 5	Total number of women who received either VAT2 or VAT3 or VAT4 or VAT5	ANC Register and/or VAT vaccination register	1500
10	Home visits	Number of homes visited which had : appropriate collection and disposal of household refuge; a latrine in good	Home visits register signed by the Health committee representatives and the village chiefs or quarter heads	2500

		state ; appropriate use of mosquito bed nets and use of portable water.		
11	Vitamin A supplementation (distribution)	Number of children 6 to 59 months who received Vit A	Vit A supplementation register, Vaccination Register	20
12	HIV positive Pregnant Women put on ARV prophylactic treatment	Number of HIV positive Pregnant Women put on ARV prophylactic treatment according to the national PMTCT protocol in the month	PMTCT Register	7000
13	Newborn management of a baby born of an HIV positive mother.	Number of babies born of HIV positive mothers who are placed on PMTCT protocol in the month according to National directive	PMTCT Register	7000
14	Voluntary Counseling and Testing for HIV/AIDS	Number of people who came to the health facility for HIV/AIDS voluntary counseling and testing and who collected their results	VCT Register	1000
15	Cases of STIs treated	Number of new cases of STIs diagnosed and correctly treated in the month according to national protocols (Syndromic approach)	Outpatient consultation register	400
16	Cases of TB diagnosed positive by Microscopy	Number of new cases diagnosed positive by Microscopy in the health facility	TB and Lab registers	10,000
17	Cases of TB treated and healed	Total number of positive TB cases on treatment who were completely healed in the month	TB register, Lab register	20,000
	Reproductive Health			
18	Normal Assisted Delivery	Total number of normal deliveries carried out by qualified (or skilled) staff (nurses) in the facility in the month	Deliveries Register (Maternity Register)	2500
19	FP : New or old acceptants on oral pills of injections	Total number of both old and new acceptants of family planning who are currently on oral pills or injections	Family Planning Register	1200
20	FP : Implants and IUD	Number of new cases of Implants and/or IUD carried out in the month	Family planning register	3000
21	Post abortive Curettage (spontaneous or induced)	Total number of new cases of curettage (post abortive) carried out in the facility in the month	Maternity and theater register	3500
22	ANC1 or ANC2 or ANC3 or ANC4	Total number of pregnant women who consulted the health facility in the month either for ANC1 or ANC2 or ANC3 or ANC4	ANC Register	500
23	IPT1 or IPT2 or IPT3	Total number of pregnant women who consulted the ANC service of the facility in the month and who took either IPT1 or IPT2 or IPT3	ANC Register	500

TABLE A2: INTERVENTION GROUP COMPARISON TABLE

	T1 **Complete PBF with performance bonuses for medical personnel**	**C1** **PBF with subsidies that are not linked to performance**	**C2** **Only supervision without bonuses or autonomy**	**C3** **Status quo**
Contract	Classic PBF contract	Contract stipulating the conditions of PBF for verification and supervision	Contract stipulating technical support in the form of supervision	No contract
Business plan	Yes	Yes	Simple business plan focused on intensified supervision	No business plan
Quality evaluation	Quality evaluation and feedback with quality taken into account in bonus payment	Quality evaluation with feedback as in T1, but no effect on payment	Quality evaluation with feedback as in T1	Quality evaluation with written feedback twice a year
Review/verification of service amounts	Review and verification of service quantities	Review and verification of service quantities	Review and verification of service quantities	Single quarterly statement without verification of the quantity of services produced
Payment	Payments tied to performance	Payments not tied to performance	No payment	No payment
Management autonomy	Management autonomy with control over all revenues.	Management autonomy with control over all revenues.	No management autonomy, continuation the status quo system	No management autonomy, continuation the status quo system
Monthly activity report submitted to district	Yes	Yes	Yes	Yes

HOUSEHOLD SURVEYS

The impact evaluation adapted the Health Results Innovation Trust Fund (HRITF) survey instruments for this impact evaluation. Household surveys were conducted in each of the 14 districts included in the impact evaluation. To select the households to be surveyed, a catchment area was first established for each of the 245 primary care facilities. GIS mapping was conducted before the baseline survey to define realistic catchment areas for health facilities. GIS mapping was necessary because the government does not have a clearly defined health map with specific catchment areas. The government defines instead "Health Areas" (similar to sub-districts) that often include several facilities. As such it was necessary to define "zones of responsibility" for each facility. This GIS mapping defined 'true' catchment areas by taking into account physical features (like terrain or water bodies) and roads that influence travel time and thereby potentially affect health facility choice. One village from each health facility catchment area was randomly selected for the household survey. Regional maps of the study health facilities are presented in Appendix Figures A2 – A4.

A village household listing exercise was first conducted to identify all village households. At baseline 16 of the households identified in the listing exercise were randomly selected to be surveyed in each village. The survey team attempted to revisit all baseline households at endline. However, many baseline households could not be located or were no longer eligible at endline. When this occurred, baseline households were replaced using the nearest neighbor as recorded in the listing exercise. An additional four households were added to the household sampling roster at endline such that a total of 20 households were sampled in each village for the endline survey. In both rounds, the primary inclusion criteria for the household survey was that the household must have contained at least one woman who had been pregnant in the 24 months preceding the survey. Though the sample was meant to be a panel, with repeat sampling of the same households at baseline and endline, only a small proportion of households sampled at endline were also sampled at baseline. For instance, only 29% of the 4,813 households from which the 6,275 pregnant women surveyed at endline resided were also sampled at baseline. Therefore, the surveys are analyzed as repeated cross-sectional surveys rather than panel data in this report.

The household survey was administered to all members of the household who were present on the day of data collection. Demographic data including educational attainment and labor force participation was collected from all adult members of the household. Data on recent illness and health care use in the past four weeks were collected from all household members, with primary care givers providing information about child health. Household level data on housing characteristics, household assets and household level income were provided by the head of household. Additionally, the household survey contained separate modules for women of reproductive age (15 – 49 years), women who had been pregnant in the 24 months before the survey, and for children under five years of age. The main health themes covered in these modules included:

- Health behaviors for MCH services
- Health seeking behaviors, barriers to use and health service use
- Household health expenditures
- General perceptions of health service quality

In addition, the survey teams weighed and measured the height of all children aged under 5 years present in the household during the survey team's visit.

FACILITY-BASED SURVEY

The facility survey was conducted in all the CMAs, CSIs and District Hospitals in the 14 districts included in the impact evaluation. All facility team visits were unannounced. The facility-based survey included multiple components. The sample of health workers, patient-provider observations and client exit interviews was selected to enable findings from these three components to be linked.

FACILITY ASSESSMENT MODULE

The facility assessment module collected data on key aspects of facility functioning and structural aspects of quality of care. The individual in charge of the health facility at the time when the survey team visited the health facility was asked to be the respondent for this survey module. The main themes that were covered by the facility assessment included:

- Facility staffing, including staff on duty at the time of the survey team's visit and staff present at the time of the survey team's visit
- Facility infrastructure and equipment
- Availability of drugs, consumables and supplies at the health facility
- Supervision
- Record keeping and reporting to the Health Management Information System
- Facility management
- Official user charges at the facility
- Revenues obtained at the health facility, and how revenues have been used

HEALTH WORKER INTERVIEW MODULE

For health facilities with more than five health workers, a list of all clinical staff who worked in the area of maternal and child health providing prenatal or under five consultations was obtained. If this list contained more than five people, study enumerators interviewed a random sample of these health workers. If the list contained fewer than five people, all clinical personnel working in maternal and child health were interviewed. The interviews focused on the following areas:

- Role and responsibilities of the interviewed health worker
- Compensation, including delays in salary payments
- Staff satisfaction and motivation

OBSERVATIONS OF PATIENT-PROVIDER INTERACTION MODULE

While the health worker interview module collects information on what health workers know, the purpose of this module is to gather information on what health workers actually *do* with their patients.

A member of the survey team observed consultations with a systematic random sample of patients under five presenting with a new condition (i.e., not for follow-up visits or routine) and new ANC clients. The observer used a structured format to note whether key desired actions were carried out. In the case of patients under five, the instruments were focused on whether IMCI protocols are followed. For ANC clients, the instruments examined whether key desired actions (including counseling) were carried out. As primary care facilities do not offer ANC services on all days of the week – typically these are offered 2 days each week – the ANC module was not conducted at all health facilities. During the baseline survey, 5 under-5 and 5 ANC observations were conducted at each facility

where these modules are implemented. After finding that many health facilities did not offer ANC on the day of the survey at baseline, during the endline survey enumerators were asked to interview as many women receiving ANC on the day of the survey as possible to increase the sample size. All health workers selected for patient-provider observations will be included in the health worker interview sample.

PATIENT EXIT INTERVIEWS

Enumerators conducted an exit interview with all patients whose consultation was observed as part of the study procedures. If the patient was a child, the child's caregiver was interviewed. The under-fives included in the patient exit sample were the same children whose consultation with a provider was observed. In addition to this, exit interviews were conducted with all ANC clients whose consultation with a provider was observed.

ANALYSIS OF HOUSEHOLD CARE SEEKING BEHAVIOR

For the household survey, a random sample of 16 to 20 households was selected in each health facility catchment area. The analysis of the household survey in this report starts from the assumption that household members seek care in the health facility closest to where they live, or in other words that people living in the catchment area of a facility obtain health care in that facility.

However, it is apparent from the baseline survey data that households do not always seek care from the closest health facility in their health zone (appendix table 8). The household survey analysis assigns a treatment group (PBF, C1, C2 or C3), to each household in the study, which represents the treatment assignment of the closest health facility to the sampled household. Additionally, both the baseline and the endline household survey included information about the name of the health facility where the household sought reproductive health care. Using this information, we created a variable that measures whether women sought care in a health facility consistent with their assigned treatment group, a health facility assigned to another treatment group, a non-randomized hospital, or a health facility outside of the study area. At baseline, for antenatal care, for example, 44.8% of women sought care in a health facility assigned to their own treatment group, but 11.1% sought care in a higher level hospital (not included in the randomization conducted for the impact evaluation) and 22.2% sought care in other health facilities beyond their own health zone (18.7% in other facilities assigned to other treatment groups in our study sample and 3.2% in facilities not included in the study sample). Another 7.8% did not seek any antenatal care and we are missing information about the service location for the remaining 15.7%. If we focus on women for whom we have information about the service location, 52.3% sought care in in their "assigned" treatment group, and if we further exclude women who did not seek any antenatal care, this percentage increases to 57.6%. This "health care shopping" behavior whereby households bypass the closest health facility is also present for deliveries and postnatal care: at baseline, focusing on women for whom we have information about the service location and who sought care in a facility, only 51.9% delivered, and only 56.1% sought postnatal care in a facility with their corresponding treatment group.

At endline, those percentages are slightly higher, but not very substantially: focusing again on women for whom we have information about the service location and who sought care in a facility, 60.9% obtained antenatal care, 55% delivered, and 60.6% sought postnatal care in a facility assigned to their treatment group.

When the household is indeed seeking care in a health facility that is consistent with their assigned treatment group, this assignment is correctly done. However, when the household seeks care in another facility, this assignment between household and study group is potentially erroneous, leading to measurement error. This measurement error would introduce statistical noise in the analysis and reduce our capacity to measure potential impacts of the interventions (attenuation bias). Another interpretation of these patterns is to see it as non-compliance with assigned treatments. The ITT model estimated remains valid. The causal estimates the ITT model creates may however not fully capture the causal effect of the treatment relative to a "clean" control, but rather measure the causal impact of having a treated facility closer compared to people living further away from a treated facility. These estimates are likely below the true causal effect of the intervention. This is a substantial limitation of the household survey analysis that needs to be kept in mind.

The statistical analysis and interpretation of the household survey would be further complicated if this health care "shopping" behavior was driven or reinforced by the introduction of PBF or the interventions implemented in C1 (additional financing) and C2 (enhanced supervision). If this was the case, this could introduce a bias in the estimates going further than the attenuation bias described above. Appendix tables 9 to 11 investigate whether the

implementation of PBF or the other interventions have directly influenced household's health care seeking behaviors. For antenatal care, deliveries and postnatal care, they report results from multinomial logit difference-in-differences regression models where the four options for the household are: not seeking care, seeking care at the assigned facility, i.e. the closest one in the health zone, seeking care in an unassigned facility, i.e. another facility of the same level potentially randomized into a different impact evaluation group, and seeking care at a non-randomized facility, generally a higher level hospital which was not included in the randomization.

Overall, the results do not suggest that the health care seeking behavior is driven or even significantly influenced by the introduction of PBF or the other interventions in C1 and C2. We also conducted this analysis using OLS with a binary outcome equal to 1 if the respondent sought care in a health facility in their assigned treatment group, and zero otherwise. We found no evidence that health care shopping was affected by PBF in this analysis, consistent with the results from the multinomial logit (Appendix Tables A4 – A6). Health care shopping behavior by households was widespread in Cameroon at baseline in 2012 and continues to be widespread at endline in 2015, but does not appear to be a consequence of the introduction of PBF. We therefore decided to keep the presentation of the household survey results in which the analysis assumes that household members seek care in the health facility closest to where they live. We recognize that this assumption is not always verified and that therefore the results might suffer from attenuation bias. In addition to the fact that overall we did not find significant evidence that health care seeking behavior was influenced by the introduction of PBF, the following other considerations motivated our choice:

1) In many cases, the data collected about which health facility was visited allowed us to find out whether the visited facility was the closest one, but in case it was not, did not allow us to ascertain to which study group the visited facility pertained.

2) When we could ascertain to which study group the visited facility pertained, such bypassing behavior is clearly endogenous and assigning to the household bypassing its closest facility the study group of the facility actually visited would lead to endogeneity bias.

3) The results from the household survey analysis are broadly consistent with the results from the health facility survey analysis, which are not affected by the measurement error introduced by the health care shopping behavior of households.

4) The health care shopping behavior prevalent in Cameroon is likely present in many other countries. Our analysis uses the same assumptions and methods as the other impact evaluation reports including household survey results. However, to our knowledge, the household survey analysis in Cameroon is, so far, the only one to have explicitly collected or used detailed information about the name of the facility visited by the household sought care. We are therefore in a position to better acknowledge this study limitation and document how our household survey results are potentially affected by this health care bypassing behavior.

APPENDIX TABLE A3: HEALTH CARE SEEKING BEHAVIOR†

Baseline	Antenatal care		Delivery care		Postnatal care	
	N	%	N	%	N	%
Did not receive the health service	214	7.78	585	21.26	1,825	66.34
Received the health service in assigned treatment group	1,231	44.75	951	34.57	440	15.99
Received the health service in different treatment group	514	18.68	361	13.12	154	5.6
Received the health service in a facility outside the study area	88	3.2	72	2.62	44	1.6
Received the health service in a non-randomized hospital	304	11.05	304	11.05	138	5.02
Missing data on service location	398	15.73	332	19.03	141	15.38
N	2,751		2,751		2,751	
Endline	Antenatal care		Delivery care		Postnatal care	
	N	%	N	%	N	%
Did not receive the health service	226	6.82	582	17.57	1,954	59
Received the health service in assigned treatment group	1,626	49.09	1,358	41	705	21.29
Received the health service in different treatment group	435	13.13	395	11.93	177	5.34
Received the health service in a facility outside the study area	315	9.51	200	6.04	102	3.08
Received the health service in a non-randomized hospital	292	8.82	321	9.69	171	5.16
Missing data on service location	414	13.43	261	12.58	194	14.38
N	3,312		3,312		3,312	

†Percentages calculated from household survey data among sampled women who had been pregnant in the 24 months before the survey.

APPENDIX TABLE A4: HEALTH CARE SHOPPING FOR ANTENATAL CARE†				
	(1)	(2)	(3)	(4)
	No ANC	ANC in assigned treatment group facility	ANC in unassigned treatment group facility	ANC in non-randomized facility
Post indicator	0.001	0.040	-0.060*	-0.014
	[0.014]	[0.045]	[0.033]	[0.018]
PBF/Post interact	0.012	0.065	-0.021	-0.019
	[0.018]	[0.055]	[0.039]	[0.024]
Control 1/Post	-0.004	0.016	0.045	-0.015
interact	[0.018]	[0.054]	[0.038]	[0.023]
Control 2/Post	0.015	-0.017	0.005	0.016
interact	[0.020]	[0.057]	[0.041]	[0.023]
N	5407	5407	5407	5407

* = p < 0.10, ** p < 0.05, *** p< 0.01

† Results from multinomial logistic difference-in-differences regression models examining the effect of PBF on facility bypassing for reproductive health care. Regression model adjusted for individual (age, marital status, education level, religion, ethnicity, working status and type of work), household control variables (number of individuals in the household, housing type, house ownership, water source, and type of sanitation) and facility-level control variables at baseline (type of health facility, urban/rural). Standard errors were clustered at the health facility level.

APPENDIX TABLE A5: HEALTH CARE SHOPPING FOR SKILLED DELIVERY†				
	(1)	(2)	(3)	(4)
	No skilled delivery	Skilled delivery in assigned treatment group facility	Skilled delivery in unassigned treatment group facility	Skilled delivery in non-randomized facility
Post indicator	-0.036*	0.064	-0.060**	-0.001
	[0.020]	[0.040]	[0.030]	[0.017]
PBF/Post interact	0.018	0.038	0.042	-0.038
	[0.028]	[0.050]	[0.033]	[0.025]
Control 1/Post	-0.000	0.003	0.072**	-0.012
interact	[0.030]	[0.050]	[0.034]	[0.023]
Control 2/Post	0.016	0.004	0.039	-0.018
interact	[0.028]	[0.048]	[0.037]	[0.022]
N	5419	5419	5419	5419

* = p < 0.10, ** p < 0.05, *** p< 0.01

† Results from multinomial logistic difference-in-differences regression models examining the effect of PBF on facility bypassing for reproductive health care. Regression model adjusted for individual (age, marital status, education level, religion, ethnicity, working status and type of work), household control variables (number of individuals in the household, housing type, house ownership, water source, and type of sanitation) and facility-level control variables at baseline (type of health facility, urban/rural). Standard errors were clustered at the health facility level.

APPENDIX TABLE A6: HEALTH CARE SHOPPING FOR POSTNATAL CARE†				
	(1)	(2)	(3)	(4)
	No postnatal care	Postnatal care in assigned treatment group facility	Postnatal care in unassigned treatment group facility	Postnatal care in non-randomized facility
Post indicator	-0.059**	0.047*	-0.031**	0.005
	[0.025]	[0.027]	[0.016]	[0.010]
PBF/Post interact	-0.030	0.056*	0.030*	-0.018
	[0.032]	[0.033]	[0.018]	[0.014]
Control 1/Post	0.013	-0.015	0.041**	-0.006
interact	[0.030]	[0.031]	[0.019]	[0.012]
Control 2/Post	0.014	0.010	0.014	-0.006
interact	[0.031]	[0.032]	[0.019]	[0.013]
N	5634	5634	5634	5634

* = p < 0.10, ** p < 0.05, *** p< 0.01

† Results from multinomial logistic difference-in-differences regression models examining the effect of PBF on facility bypassing for reproductive health care. Regression model adjusted for individual (age, marital status, education level, religion, ethnicity, working status and type of work), household control variables (number of individuals in the household, housing type, house ownership, water source, and type of sanitation) and facility-level control variables at baseline (type of health facility, urban/rural). Standard errors were clustered at the health facility level.

APPENDIX TABLE A7: RESULTS OF HOUSEHOLD ANALYSIS WITH STRATIFICATION ON BASELINE FACILITY BYPASSING

	Skilled delivery			ANC			Postnatal care		
	Above median	Below median		Above median	Below median		Above median	Below median	
	β1	β2	p-value	β1	β2	p-value	β1	β2	p-value
Post indicator	0.080***	0.022		0.033	0.017		0.149***	0.036	
	[0.029]	[0.020]		[0.021]	[0.018]		[0.043]	[0.047]	
PBF/Post interact	-0.060	-0.027		0.030	-0.024		-0.034	-0.008	
	[0.042]	[0.034]	0.515	[0.030]	[0.024]	0.149	[0.052]	[0.077]	0.779
Control 1/Post	0.037	-0.004		-0.040	-0.004		-0.108**	0.045	
interact	[0.052]	[0.034]	0.495	[0.025]	[0.029]	0.334	[0.054]	[0.063]	0.058
Control 2/Post	-0.091**	-0.018		-0.051*	-0.040		-0.085	-0.056	
interact	[0.039]	[0.039]	0.172	[0.026]	[0.028]	0.762	[0.053]	[0.057]	0.702
p-value PBF vs.							0.102	0.485	
C1	0.069	0.565		0.016	0.489				
p-value PBF vs.							0.254	0.511	
C2	0.443	0.841		0.007	0.552				
p-value PBF vs.							0.516	0.914	
C3	0.149	0.431		0.315	0.326				
	2797	3128		3301	2872		3455	2447	

BALANCE AT BASELINE TABLES

APPENDIX TABLE A8: INDIVIDUAL LEVEL CHARACTERISTICS OF HOUSEHOLD MEMBERS SAMPLED AT BASELINE*

Individual all household members	*Mean T1*	*Mean C1*	*Mean C2*	*Mean C3*	*Mean total*	*p-value T1/C3*	*p-value C1/C3*	*p-value C2/C3*	*F-statistic*	*N*
Age	18.41	18.76	18.00	18.35	18.37	0.846	0.197	0.245	0.155	19232
Catholic	0.44	0.37	0.32	0.36	0.37	0.140	0.858	0.416	0.150	19196
Protestant	0.36	0.40	0.42	0.43	0.40	0.099	0.498	0.701	0.379	19196
Other religion	0.15	0.15	0.13	0.14	0.14	0.582	0.701	0.886	0.875	19196
Muslim	0.05	0.08	0.13	0.07	0.09	0.437	0.696	0.114	0.143	19196
Kom	0.08	0.08	0.05	0.14	0.09	0.307	0.317	0.111	0.460	19178
Banso	0.06	0.12	0.09	0.05	0.08	0.844	0.194	0.435	0.559	19178
Other ethnicity	0.86	0.80	0.86	0.81	0.83	0.444	0.951	0.439	0.726	19178
Adults > 18 years										
Years of school	5.65	5.70	5.57	5.52	5.61	0.481	0.353	0.779	0.785	6807
Literacy	0.74	0.75	0.72	0.73	0.73	0.793	0.580	0.857	0.882	7991
Any school	0.88	0.87	0.87	0.85	0.87	0.299	0.636	0.651	0.775	7984
Work	0.74	0.72	0.73	0.71	0.73	0.283	0.936	0.432	0.609	7812
Agricultural work	0.60	0.57	0.58	0.58	0.58	0.704	0.858	0.953	0.945	5698
Work in retail	0.14	0.16	0.16	0.15	0.15	0.593	0.903	0.899	0.891	5698
Other type of work	0.19	0.19	0.19	0.19	0.19	0.882	0.871	0.860	0.998	7737
Never married	0.17	0.18	0.18	0.18	0.18	0.505	0.959	0.838	0.895	8038
Monogamous marriage	0.45	0.46	0.47	0.45	0.46	0.905	0.853	0.663	0.939	8038
Polygamous marriage	0.07	0.08	0.08	0.07	0.07	0.681	0.783	0.665	0.865	8038
In union	0.21	0.17	0.17	0.19	0.18	0.584	0.616	0.692	0.708	8038
Divorced or widowed	0.11	0.11	0.10	0.10	0.10	0.839	0.520	0.647	0.754	8038

* Standard errors were adjusted for facility-level clustering of observations

APPENDIX TABLE A9: HOUSEHOLD LEVEL CHARACTERISTICS OF HOUSEHOLDS SAMPLED AT BASELINE*

Household	*Mean T1*	*Mean C1*	*Mean C2*	*Mean C3*	*Mean total*	*p-value T1/C3*	*p-value C1/C3*	*p-value C2/C3*	*F-statistic*	*N*
Total number of individuals in the household	5.61	5.47	5.57	5.64	5.57	0.895	0.328	0.696	0.769	3457
Number of women 15 - 49	1.39	1.33	1.35	1.39	1.37	0.894	0.145	0.309	0.357	3457
Number of kids under 5	1.52	1.48	1.58	1.55	1.53	0.491	0.156	0.654	0.288	3457
House with multiple flats	0.07	0.09	0.08	0.09	0.08	0.517	0.715	0.813	0.841	3457
Building with apartments	0.10	0.10	0.11	0.09	0.10	0.685	0.750	0.518	0.932	3454
Compound	0.26	0.22	0.27	0.26	0.25	0.967	0.381	0.882	0.710	3113
House	0.25	0.22	0.23	0.23	0.23	0.391	0.878	0.758	0.775	3457
Shack	0.03	0.04	0.02	0.02	0.03	0.526	0.069	0.967	0.281	3457
Other housing type	0.01	0.01	0.01	0.02	0.01	0.547	0.205	0.218	0.520	3455
Owner occupied dwelling - with mortgages	0.04	0.04	0.03	0.04	0.04	0.925	0.924	0.462	0.813	3457
Owner occupied dwelling - without mortgages	0.65	0.60	0.62	0.64	0.63	0.895	0.409	0.738	0.767	3457
Rented housing (not tied to the job)	0.14	0.16	0.15	0.15	0.15	0.750	0.807	0.980	0.957	3457
Housing rent free (other owner)	0.10	0.13	0.14	0.11	0.12	0.326	0.401	0.299	0.099	3457
Other housing payment type	0.05	0.06	0.05	0.03	0.05	0.550	0.342	0.570	0.783	3043
Piped water into the dwelling	0.03	0.02	0.03	0.03	0.03	0.978	0.482	0.966	0.689	3457
Piped water into yard/plot	0.09	0.07	0.07	0.06	0.07	0.328	0.788	0.885	0.754	3457
Piped water from public tap/standpipe	0.33	0.37	0.34	0.28	0.33	0.348	0.131	0.277	0.452	3457
Water from a well or borehole	0.06	0.09	0.14	0.07	0.09	0.718	0.567	0.055	0.146	3457
Water from a protected well	0.05	0.03	0.06	0.05	0.05	0.876	0.403	0.413	0.438	3457

Household	*Mean T1*	*Mean C1*	*Mean C2*	*Mean C3*	*Mean total*	*p-value T1/C3*	*p-value C1/C3*	*p-value C2/C3*	*F-statistic*	*N*
Water from an unprotected well	0.06	0.05	0.06	0.05	0.05	0.546	0.829	0.443	0.865	3457
Water from a protected spring	0.08	0.09	0.07	0.09	0.08	0.742	0.966	0.565	0.915	3457
Water from an unprotected spring	0.22	0.19	0.14	0.25	0.20	0.487	0.167	0.010	0.050	3457
Surface water puddles lakes rivers	0.09	0.08	0.07	0.12	0.09	0.429	0.333	0.165	0.580	3457
Latrine pit with a slab	0.30	0.33	0.31	0.26	0.30	0.410	0.118	0.307	0.456	3455
Latrine pit without a slab	0.59	0.58	0.56	0.61	0.59	0.748	0.602	0.427	0.878	3455
Other sanitation type	0.11	0.09	0.13	0.13	0.11	0.594	0.206	0.995	0.364	3455

* Standard errors were adjusted for facility-level clustering of observations

APPENDIX TABLE A10: FACILITY LEVEL CHARACTERISTICS AT BASELINE*

Facility	*Mean T1*	*Mean C1*	*Mean C2*	*Mean C3*	*Mean total*	*p-value T1/C3*	*p-value C1/C3*	*p-value C2/C3*	*F-statistic*	*N*
Number of beds in the health facility	8.07	9.98	11.70	9.26	9.84	0.453	0.788	0.226	0.244	185
Electricity in the health facility	0.70	0.78	0.69	0.77	0.73	0.410	0.914	0.363	0.626	206
Piped water in the health facility	0.40	0.38	0.35	0.35	0.37	0.665	0.792	0.927	0.947	206
Facility has an incinerator	0.08	0.22	0.24	0.23	0.19	0.033	0.914	0.932	0.027	206
Latrine in the health facility	0.85	0.84	0.85	0.79	0.83	0.457	0.540	0.409	0.853	206
Facility open 24 hours	0.66	0.72	0.64	0.71	0.68	0.607	0.899	0.439	0.775	206
Water towel and soap in Examination Room	0.46	0.43	0.47	0.45	0.45	0.897	0.858	0.805	0.976	199
Secure Box for Sharps	0.80	0.86	0.80	0.83	0.82	0.708	0.715	0.668	0.832	200
User Fees for Consultation Posted	0.38	0.32	0.36	0.35	0.35	0.810	0.723	0.921	0.939	206
User Fees for Laboratory Services Posted	0.34	0.35	0.37	0.23	0.32	0.257	0.225	0.154	0.460	195
Child Weighing Scale	0.87	0.88	0.94	0.83	0.88	0.595	0.460	0.069	0.222	202
Height Measure	0.41	0.43	0.45	0.53	0.46	0.253	0.303	0.430	0.663	191
Tape Measure	0.96	0.98	1.00	0.96	0.98	0.903	0.539	0.153	0.167	204
Blood Pressure Instrument	0.86	0.90	0.87	0.85	0.87	0.836	0.468	0.777	0.895	199
Thermometer	0.98	0.94	0.95	1.00	0.97	0.317	0.079	0.079	0.067	204
Stethoscope	0.96	0.92	0.91	0.91	0.93	0.353	0.951	0.918	0.630	202
Lab services	0.74	0.80	0.82	0.77	0.78	0.686	0.727	0.557	0.760	206
Blood test	0.34	0.42	0.48	0.54	0.45	0.084	0.314	0.574	0.347	159
Malaria test	0.97	1.00	0.91	0.97	0.96	0.970	0.317	0.223	0.101	160
TB test	0.13	0.28	0.20	0.19	0.20	0.501	0.375	0.864	0.461	159
HIV test	0.11	0.23	0.18	0.22	0.18	0.194	0.880	0.703	0.402	158
Facility provided immunization	0.98	0.96	0.95	0.98	0.97	0.944	0.582	0.366	0.734	206
Facility provides ANC	0.98	0.98	0.98	1.00	0.99	0.318	0.317	0.318	0.392	206

* Standard errors were adjusted for facility-level clustering of observations

APPENDIX TABLE A11: BASELINE HEALTH SERVICE COVERAGE

Health service coverage	*Mean T1*	*Mean C1*	*Mean C2*	*Mean C3*	*Mean total*	*p-value T1/C3*	*p-value C1/C3*	*p-value C2/C3*	*F-statistic*	*N*
Skilled delivery	0.77	0.75	0.78	0.76	0.77	0.864	0.920	0.790	0.981	2878
At least two ANC visits	0.86	0.90	0.91	0.87	0.88	0.604	0.216	0.140	0.164	2969
Tetanus vaccination during ANC	0.86	0.88	0.86	0.86	0.87	0.881	0.428	0.830	0.862	2971
Postnatal care	0.34	0.31	0.34	0.31	0.33	0.323	0.847	0.347	0.674	2966
Use of modern contraception	0.36	0.30	0.30	0.35	0.33	0.837	0.234	0.137	0.180	2029
Full vaccination (documented)	0.52	0.61	0.54	0.58	0.56	0.079	0.840	0.795	0.234	796
Full vaccination documented/self-report	0.53	0.66	0.59	0.62	0.60	0.355	0.593	0.495	0.497	1201
Growth monitoring	0.04	0.04	0.05	0.04	0.04	0.085	0.542	0.530	0.137	3541
Bed net use (< 5 yrs)	0.74	0.80	0.81	0.80	0.79	0.883	0.828	0.771	0.972	5786

* Standard errors were adjusted for facility-level clustering of observations

FINANCING FIGURES

FIGURE A5: TOTAL PAYMENT PROVIDED TO T1 AND C1 HEALTH FACILITIES IN NORTH WEST

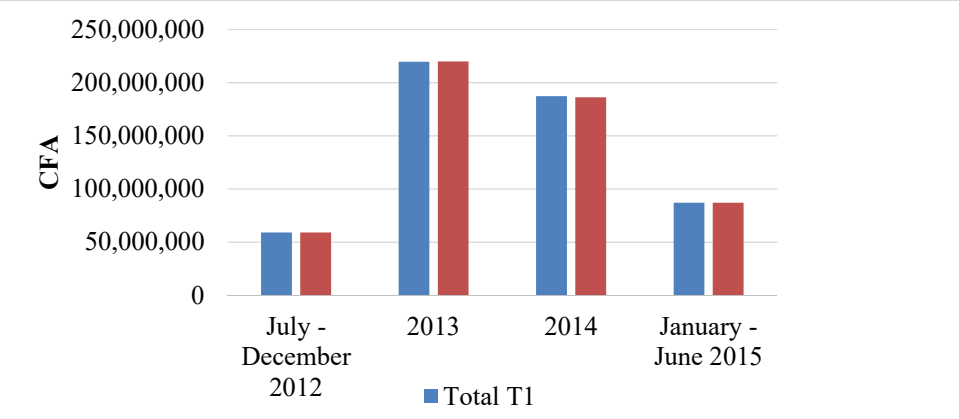


FIGURE A6: TOTAL PAYMENTS PROVIDED TO T1 AND C1 HEALTH FACILITIES IN SOUTH-WEST

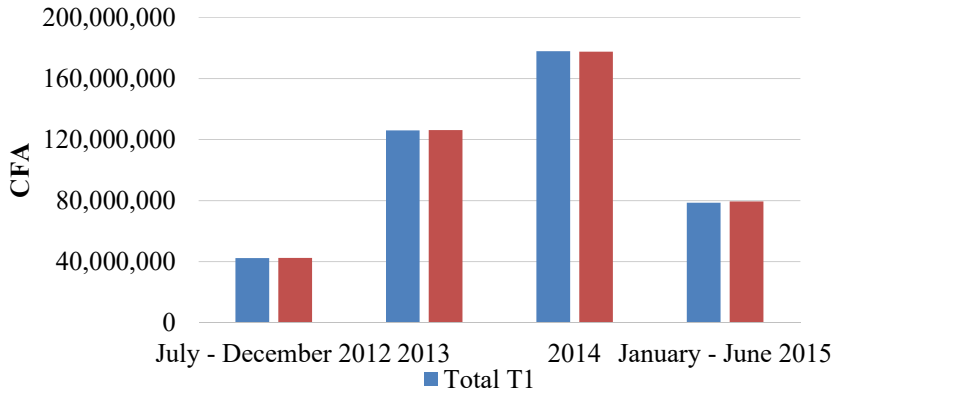
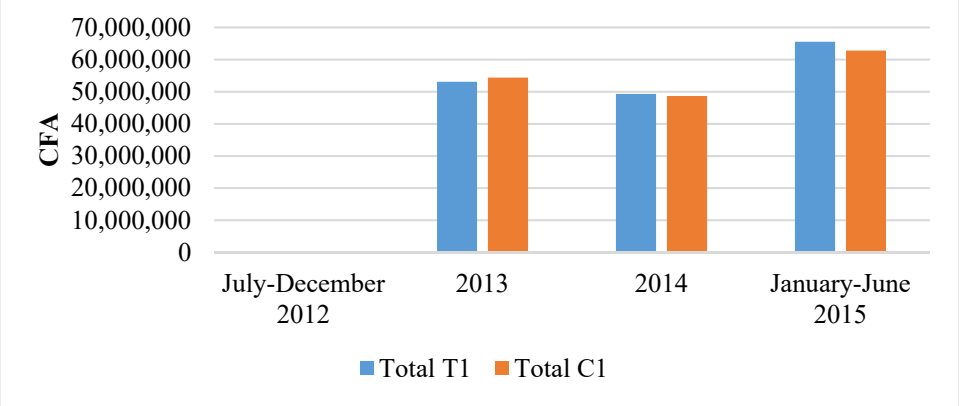
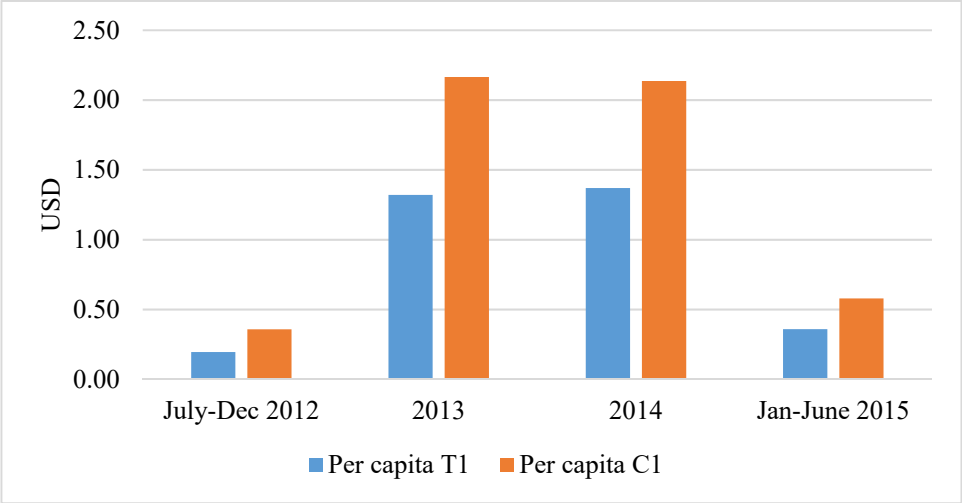


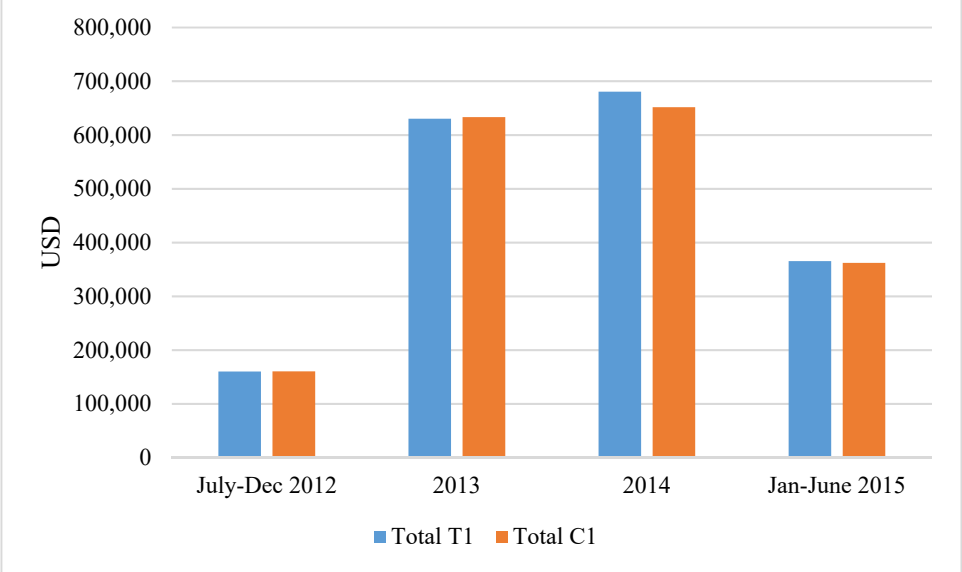
FIGURE A7: TOTAL PAYMENTS PROVIDED TO T1 AND C1 HEALTH FACILITIES IN EAST



APPENDIX FIGURE A8: PER CAPITA SUBSIDY PAYMENTS, T1 AND C1 FACILITIES, $US



APPENDIX FIGURE A9: TOTAL SUBSIDY PAYMENTS, T1 AND C1 FACILITIES, $US



PRINCIPAL RESULTS COMBINING GROUPS T1 & C1

Table A12: Coverage of reproductive health services† and provision of modern family planning‡					
	(1)	(2)	(3)	(4)	(5)
	Skilled delivery	At least two ANC visits	Tetanus vaccine during pregnancy	Postnatal care	Modern contraception
Post indicator	0.052***	0.022	0.001	0.105***	0.002
	[0.019]	[0.014]	[0.019]	[0.031]	[0.044]
PBF & T1/Post interact	-0.013	-0.006	0.014	-0.024	-0.045
	[0.025]	[0.017]	[0.021]	[0.036]	[0.049]
Control 2/Post interact	-0.050*	-0.044**	0.010	-0.070*	0.000
	[0.029]	[0.019]	[0.023]	[0.039]	[0.053]
p-value PBF/T1 vs. C2	0.174	0.031	0.841	0.154	0.274
p-value PBF/T1 vs. C3	0.600	0.711	0.503	0.504	0.352
Baseline mean C3	0.784	0.894	0.878	0.323	0.180
N	5858	5974	5975	5966	4498

* = p < 0.10, ** p < 0.05, *** p< 0.01

† Results from difference-in-differences regression models examining the effect of PBF on reproductive health service use among female respondents included in the household survey who had been pregnant in the previous 24 months. ‡ Results from difference-in-differences regression models examining the effect of PBF on modern contraceptive use among female respondents of reproductive age (15 – 49) included in the household survey. Regression models adjusted for individual (age, marital status, education level, religion, ethnicity, working status and type of work) and household control variables (number of individuals in the household, housing type, house ownership, water source, and type of sanitation). Standard errors were clustered at the health facility level.

Table A13: Full vaccination coverage, growth monitoring, bednet use, stunting, underweight and wasting among children†							
	(1) Fully vaccinated documented by vaccine card	(2) Fully vaccinated by vaccine card or self-report	(3) Growth monitoring in the last month	(4) Slept under a bednet	(5) Stunting	(6) Under-weight	(7) Wasted
Post indicator	0.126*	0.107**	-0.014	-0.181***	-0.008	-0.010	0.047**
	[0.072]	[0.052]	[0.013]	[0.025]	[0.025]	[0.022]	[0.021]
PBF & T1/Post interact	0.056	0.076	0.014	-0.002	0.009	0.044*	-0.016
	[0.085]	[0.061]	[0.015]	[0.034]	[0.030]	[0.026]	[0.025]
Control 2/Post interact	0.019	0.029	0.022	0.003	0.037	0.018	-0.028
	[0.092]	[0.073]	[0.019]	[0.036]	[0.033]	[0.028]	[0.027]
p-value PBF/T1 vs. C2	0.611	0.445	0.655	0.894	0.315	0.255	0.561
p-value PBF/T1 vs. C3	0.508	0.214	0.348	0.959	0.772	0.087	0.524
Baseline mean C3	0.599	0.645	0.048	0.809	0.444	0.147	0.067
N	1569	2448	7055	10107	8711	8672	8480

* = p < 0.10, ** p < 0.05, *** p< 0.01

† Results from difference-in-differences regression models examining the effect of PBF on child vaccination among children (12 – 23 months), growth monitoring in the last month among children (12 – 59 months), having slept under a bednet the night before the survey and on child anthropometric outcomes (stunting, underweight and wasting) among children under 5 years of age included in the household survey. Regression model adjusted for individual (age, father in the household, religion, ethnicity) and household control variables (number of individuals in the household, housing type, house ownership, water source, and type of sanitation). Standard errors were clustered at the health facility level

Table A14: Provision of reproductive and child health services†						
Panel A	(1)	(2)	(3)	(4)	(5)	(6)
	Skilled delivery	ANC	Tetanus vaccine during pregnancy	Postnatal care	Modern contraception	Third dose of polio vaccine
Post indicator	0.516	3.209	-16.895***	-3.799*	0.670	-5.284***
	[0.797]	[3.934]	[5.563]	[2.175]	[1.006]	[1.984]
PBF & T1/Post interact	1.602*	2.303	18.858***	4.888**	7.581***	3.708*
	[0.918]	[5.581]	[5.945]	[2.219]	[1.728]	[2.134]
Control 2/Post interact	0.041	4.014	8.783	3.498	3.369	1.106
	[1.357]	[5.438]	[7.694]	[2.498]	[2.061]	[3.952]
p-value PBF/T1 vs. C2	0.177	0.733	0.079	0.277	0.063	0.459
p-value PBF/T1 vs. C3	0.083	0.680	0.002	0.029	<0.001	0.084
Baseline mean C3	7.76	20.57	32.84	10.22	3.02	23.90
N	2182	2220	2220	2220	2220	2220
Panel B	(7)	(8)	(9)	(10)	(11)	
	Meningitis vaccine	Measles vaccine	HIV testing	PMTCT	ART	
Post indicator	-45.963***	-3.741*	4.214	-3.551	1.023*	
	[9.769]	[2.250]	[3.068]	[3.323]	[0.610]	
PBF & T1/Post interact	20.432*	2.860	56.470***	2.223	-1.077	
	[11.522]	[2.441]	[11.762]	[3.431]	[0.679]	
Control 2/Post interact	8.430	-0.714	6.730	1.644	-0.692	
	[13.555]	[3.546]	[5.773]	[3.161]	[0.596]	
p-value PBF/T1 vs. C2	0.249	0.221	<0.001	0.703	0.322	
p-value PBF/T1 vs. C3	0.078	0.243	<0.001	0.518	0.114	
Baseline mean C3	46.65	20.90	9.98	9.86	0.012	
N	2220	2220	2220	2220	2220	

* = p < 0.10, ** p < 0.05, *** p< 0.01

† Results from difference-in-differences regression models examining the effect of PBF on reproductive health service provision reported in facility registers. Monthly number of services provided during the six months before the baseline and endline surveys used as the dependent variable. Regression models adjusted for facility controls (type of health facility public/private/religious, urban/rural). Standard errors were clustered at the health facility level.

Table A15: Health care spending as reported in household data†				
	(1) Official provider fee	(2) Unofficial provider fee	(3) Lab and x-ray fees	(4) Transportation fees
Post indicator	1801.95	2054.42*	1040.46	123.38
	[1473.53]	[1057.44]	[711.56]	[200.81]
PBF & T1/Post interact	-913.24	-2494.01*	-1002.42	-475.05**
	[1458.26]	[1363.51]	[792.18]	[231.22]
Control 2/Post interact	-1369.69	-1424.04	-631.38	-369.07
	[3966.29]	[1244.36]	[884.46]	[236.28]
p-value PBF/T1 vs. C2	0.903	0.307	0.505	0.580
p-value PBF/T1 vs. C3	0.532	0.069	0.207	0.041
Baseline mean C3	1689.22	2183.33	1603.09	910.30
N	2374	2261	2292	2365

* = p < 0.10, ** p < 0.05, *** p< 0.01

† Results from difference-in-differences regression models examining the effect of PBF on health care spending in the last 4 weeks among respondents in the household survey. Regression model adjusted for individual (age, sex) and household control variables (number of individuals in the household, housing type, house ownership, water source, and type of sanitation). Standard errors were clustered at the health facility level

WHO WELL-BEING INDEX

(8)	WHO well-being index	
	Now I will read five statements about how a person might be feeling. For each of the five statements, please indicate whether in the last two weeks, you have been feeling this way most of the time, more than half of the time, less than half of the time, only rarely, or never. PLEASE SHOW AND ASK TO PICK OUT THE COLORED AND NUMBERED CARDS RESPONSE CODE MOST OF THE TIME 1 MORE THAN HALF OF THE TIME 2 LESS THAN HALF OF THE TIME 3 ONLY RARELY 4 NEVER 5	RECORD RESPONSE CODE
(8.01)	In the past two weeks, I have felt cheerful and in good spirits…..	
(8.02)	In the past 2 weeks, I have felt calm and relaxed…	