

Impact Evaluation of Nigeria State Health Investment Project

Acknowledgements

This report was produced and written by a task team consisting of Eeshani Kandpal (PI, Economist, DECPI), Benjamin Loevinsohn (TTL, Lead Public Health Specialist, GHN07), Christel Vermeersch (co-PI, Senior Economist, GHN04), Elina Pradhan (co-PI, Young Professional, GHN07), Madhulika Khanna (Research Associate, Consultant, DECPI), Mark Conlon (Research Assistant, Consultant, DECPI) and Wu Zeng (Consultant).

Project team members, Ayodeji Oluwole Odutolu (TTL, GHN07), Gyorgy Fritsche (GHN07) and Opeyemi Fadeyibi (Consultant, GHN07) provided comments on multiple drafts as well as support and inputs during the design and implementation of the survey and preliminary analysis.

This work has been supported by the Federal Ministry of Health of Nigeria. The team particularly acknowledges the support provided by Emmanuel Meribole.

The data collection was implemented by the Nigerian Bureau of Statistics and the National Population Commission. Kevin McGee (Economist, DECDG) and Lena Nguyen (Survey Specialist, DECDG) provided technical support and supervision in the development of instruments and deployment of Survey Solutions using Android tablets and oversaw data management. Asmelash Tsegay (Consultant, DECDG) and Wondu Kassa (Consultant, DECG) supervised the training of the field teams and the field work.

Facility data were collected by the National Bureau of Statistics. Samuel Adebayo and Florence Oke of the National Bureau of Statistics managed the collection of facility data. Household data were collected by the National Population Commission. Inuwa Jalingo, Raphael Ologun and Amarachukwu Onwuzurumba of the National Population Commission of Nigeria managed the collection of household data.

The financial contributions of the Health Results Innovation Trust Fund (HRITF) are gratefully acknowledged.

This version: December 2018

ABSTRACT

Background: In response to slow progress in improving health service delivery, the Government of Nigeria requested World Bank support in testing performance-based financing (PBF) and decentralized facility financing (DFF) as part of the Nigeria State Health Investment Project (NSHIP). PBF provides funding directly to health facilities based on the quantity and quality of services they deliver. Funds are transferred electronically to each facility's bank account and they have substantial autonomy in how they use the funds. Up to half the funds can be used to pay performance bonuses to staff. Supervision was substantially strengthened. DFF was similar to PBF except that facility earnings were NOT based on the quantity and quality of services they delivered, they could not pay performance bonuses to health workers and the amount they received was, by design, half the amount earned by PBF facilities. This paper reports an impact evaluation (IE) of the PBF-DFF pilot.

Methods: A three-armed trial with experimental and quasi-experimental components was used to assess the effectiveness of PBF, DFF, and a control arm ("business as usual") to improve the quantity of key maternal and child health (MCH) services delivered and the quality of care (QOC) provided in public facilities. All the local government areas (LGAs) in three states were randomly allocated to either PBF or DFF, yielding a truly randomized comparison of PBF-DFF. Control states and LGAs were chosen by matching population demographic characteristics to those in the treatment states, generating a "difference-in-difference" comparison of the NSHIP arms to the control. Baseline health facility and household surveys were conducted in 2014 and at endline in 2017. The study also incorporated a cost-effectiveness analysis (CEA) and an independent review of financial management.

Findings: Both PBF and DFF had a practically and statistically significant impact on the quantity of key MCH services. Of the 8 quantity indicators identified in the IE concept note, 7 showed positive adjusted differences in difference (DiDs, i.e., endline – baseline in the experimental arm minus endline-baseline in the control arm) and 3 were statistically significant. For example, NSHIP increased fully immunized child (FIC) coverage by 14 percentage points (pp) and modern contraceptive prevalence by 4.5 pp. PBF and DFF also had a sizeable and consistent effect on QOC. Of the 26 indicators of quality identified in the IE concept note before data was collected, 20 (77%) showed statistically significant estimates of program impact. For example, availability of essential drugs, contraceptive supplies, basic equipment, hand-washing stations, proper waste management all improved much more in the NSHIP arms than the control arm. Process measures of QOC did not improve consistently. There was little difference between the PBF and DFF arms in terms of QOC and only modest differences in the coverage of key services. Specifically, PBF led to an 11 pp increase in skilled birth attendance; but DFF outperformed PBF on Penta3 coverage and insecticide-treated net usage by children younger-than-5. Other key outcomes (curative care for children younger-than-5, modern contraceptive prevalence, and antenatal care utilization) were not significantly different between PBF and DFF. The largest gains under NSHIP occurred in the upper income quintiles in our sample, which is more rural and poorer than the national average. The financial review generally found that funds sent to the NSHIP facilities were appropriately used for their intended purpose and that financial management, while needing to be improved, was better in NSHIP facilities than in the control group. The CEA showed that both PBF and DFF were cost-effective interventions in comparison to Nigeria's per capita GDP.

Interpretation: Both PBF and DFF had important effects on the coverage and quality of MCH services while the control arm, like the rest of Nigeria, made only modest progress. The improvements were accomplished at a cost that is affordable, particularly if the Basic Health Care Provision Fund (BHCPF) is implemented and funded as envisaged in the National Health Act. Both interventions are cost-effective and were likely successful due to decentralization of funds, autonomy given to the facilities, improved supervision, and investments in health systems management. While the progress achieved by PBF and DFF is encouraging, further improvements are required on coverage, process quality of care, and improving access to services for the poor. These challenges suggest that further innovation is required particularly on strengthening management at the facility level to take full advantage of the decentralized resources and efforts on the demand-side targeted at the poor. The similar results achieved by PBF and DFF suggests that the incentive structure of PBF could be improved and that DFF may be sufficient if there are robust efforts to improve supervision, management, and monitoring and evaluation. The relatively low absolute levels of utilization of preventive and promotive services suggests that demand-side interventions may further strengthen the impact of NSHIP.

EXECUTIVE SUMMARY

1. **Rationale:** In response to slow progress in the delivery of primary health care (PHC) services, the Federal Government of Nigeria (FGON) requested World Bank Group (WBG) support in testing performance-based financing (PBF). PBF provides funding directly to health facilities based on the quantity and quality of services they deliver. This report presents the results of an impact evaluation (IE) of this PBF pilot that incorporates experimental and quasi-experimental comparisons. There are several reasons why this study is important, including: (i) it was implemented on a large scale (covering more than 9 million people), for a reasonable length of time (an average of three years of implementation), and at an incremental cost (about \$2.60 per capita per year) that is likely affordable using domestic resources; (ii) the study examines approaches that have been incorporated into the design of the Basic Health Care Provision Fund (BHCPF) stipulated under the national health act (NHAct); and (iii) it compares PBF to Decentralized Facility Financing (DFF) and a control group (“business as usual”). The DFF arm provides the same levels of autonomy, enhanced supervision, and decentralization as the PBF arm and thus tests the impact of one form of incentive-based payments to health workers.

A. Background

2. **Progress on Health Indicators:** Nigeria’s progress on health outcomes has been mixed. The under-five mortality rate (U5MR) in Nigeria remains high, especially by comparison to other lower-middle income countries (LMICs) but has come down about 37% over the last 13 years. A poor child in Nigeria still faces the highest risk of dying before its fifth birthday in all West Africa. The maternal mortality ratio remains static at 576 deaths per 100,000 live births (NDHS 2013) and nutritional status (stunting among children under-5) has not improved over the last decade. The total fertility rate (TFR), at 5.8 in 2016 (MICS), has changed little since 1990. The lack of clear progress on health impact indicators is consistent with slow progress on PHC service delivery. In 1990 a demographic and health survey (DHS) found immunization coverage (DPT3) to be 33%. It was also 33% in the 2016 multiple indicator cluster survey (MICS). Similarly, slow progress has been noted for antenatal care (ANC), skilled birth attendance (SBA), and the contraceptive prevalence rate (CPR). There has, however, been some notable progress in certain aspects of PHC. A large improvement has been achieved in malaria prevention (e.g. use of long-lasting insecticide treated bed nets [LLINs] by children under-5) and treatment. Transmission of wild polio virus has been interrupted and there has not been a wild polio case in more than two years. Nigeria was able to contain Ebola Virus Disease quickly with a case fatality of about 30% (substantially lower than in other affected countries).

3. **Impediments to Progress:** A variety of constraints have prevented improvements in Nigeria’s health outcomes: (i) public investment in health is lower as a proportion of GDP than in any other country in the world; (ii) the public health system suffers from limited accountability, weak management, and inadequate supervision; (iii) PHC facilities lack operating budgets and have limited autonomy; and (iv) health workers are poorly motivated and suffer from low productivity. Unlike many other countries in sub-Saharan Africa (SSA), Nigeria does NOT suffer from a shortage of health workers or physical access to health facilities. 85% of the population lives within an hour’s walk of a health facility and the health worker to population ratio is twice the SSA average.

4. **Response to PHC Challenges:** To address these challenges in PHC, the FGON, with support from the WBG and the Health Results Innovation Trust Fund (HRITF) launched in 2014, a PBF project called the Nigeria State Health Investment Project (NSHIP), initially implemented in three states but now expanded to five other states, covering all the North East states. The FGON also established the Basic Health Care Provision Fund (BHCPF) as envisaged in the National Health Act of 2014. The BHCPF promises significant additional domestic resources for PHC and its design employs results-based approaches similar to NSHIP's.

B. Project Design

5. **Objectives:** The project development objectives (PDOs) of NSHIP were to “to increase the delivery and use of high impact maternal and child health interventions and to improve the quality of care at selected health facilities in the participating states”. The project appraisal document (PAD) for NSHIP identifies five indicators that were to be tracked to measure progress towards the PDO: (i) proportion and number of 12-23 months-old children fully immunized; (ii) proportion of births attended by skilled health providers; (iii) average health facility score on quality of care; (iv) number of curative care visits by children under five; and (v) number of Direct Project Beneficiaries who are women. The concept note (i.e. protocol) for the IE listed a series of 26 specific measures of quality of care that included both structural measures, like availability of essential drugs and contraceptives, and process measures, such as health worker knowledge and the extent to which national protocols were followed. In addition, the IE concept note identified a further five quantity indicators as primary outcomes: (i) modern contraceptive prevalence rate; (ii) antenatal care; (iii) institutional delivery; (iv) Penta3 immunization of children 12-23 months; and (v) use of long-lasting insecticide treated nets (LLINs) by children under-5.

6. **Interventions:** NSHIP supported two different approaches to improving PHC service delivery, PBF and DFF, that allowed a comparison to “business as usual.”

(i) **Performance-Based Financing (PBF):** Most of the PHC facilities in local government areas (LGAs) assigned to PBF received a quarterly payment based on the quantity of pre-defined services they provided. Each type of service had a tariff associated with it and the facility received a payment that reflected the number of services provided multiplied by the tariff. (For example, if a PHC facility fully immunized 100 children in the quarter and the tariff was \$5 per child immunized, the facility would receive \$500.) The quantity of services was reported monthly and verified quarterly by an external verification agency. To ensure quality of care was addressed, a quantitative supervisory checklist (QSC) that assessed structural and process quality of care, was used by LGA supervisors and formed the basis of a quality bonus. The QSC was also verified by the external verification agency. An additional bonus was tied to the remoteness of the facility. The amount earned by the facility was transferred electronically to the facility's bank account for which the signatories were the officer-in-charge (OIC) and the chair of the Ward Development Committee (WDC). Facilities could use these funds for: (i) health facility operational costs (about 50%), including maintenance and repair, drugs and consumables, outreach, and other quality enhancement measures; and (ii) performance bonus for the health workers (up to 50%).

(ii) **Decentralized Facility Financing (DFF):** DFF was similar to PBF except that the payments to the PHC facilities were NOT linked to the quantity or quality of services they delivered. By design the DFF

facilities received half of the amount the PBF facilities earned since they were not allowed to pay performance bonuses to their staff. DFF facilities were also NOT subjected to third party verification of quantity or quality. However, the DFF facilities had the same level of autonomy in using their funds as PBF, they were supervised in a similar way, and they also received funds into their bank accounts through electronic transfer.

7. **Geographic Scope:** NSHIP was implemented in three states, Adamawa (North East), Nasarawa (North Central) and Ondo (South West). These states were purposively selected on the basis of their performance on health outcomes along six Project Development Objectives but also their willingness to accept the intervention (World Bank, 2011).

C. Study Methodology

8. **Three-Armed Trial:** The IE randomly allocated all the 52 LGAs in the experimental states (Adamawa, Nasarawa, and Ondo) to either the PBF or DFF arms thus constituting a randomized trial comparing the two approaches. A control group (“business as usual”) was established by selecting states similar to the experimental states along thirteen observable demographic characteristics that are associated with maternal and neonatal health. The three control states were neighbors of the experimental states in the same geopolitical zone and comprised Taraba (North East), Benue (North Central), and Ogun (South West). The comparison between NSHIP and the control states was, thus, a “difference-in-difference” assessment. Before and after household surveys and health facility surveys were undertaken in all LGAs of the three experimental states and the three control states. Baseline data were collected between February and April of 2014. The endline data were collected between August and October of 2017. Thus, the study assesses about three years of implementation.

9. **Research Questions:** The IE was originally intended to answer the following questions: (i) whether NSHIP’s interventions (PBF and DFF) improved the availability, utilization, and coverage of maternal, child, and reproductive health services, particularly among the poor; and (ii) whether they improved the quality of care. Subsequently, the scope of the IE was expanded to address: (i) the cost-effectiveness of the three arms; and (ii) whether NSHIP financial resources were used appropriately.

10. **Data Collection:** Data was collected using the following tools:

- (i) **Health Facility Surveys:** In each study LGA, one health facility was randomly chosen per ward, resulting in baseline and endline health facility surveys being administered at 786 health facilities across the six project and comparison states. These surveys included: (a) an assessment of the health facility itself (availability of drugs, equipment, etc.); (b) an interview with a health care provider at the facility; (c) direct observation of patient-health care provider interactions in the context of antenatal care (ANC) and under-5 curative care visits; and (d) exit interviews with ANC and Under-5 patients. The baseline sample includes 2,250 health care provider interviews, as well as observation of and exit interviews with 1,959 antenatal care patients and 1,778 under-5 five curative care

patients. The endline survey includes exit interviews and clinical observations on 2,640 antenatal care patients and 2,519 patients seeking under-5 curative care.

- (ii) **Household Surveys:** The National Population Commission of Nigeria defined enumeration areas across the country for the 2006 census. In 2008, the Federal Ministry of Health (FMOH) used these enumeration areas to create facility catchment areas. All households in the catchment areas of the sample facilities were listed at baseline in 2014 and again at endline in 2017 for this study. From the full household listing, we sampled all households with at least one pregnant woman or a birth in the past 24 months on a probability proportional to size (PPS) basis. This selection process resulted in 7,683 households being interviewed at baseline and 7,527 households being interviewed at the endline. The household questionnaire collected data on socio-economic and demographic conditions along with detailed health histories, anthropometry, and the woman's pregnancy and delivery.
- (iii) **Cost-Effectiveness Analysis:** The cost-effectiveness analysis (CEA) was conducted from a health system perspective and examined financial costs rather than economic costs. The CEA focused on the incremental (additional) costs) and compared PBF, DFF, and NSHIP (i.e., both PBF and DFF) to the control group. It also compared PBF to DFF. The additional costs included in the analysis were: (i) NSHIP implementation costs in the Nigeria, and (ii) the World Bank headquarters' cost for designing, implementing and monitoring the project. All costs were measured in US 2015 dollars, and a discount rate of three percent was applied.
- (iv) **Financial Review:** An independent financial review of NSHIP was undertaken by the auditing firm Ernst and Young (E&Y). Their staff reviewed the financial records of the state primary health care development agencies (SPHCDA) in all 3 intervention states and the 3 control states. They also visited 3 LGA primary health care departments in each of the intervention states (9 in total) and 10 PHC facilities in each state (30 in total). In each control states 1 LGA was visited (3 in total) and 5 PHC facilities (15 in total) were reviewed.

11. Estimating Impact Using Difference-in-Difference (DiD):

To estimate the impact of NSHIP, PBF, and DFF, this study uses a difference-in-difference approach, sometimes called a "double-difference" method. This approach simply compares the changes from baseline to endline in the experimental group to the same change in the control group. For example, the coverage of fully immunized children (FIC) went from 31.0% in the NSHIP area at baseline to 46.8% at endline for a difference of 15.8 percentage points. In the control states FIC coverage went from 24.6% at baseline to 26.3% at endline for a difference of 2.1 percentage points. The DiD is calculated as 15.8 percentage points – 2.1 percentage points = 13.7 percentage points. The comparison of PBF and DFF LGAs exploits the random assignment with equal probability of LGAs to these two arms. However, for the comparison to control states, the use of DiD is based on the assumption that the rates of changes in the key indicators are parallel, i.e. that they are likely to change at the same rate if the intervention had not been carried out. This hypothesis was tested using results from the 2008 and 2013 DHS and found not to hold for eight of twenty two key household characteristics, including skilled birth attendance. Thus, the DiD estimates have to be adjusted to take account of that fact. This adjustment was done by matching treatment and control units on baseline characteristics to estimate the probability that a unit falls into either a project or a control state. Then, this probability is used to reweight the DiD regression equation, placing a greater weight on the observations that are in project states but should have been in control states. However, it is worth noting that in only one out of twenty-two comparisons are control states

worse off than experimental states, suggesting that, if anything, unadjusted DiDs underestimate the true impact of NSHIP. A comparison group that is consistently becoming worse while the treatment group improves might raise concerns that we are overestimating the impact of the program by comparing it to an artificially low and deteriorating counterfactual.

D. Main Results

12. **Quantity of Services:** Table 1 describes the changes in the quantity indicators highlighted in the PAD and the IE concept note. It's important to note that, on average, the NSHIP and Control arms were reasonably similar at baseline. NSHIP's quantity-related PDO indicators all showed positive adjusted difference-in-differences (DiD's) and two of the three were statistically significant at $p < 0.05$. The improvements in these indicators ("effect sizes") were sizeable. For example, the number of consultations for children under-5 almost tripled (increased 2.7-fold) in the NSHIP arm compared to 31% increase in the control arm. Of the 5 additional quantity indicators mentioned in the IE concept note, 4 has positive adjusted DiD's and one was statistically significant. With the exception of ITN use by children under-5, the control group made little progress. In the Nigerian context, where there has been limited progress in population coverage of these services over the last 25 years, the improvements seen in the NSHIP arms are encouraging. They also compare favorably to the mean annual level of improvement observed for the same indicators in other developing countries (see Table 2 below). Nonetheless, the endline coverage rates remain mediocre compared to Nigeria's neighbors. For example, the endline coverage of Penta3 immunization at 56.4% is still lower than the greater than 80% coverage achieved in Cameroon, Ghana, and Senegal (latest DHSs).

Table 1: Baseline, Improvements, and Difference-in-Differences for Quantity Indicators

Indicator	Baseline Coverage		Change from Baseline % points		DiD	Adjusted DiD
	NSHIP	Control	NSHIP	Control		
Skilled Birth Attendance (%)	57.6	64.2	14.4	2.1	12.3	4.4
Fully Immunized Child (%)	31.0	24.6	15.8	2.1	13.7	14.1***
Consultations by under-5's in the last month per public health facility (number)	29.6	25.9	50.4§	8.1§	42.3	42.0***
Contraceptive Prevalence Rate, Modern (%)	17.0	18.9	8.8	2.2	6.6	4.5*
Penta3 Immunization (%)	40.4	35.9	16.0	-0.3	16.3	16.3***
Institutional Delivery (%)	50.9	55.8	13.0	4.2	8.8	2.0
Antenatal care 4 visits (%)	52.0	46.0	3.9	4.1	-0.2	-3.6
ITN use by children under-5 (%)	44.6	45.6	17.7	18.9	-1.2	1.0

Note: * $p < 0.1$ ** $p < 0.05$ *** $p < 0.01$. § represents numbers not percentage points

Table 2: Improvements seen in NSHIP Arms (Percentage Points) per Year Compared to Global Mean

	Penta3	SBA	mCPR	ANC	ITN Use <5
Baseline % for NSHIP	40.37	57.6	16.98	52.1	44.56
Change in % points-baseline to endline	15.96	14.39	8.77	3.9	17.70
NSHIP Avg. changer per year in % points	5.3	4.8	2.9	1.3	5.9
Global Mean change per year % points¹	5.0	1.5	1.7	1.2	5.3
Median change per year, Global ¹	3.0	1.2	0.9	0.8	5.5
Years for Global comparison	1980-2008	1985-2008	1987-2008	1987-2008	1999-2008

¹Source: Arur et al, 2011 Setting Targets in Health, Nutrition, and Population Projects

13. **Comparison of PBF to DFF on Quantity Indicators:** Overall, there was little difference between the PBF and DFF arms in terms of quantity of services delivered. Of the 8 quantity indicators included in the IE, DFF achieved larger adjusted DiDs on 4, however, PBF achieved statistically greater improvements in skilled birth attendance and the related institutional delivery rate. PBF may have also done better on modern contraceptive prevalence rate but DFF likely achieved better results on immunization and ITN use.

14. **Power to detect differences:** The IE was designed to detect a treatment effect of 10 pp with 97% certainty for facility-level outcomes and with 90% certainty for household-level outcomes. For example, we detect a 7 pp increase in the coverage of institutional deliveries in PBF versus control arms with a 95% confidence interval ranging from -0.4 pp to +14 pp and a 10 pp increase in PBF versus DFF arms with 95% confidence interval ranging from 3 pp to 17 pp. On the other hand, we are unable to detect (at usual levels of confidence) some differences due to smaller-than-expected sample sizes, including a 4.4 pp increase in skilled birth attendance in NSHIP versus control facilities (95% confidence interval -2 pp to +11 pp) or a 6.6 pp increase in Penta3 coverage in DFF versus PBF (95% confidence interval [-2 to 2.3pp]). Results are robust to Kling and Liebman (2004) average effects correction for multiple hypothesis testing and suggest that while NSHIP improved structural quality, the DFF arm drove most of these gains. In terms of outcomes, PBF improved maternal health outcomes, but DFF increased child health outcomes.

Table 3: Baseline, Improvements, and Difference-in-Differences for Quantity Indicators, PBF vs. DFF

Indicator	Baseline Coverage		Change from Baseline % points		DiD
	PBF	DFF	PBF	DFF	
Skilled Birth Attendance (%)	53.26	62.25	19.78	8.55	11.23***
Fully Immunized Child (%)	37.68	24.56	12.91	18.13	-5.21
Consultations by under-5's in the last month per public health facility (number)	33.24	26.43	53.51	48.37	5.14
Contraceptive Prevalence Rate, Modern (%)	17.69	16.20	9.81	7.64	2.18
Penta3 Immunization (%)	47.88	33.05	11.91	19.41	-7.5*
Institutional Delivery (%)	47.71	54.38	17.96	7.68	10.30***
Antenatal care 4 visits (%)	51.91	52.14	2.84	5.13	-2.5
ITN use by children under-5 (%)	46.59	42.45	15.72	19.7	-5.6*

Note: * p<0.1 **p<0.05 ***p<0.01

15. **Equity:** As can be seen in Table 4, the largest improvements seen under NSHIP in SBA and FIC (the PDO indicators that were to be broken down by income quintile) were observed in the 3rd, 4th, and 5th income quintiles. For example, there is nearly an 11-percentage point double difference in SBA for PBF in the fourth income quintile. On the other hand, for immunization even the lower income quintiles saw some improvement. Improvements in coverage were found to be greater in urban than rural areas although not by much for immunization. However, the IE sample is more rural and poorer than the national average. Indeed, as Table 5 shows, the sample for this IE is generally poorer than the Nigerian DHS sample which is nationally representative. Indeed, comparing wealth distributions of the IE and DHS surveys, the observed gains do not come from the 4th and 5th quintiles of the wealth distribution of the whole country, but rather from the middle of the national wealth distribution.

Table 4: Adjusted DiD's in Percentage Points in Comparison to the Control Group by Income Quintile

Indicator	Group	Q1 (poorest)	Q2	Q3	Q4	Q5 (richest)
Skilled Birth Attendance (SBA)	NSHIP	-2.6	-2.0	6.8	2.4	3.2
	PBF	-2.5	4.5	11.8**	11.0**	6.1
	DFF	-1.2	-9.2	0.7	-6.5	0.2
Fully Immunized Child (FIC)	NSHIP	6.6	6.2	3.6	18.3***	22.8***
	PBF	3.5	6.4	0.3	13.8*	18.1**
	DFF	10.5	7.4	8.0	20.4***	24.9***

Note: * p<0.1 **p<0.05 ***p<0.01

Table 5: Comparison of Household Wealth in NSHIP IE and Nigerian DHS

	DHS		NSHIP IE Survey	
	2008	2013	2014	2017
Radio	76.9	68.3	64.39	61.23
Television	39.3	47.6	44.37	43.4
Mobile Phone	49.2	75.1	60.39	77.7
Refrigerator	16.6	18.3	12.33	11.98
Canoe	2.8	2.3	5.7	2.16
Bicycle	28.7	18.3	10.18	7.18
Animal-drawn cart	4	3.6	1.26	1.15
Motorcycle/scooter	29.8	31.2	39.09	36.72
Car/truck	9.5	8.7	6.24	6.34
Motor boat	0.5	0.8	0.9	1.2
Ownership of agricultural land	68.3	57.8	36.15	38.38
Ownership of farm animals	60.4	49.6	33.88	62.93

16. **Quality of Care:** Overall, the quality of care (QOC) indicators described in the IE concept note increased much more in the NSHIP arms than in the control arm. Of the 26 QOC indicators in the IE concept note 21 (81%) adjusted DiDs favored NSHIP and 20 (77%) were statistically significant (p<0.05). As can be

seen in Table 6, significant improvements were seen in structural QOC such as availability of drugs, equipment, proper handwashing stations, and healthcare waste management. NSHIP facilities also carried out much more outreach. On process quality of care the results were more mixed. The proportion of health workers following national protocols for under-5 examinations declined slightly (but not as much as in the control arm) and antenatal care (ANC) protocol completion improved only a little. In addition, health worker knowledge did not improve under NSHIP. The results on process QOC indicators demonstrates that there is still a lot of work to do in this area. (A complete listing of QOC indicators is in the main body of the text.) The DFF arm drove most of the gains on QOC.

Table 6: Selected Quality of Care Indicators – NSHIP vs. Control with Adjusted DiDs

Indicator	Baseline (%)		Improvements from Baseline		Adjusted DiD
	NSHIP	Control	NSHIP	Control	
> one female clinical staff present on day of survey	84.0	86.3	8.0	-8.4	19.6***
% of health facilities (HFs) with water for hand washing, soap, clean towel in patient area	68.0	73.0	26.0	-4.0	43.0***
% of HFs with basic delivery equipment	17.1	15.4	57.0	-2.0	59.8***
Number of essential drugs available on day of survey	6.5	7.5	8.1	0.7	7.8***
Average number of contraceptive methods in stock on the day of survey	1.4	1.8	1.4	0.1	1.5***
% of HFs that have a working waste disposal system (bin, pit or incinerator) in use & safety box for sharps	59.0	66.0	36.0	4.0	34.0***
Average health worker clinical knowledge score	45.9	46.1	-5.4	-8.6	3.9
Under-five examination quality score (based on IMCI protocols)	55.2	63.9	-7.4	-21.0	14.6***
ANC examination quality score (based on national ANC protocols)	45.1	44.8	4.4	-4.5	8.2*
% of HFs that conduct outreach for key MCH services	49.0	35.3	22.2	-5.8	22.2***

Note: * p<0.1 **p<0.05 ***p<0.01

17. **Challenges to Inference:** Inferences drawn from differences between NSHIP and the control group are somewhat limited by the fact that project states were purposively chosen. This is not an issue for the PBF-DFF comparison. External validity is reduced by the fact that households were selected from the catchment areas of our study facilities not from the LGA as a whole. The rest of the LGAs may have different levels of household level coverage. Finally, while the PBF versus DFF relies on randomized assignment of LGAs to the two arms, the comparisons to the control are based on purposively selected states and are quasi-experimental in design.

18. **Mechanisms of Action:** The study also examined the causal and behavioral mechanisms through which the PBF and DFF arms could have achieved the observed gains. The influx of operating funds to facilities was associated with increases in availability of inputs and conduct of outreach. PBF workers who were aware of NSHIP incentives saw more patients than did DFF workers who were also aware of NSHIP. This result suggests that awareness of the incentive payment may have succeeded in increasing the number of patients seen. However, overall levels of awareness of NSHIP were low, suggesting that the full impact of PBF was not realized.

19. **Costs:** Of the total funds spent on NSHIP up until June 2017, 63.5% was disbursed to PBF and DFF health facilities (of which a small proportion was provided to LGA PHC department and hospital management boards that oversee these health facilities). The remaining 36.5% was spent on health systems governance and management at the LGA, State and Federal levels. Overall, annual per capita expenditure on NSHIP was \$2.62. PBF cost \$3.49 per capita per year and DFF cost \$1.74. Of the funds provided to PBF facilities, 76% were accounted by six services: family planning (19%), delivery care (17%), curative care (12%), HIV/STD care (10%), vaccination (9%), and household visits (9%).

20. **Cost-Effectiveness Analysis:** The incremental cost-effectiveness ratios (ICERs) of PBF as compared to DFF and control were \$698 and \$796/QALY gained, respectively, without quality of care adjustment. These ratios fell to \$458 and \$300/QALY gained if the quality of care is taken into account. The ICER for NSHIP was estimated to be \$901 and \$293/QALY gained, without and with quality adjustment, respectively. Both PBF and DFF are cost-effective when compared with Nigeria's GDP per capita¹. PBF is found to be relatively more cost-effective than DFF primarily due to the improvements in skilled birth attendance and the use of modern contraceptives. However, this difference disappears once the estimates are adjusted for impact of quality.

21. **Financial Review:** The main findings of the E&Y financial review included: (i) the SPHCDA arranged for the transfer of the correct amount of funds to each facility and the average payment was accomplished in 51 days (compared to the 45 days standard in the project implementation manual); (ii) there was no evidence of "phantom" health facilities receiving funds or non-NSHIP facilities receiving any transfers; (iii) NSHIP funds accounted for about 95% of all funds in PBF and DFF facilities and were generally being used appropriately to meet operational expenses; (iv) financial management in NSHIP facilities needed to be improved as some expenses were not recorded properly, vendors were sometimes paid in cash, and in some facilities the system of signatories was not being followed; (v) financial management in control facilities was almost non-existent even though they had some cash income from user-fees. It appears that decentralizing funds to facilities is likely to result in less corruption than maintaining the same funds at state or federal levels.

E. Discussion and Policy Implications

22. The results of this IE suggest the following:

- (i) **NSHIP should be expanded:** The study demonstrates that both PBF and DFF had important effects on the coverage and structural quality of MCH services while the control arm, like the rest of Nigeria, made only modest progress. Under real-world conditions and at large scale PBF and DFF appear to be practical and scalable interventions in the Nigerian context.
- (ii) **NSHIP is affordable and cost-effective:** The improvements seen under NSHIP were accomplished at a cost that is affordable using domestic resources, particularly if the BHCPF is implemented and funded as envisaged in the National Health Act. Both PBF and DFF are cost-effective compared to Nigeria's per capita GDP.

¹ WHO recommends comparing ICER to GDP/Capita. GDP/Capita proxies for the productivity a person in a year. If an intervention could save more than what a person produces in a year, it is regarded as highly cost-effective. Overall, PBF program is very cost-effective, whether it is compared to DFF or the control group.

- (iii) **Further improvements are needed:** While the NSHIP results are encouraging, there are 3 important challenges: (a) the endline coverage of MCH services remains mediocre by comparison to Nigeria's neighbors; (b) the process measures of quality of care need to improve significantly to have real health impact; and (c) there is a clear need to improve services for the poor and those living in remote rural areas (often the same people).
- (iv) **Further innovations are required:** Given these challenges, there is need for further innovation in the following areas: (a) health facility management needs strengthening so that health facilities can take full advantage of the resources and autonomy provided under PBF and DFF; (b) taking greater advantage of the private sector, particularly in rural areas; (c) exploring other forms of monetary incentives for health workers which might achieve better results under PBF; and (d) demand-side efforts aimed at increasing utilization by the poor.
- (v) **Decentralized Funds, Autonomy and Strengthened Feedback were likely key:** The similar results achieved by PBF and DFF suggest that providing operating budgets to health facilities, allowing them to spend the funds on their perceived priorities, systematic feedback using a QSC, and strengthened management and governance at LGA, state and federal levels may have been key reasons for the success of NSHIP. However, we do not find the quality of internal or external supervision at the facility-level to have driven observed gains in either project arm. Further, the fact that most health workers in NSHIP facilities did not know about the program, including most in PBF facilities who received financial incentives from the program, suggests room for strengthening facility-level management and supervision.
- (vi) **DFF and PBF may be used in combination:** Of the eight outcomes considered, four are significantly different between PBF and DFF, two in each direction, suggesting that both arms were successful to some extent. Results suggest that while NSHIP improved structural and process quality, the DFF arm drove most of these gains. In contrast, while PBF improved maternal health outcomes, DFF was more effective at improving child health outcomes. However, it is important to note that DFF may have benefited from the concurrent implementation of PBF. It is of course possible that if supervision is strengthened, management improved, and monitoring and evaluation robustly implemented, the DFF model can achieve similar results at lower cost. PBF, on the other hand, is more effective at improving institutional delivery. Finally, although a majority of health workers were not aware of NSHIP, the impact of PBF was strongest among those who were aware of the project. The impact of NSHIP may thus be strengthened by increasing awareness of the program and improving other aspects of facility-level management. DFF may be a cost-effective option to providing most services, while PBF may be used for targeted indicators including institutional delivery. However, the absolute levels of coverage of skilled birth attendance, antenatal care, and modern contraceptive use remain low, suggesting that demand-side barriers may constrain utilization. Incentives—financial and otherwise— geared at overcoming social norms and increasing demand may be explored as a complement to supply-side incentives.
- (vii) **Extrapolating from these results:** This IE examines impact after three years of program exposure, while most comparable studies look at impacts of PBF pilots after 18-24 months. It's possible that long-term effects on health system strengthening have not yet been realized. However, it's also possible that the results achieved by this large-scale pilot may not be accomplished if PBF and DFF were implemented nationwide. In addition, the low utilization levels and the fact that the gains are not among the poorest of the poor beg the question whether these impacts represent an upper bound to the effectiveness of NSHIP.

Introduction

Nigeria's progress on health outcomes has been mixed. The under-five mortality rate (U5MR) in Nigeria remains high, especially by comparison to other lower-middle income countries (LMICs) and a poor child in Nigeria still faces the highest risk of dying before its fifth birthday in all West Africa. The maternal mortality ratio remains stagnant at 576 deaths per 100,000 live births (NDHS 2013) and nutritional status (stunting among children under-5) has not improved over the last decade. The total fertility rate (TFR), at 5.8 in 2016 (MICS), has changed little since 1990. The lack of clear progress on health impact indicators despite 85 percent of the population living within an hour's walk of a PHC is consistent with slow progress on service delivery.

Several factors constrain Nigeria's health outcomes, including public investment in health that is lower as a proportion of GDP than in any other country in the world, and poor accountability, management and supervision of the public health system suffers from limited accountability. Centralized resources and management also mean that PHC facilities lack operating budgets and have limited autonomy. This lack of resources and autonomy further translates into low health worker motivation (World Bank, 2012).

To address these challenges, in 2014, the Government of Nigeria in collaboration with the HRITF and the World Bank launched a performance-based financing (PBF) project called the Nigeria State Health Investment Project (NSHIP) in three states chosen for their poor performance on maternal and neonatal health (MNH) outcomes: Adamawa, Nasarawa and Ondo. The four key elements of NSHIP draw from the PBF structure to (a) provide additional finances to the frontlines, (b) augment facility level supervision, (c) decentralize financial responsibilities by providing autonomy to the facilities; and (d) incentivize health workers to increase the quantity and quality of services.

This report assesses the impact of NSHIP on the quantity and the quality of MNH care provision.

Evaluating NSHIP is important because (i) it is one of the largest PBF pilots of its kind, covering more than 9 million people, (ii) the evaluation period was longer than most comparable IEs (an average of three years of implementation as opposed to the typical 18-24 month window of HRITF-funded PBF IEs), and at an incremental cost (about \$2.60 per capita per year) that is likely affordable using domestic resources. Further, this IE compares PBF to Decentralized Facility Financing (DFF) and a control group ("business as usual"). The DFF arm provides the same levels of autonomy, enhanced supervision, and decentralization as the PBF arm and thus tests the impact of a particular type of incentive-based payments to health workers, making it one of few PBF IEs to do so (notable exceptions being de Walque et al., 2017 and Shen et al., 2017).

These results are also important for in-country decision-making as this IE examines approaches that have been incorporated into the design of the Basic Health Care Provision Fund (BHCPF) stipulated under the national health act (NHAct), which is being used to overhaul Nigerian health infrastructure in a bid to achieve universal health coverage. The NHAct, passed in 2014, provides a legal framework for the re-organization and management of the health systems in the country. NSHIP is a part of the health reforms that are underway since the passage of the National Health Act and has already been expanded to the remaining five states in the North East.

2 The Nigeria State Health Investment Project

2.1 Design

NSHIP shifted the emphasis from purchasing inputs to paying for the outputs generated by the existing health care system. Primary health centers received additional funds based on the quantity and quality of services they delivered. States and local government areas (LGAs) received funds to strengthen institutional capacity. Lessons from prior World Bank's investments in the country were incorporated in the design. Facility-level autonomy, better supervision, and fiscal decentralization were key components. The primary objective was "to increase the delivery and use of high impact maternal and child health interventions and to improve the quality of care at selected health facilities in the participating states" (World Bank, 2011). The following three criteria guided the selection of three participating states under NSHIP:

- (1) strong governance capability and commitment,
- (2) greater health needs,
- (3) willingness to use PBF approaches, and
- (4) geo-political representation and filling gaps in donor presence.

Adamawa State from the North East Zone, Nasarawa State from the North Central Zone and Ondo State from the South West Zone were selected for their poor performance on MNH outcomes (World Bank, 2012). NSHIP was thus designed to cover about 400,000 pregnant women and 1.8 million children in these three states.

Two intervention packages were introduced under NSHIP. Half the LGAs in the three states received a quarterly payment contingent upon the delivery of pre-defined set of health services under PBF. These indicators were reported monthly by the health facilities and verified quarterly by an external verification agency. In addition, a quarterly audit assessed structural and process quality at each facility using a

comprehensive checklist. Bonuses were tied to quality indicators as measured through a comprehensive quality check list, with further bonuses for facility remoteness. Table 1 describes an example of the how payments to a given health facility were calculated under PBF. In this example, if a health facility fully immunizes 50 children in a quarter, they could earn US\$100 (100 x US\$2 per child fully vaccinated). Under NSHIP, 20 specific services were incentivized in primary health facilities. In this example the facility would have earned \$1,600. The total would have to be adjusted by a quality score based on a quantitative checklist administered at the facility every quarter. This facility would have earned 50 percent times 25 percent of its quantity payment, i.e. \$200. Money would be transferred electronically to the facility's bank account. Facilities could use these funds for: (i) health facility operational costs (about 50%), including maintenance and repair, drugs and consumables, outreach and other quality enhancement measures; and (ii) performance bonus for health workers (up to 50%). To incentivize improvements in quality of care at the secondary level, including referral from primary health facilities, the project tested a similar PBF approach in secondary hospitals. However, given the small number of secondary facilities and the focus on primary health care, this IE focuses on results at the primary health facility level.

Table 1: Example of PBF in a Health Facility

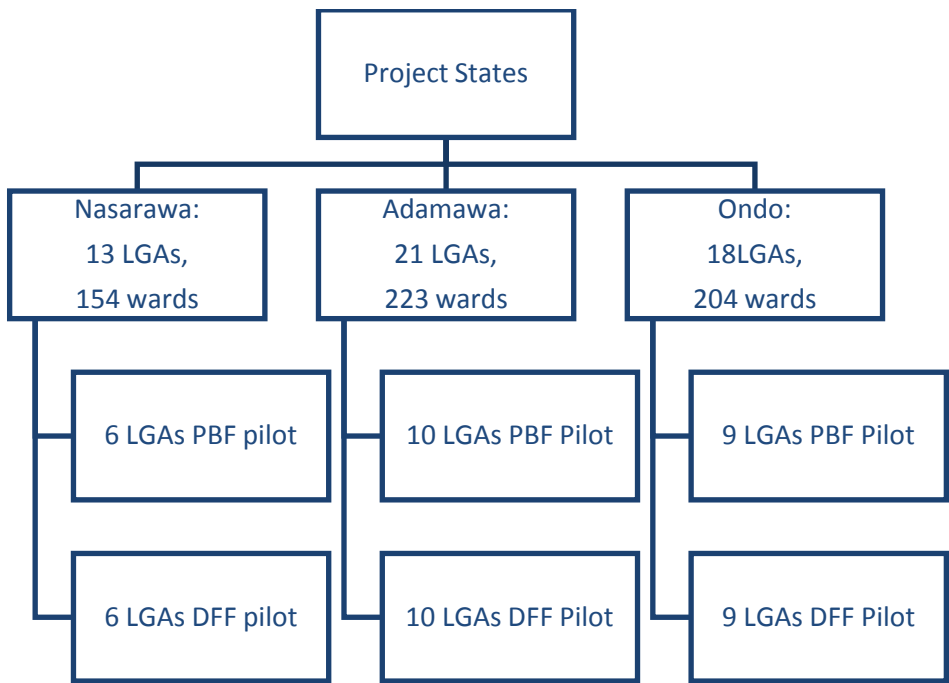
<u>Service</u>	Number Provided Last Quarter	Unit Price	Total Earned
Child fully vaccinated	50	US\$2	US\$100
Skilled birth attendance	60	US\$10	US\$600
Curative care patient visit	1,800	US\$0.5	US\$900
Sub-Total			US\$1,600
Quality bonus	Score (50%) x 25% of volume		US\$200
		Total	US\$1,800
<u>Use of Funds</u>			
Drugs and consumables			US\$400
Outreach expenditures			US\$150
Repairs & maintenance of health facility			US\$150
Bonuses to staff in the facility			US\$900
Savings			US\$200

The other half of the LGAs in each state received funding under the DFF approach. Quarterly DFF payments were calculated to be equal to the average funds earned by the PBF facilities *net* of the performance bonus, (i.e. they were 50% of the average PBF payment). The payments were provided directly to the facilities. The funds could be used to finance operational costs but not performance bonuses for staff. Facilities were paid after the conditions in the DFF contract such as management arrangements, previous period fund utilization and transparent use of funds had been met. As is the case for PBF facilities, DFF

facilities sent monthly reports of a comprehensive quality score checklist, and these reports are verified on a quarterly basis.

Thus, there are four sub-components of the PBF intervention package that can stimulate improvements in quantity and quality of care- (i) additional financing, (ii) external supervision, (iii) autonomy and decentralized financing and (iv) incentives to health care workers. Note that DFF package includes the first three sub-components, and extent of these in DFF facilities is same as the extent in PBF facilities on average. Financial incentives to health workers are a crucial component of the innovations usually proposed under the PBF approach, and comparing the PBF and DFF arms allows us to determine whether these incentives deliver additional gains.

Figure 1: Project Description



LGAs are divided into wards; at least one health facility was chosen from each ward to participate in the program. Figure 1 illustrates the NSHIP setup, while Table 2 describes PBF and DFF interventions and also presents a comparative assessment of the implementation modality of the two approaches.

Table 2: Comparison of DFF and PBF interventions

Characteristic	PBF	DFF	Notes
Health facilities per ward	≥1	≥1	Selected using the same criteria
Health Facility Management Committee	Yes	Yes	Functions according to same rules and standards comprising the officer in charge of the facility and a representative of the ward development committee.
Autonomy of HF	Yes	Yes	Same amount of autonomy in use of funds, HR functions etc. except for bonuses to staff
Funds can be used to provide bonuses to their staff	Yes	No	DFF centers cannot use their funds to pay bonuses to their staff
Use of standard HMIS forms	Yes	Likely	DFF centers are encouraged to use agreed HMIS formats as part of LGA scorecard it is a condition of payment in PBF LGAs.
Quarterly invoice	Yes	No	Provided to steering committee
Monthly verification of quantity	Yes	No	SPHCDA of PBF-TA visits PBF facilities monthly to verify quantity; no verification visits in DFF
Quarterly supervision	Yes	Yes	Attached to LGA scorecard in DFF centers but not tied to bonuses
3rd party verification of quantity	Yes	No	
3rd party verification of quality	Yes	No	
Average amount of funding intended to be provided to HF, not including the administrative and the operational costs incurred at the state and federal levels	\$2	\$1	The DFF rate is pegged at the average of what a PBF facility can receive per capita <i>net</i> of the performance bonus
Bank accounts managed by facility committee	Yes	Yes	
Training and other support provided to strengthen the management	Yes	Yes	

Along with its attempts at improving service delivery through PBF and DFF, NSHIP also aims at strengthening institutional performance by using disbursement linked incentives (DLI) to the states and local governments on an annual basis. For example, local governments are rewarded for strengthening supervision of health facilities and publishing their annual budgets. States are rewarded for implementing PBF, and for increasing immunization coverage. States and LGAs are reimbursed for their eligible expenditure programs upon achievement of institutional performance outputs through the DLI approach.

The LGA performance scorecard comprises the following process indicators: health budget approved and published (20%), quarterly supervisions facilitated (20%), quarterly HMIS reports on four key indicators prepared (20%), staffing norms maintained (20%), and quality of drugs assured (20%). The State performance scorecard comprises: the proportion of PBF Health facilities receiving payments in a timely manner (20%), outpatient visits per capita (15%), full immunization of children (15%), institutional deliveries (15%), health budget approved and published (15%), annual State DLI Report and LGA DLI scorecard published (10%), and LGA DLI grants released in a timely manner (10%) (World Bank, 2011).

Since all the facilities in a project state receive NSHIP funding, in one form or other, facilities in other states form our reference point for understanding the impact that the entire intervention package had on the quantity and the quality of health care. Section 3 explains the details of the methodology employed to assess this impact.

2.2 Primary Project Objectives

The project development objective (PDO) is “to increase the delivery and use of high impact maternal and child health interventions and to improve the quality of care at selected health facilities in the participating states”. The project appraisal document outlines six indicators that will be tracked to measure the PDO (World Bank, 2011):

- (1) Proportion and number of 12-23 months-old children fully immunized
- (2) Proportion and number of birth attended by skilled health providers
- (3) Average health facility quality of care score
- (4) Number of curative care visits by children under five
- (5) Number of Direct Project Beneficiaries who are women

These indicators relate to the criteria used to supervise performance under PBF and DFF and to incentivize performance under PBF. This paper studies NSHIP impact on these outcomes.

The concept note for the impact evaluation of NSHIP, which was approved in 2011, listed several secondary indicators that relate to priority maternal and child health service availability, utilization and coverage, and technical quality of care for priority maternal and child health services. These outcomes are described in Tables 3 and 4.

Table 3: Priority MCH service availability, utilization and coverage outcomes of interest

Priority MCH Indicator	Service Availability	Service Utilization	Service Coverage
Skilled Birth Attendance	% of health facilities that offer delivery services in facility or skilled birth attendance in the community	# of facility deliveries or deliveries in communities attended by skilled personnel in the 30 days preceding the survey	% of births attended by skilled personnel in the two years preceding the survey % of births at a health facility in the two years preceding the survey
Immunization	% of health facilities that offered EPI immunizations on the week of the survey	# of immunizations offered in the 30 days preceding the survey	% of children 12-23 months old who are fully immunized
ITN Distribution			% of children under-5 years who slept under an ITN the night preceding the survey
ANC	% of health facilities that offered ANC services on the week of the survey	# of ANC visits in the 30 days preceding the survey	% of pregnant women who received 4 or more ANC visits (in the two years preceding the survey)
Curative Care for Children		# of curative care visits from children aged under five in the 30 days preceding the survey	

Table 4: Primary outcomes of interest- Technical and structural quality of care for priority MCH services

1	Proportion of on-duty technical staff present at health facility on the day of survey
2	At least one female clinical staff present on the day of survey
3	Proportion of health facilities with water for hand washing, soap and clean towel in patient examination area
4	Proportion of health facilities with at least one clean and functioning latrine
5	Proportion of health facilities with basic EPI equipment
6	Proportion of health facilities with EPI vaccines in stock on the day of the survey
7	Proportion of health facilities with basic ANC equipment
8	Proportion of health facilities with basic clinical equipment
9	Proportion of health facilities with basic delivery equipment
10	Number of essential drugs available on the day of the survey

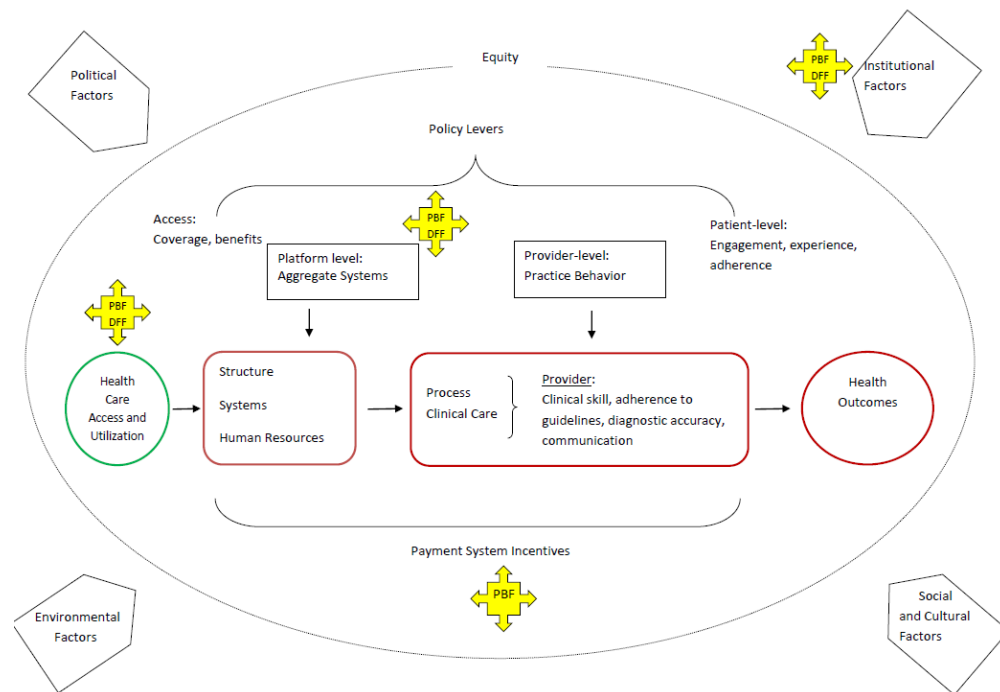
11	Average number of contraceptive methods in stock on the day of survey
12	Proportion of health facilities with bednets in stock on the day of the survey
13	Proportion of facilities with an up-to-date EPI register
14	Proportion of facilities with an up-to-date ANC and delivery register
15	Proportion of facilities with completed HMIS monthly report
16	Proportion of facilities that have a working waste disposal system (bin, pit or incinerator) in use and safety box for sharps
17	Proportion of facilities that can perform lab tests for malaria, TB, HIV and full blood count on the day of the survey
18	Proportion of facilities with working means of communication (radio, mobile phone, landline)
19	Proportion of facilities with a working vehicle to transport patients for referral
20	Proportion of health workers who report receiving their full salary on time
21	Proportion of health facilities that conduct outreach for key MCH services
22	Average health worker clinical knowledge score
23	Under-five examination quality score (based on IMCI protocols)
24	ANC examination quality score (based on national ANC protocols)
25	Average client satisfaction score
26	Proportion of clients who report that facility opening hours are convenient

NSHIP targets a large set of interrelated and multidimensional outcomes, and often lend themselves to data mining to cherry pick the results. Indeed, this is one of criticisms against the existing body of literature on PBF projects (Ireland et al., 2011). Therefore, we restrict our analysis to the indicators listed in the concept note which served as the study protocol. The indicators listed in the evaluation concept note are mapped to this framework.

2.3 Theory of Change

We focus on two features of the outputs generated under a health care system-- quantity and quality of care. Figure 2 describes the conceptual framework that guides the analysis in the following sections. Following Peabody et al. (2017), we define health care access and utilization as the first step towards the greater policy goal of improving maternal and child health care outcomes. These reflect two aspects of the quantity of care. By providing additional financing to health facilities, and also through additional supervision, NSHIP may improve the access to basic maternal and child care services. Access, however, is only one side of the coin; utilization is the other.

Figure 2: A General Conceptual Framework on the Effects of NSHIP



To understand the key factors that drive NSHIP’s performance, we employ the Donabedian conceptual framework of structure, process, and outcomes when assessing quality of care. Structure refers to stable, material healthcare assets (infrastructure, tools, technology, implements), the resources of the organizations providing care, and the financing of that care (levels of funding, staffing, payment schemes, incentives). Process denotes the interaction between caregivers and patients, including the provider’s cognitive skills and communication. Outcomes refer to direct measures of health status, death, or disability-adjusted life years as well as patient satisfaction or patient responsiveness to the healthcare system. By enhancing the incentives of the health facilities to improve their services, PBF and DFF components of NSHIP alter the aggregate systems and may induce improvements in the structural and procedural quality of care. Improvements in the quantity of care, and structures and processes may ultimately enhance the health status of the population.

By incentivizing health providers, the PBF component may have an additional impact on the structural and procedural quality of care. These impacts would ultimately be shaped by the local socio-economic factors that interact with the mechanisms that propel improvements under NSHIP.

3 Estimation and Data

3.1 Estimation Strategy

This impact evaluation uses a three-armed trial that yields experimental and quasi-experimental comparisons. The treatment states, Nasarawa, Adamawa and Ondo, were purposively chosen for the implementation of the project, but all LGAs in these states were randomly assigned to either PBF or DFF. To estimate the impact of NSHIP compared to business as usual, we need a control group of facilities that do not receive any financing or other interventions under NSHIP. These facilities serve as the benchmark against which the performance of NSHIP facilities would be measured. The concept note identifies three control states, one for each project state, chosen to be like the project states. Specifically, the control states were chosen by minimizing Euclidean distance to the treatment states on thirteen socio-economic and health indicators related to the utilization of maternal and neonatal health services available from DHS data. The control states are Benue for Nasarawa, Taraba for Adamawa and Ogun for Ondo (World Bank, 2011). Within control states, a comparison LGA was chosen for each treatment LGA using the same algorithm as state selection. We thus have six treatment LGAs in Nasarawa and six control LGAs in Benue, ten each in Adamawa and Taraba, and eight each in Ondo and Ogun. This IE thus provides quasi-experimental evidence of the effectiveness of PBF and DFF against business as usual and experimental evidence on PBF versus DFF.

This design allows us to carry out four sets of comparisons to understand how effective NSHIP was in achieving its goals. The first comparison is between the project states and business as usual or control states to estimate the impact of NSHIP as a complete package. The second is between facilities that receive PBF grants under NSHIP and control facilities. The third is between the facilities that receive DFF funds and control. The fourth comparison is between those LGAs that were randomly assigned to either the PBF or the DFF arms. This comparison, effectively a randomized control trial within treatment states, helps us understand if the incentive component of a typical supply-side PBF intervention is crucial in improving relevant outcomes, or if enhanced financing, decentralization, autonomy, and supervision can improve key MNH outcomes.

We use a double-difference method to estimate all four sets of comparisons:

$$y_{ijst} = \beta_0 + \beta_1 * Treatment_{js} + \beta_2 * Post_t + \beta_3 * Treatment_{js} * Post_t + \delta_s + \varepsilon_{ijst}$$

- Equation 1

where y_{ijst} is the outcome for household/health facility i in LGA j in state s in period t . $Treatment_{js}$ takes value 1 if the observation belongs to a project state and zero otherwise for the first three sets of comparisons. δ_s capture state fixed effects. For household level outcomes, standard errors are clustered at the level of primary sampling unit (PSU) – enumeration area (EA). For health facility level outcomes, we cluster the standard errors at the LGA level².

The double-difference method uses the trend in the comparison LGAs as an estimate of the counterfactual for the trend in the treatment LGAs. The validity of this approach crucially depends on the assumption that the average change in the comparison LGAs reflects the counterfactual change in the treatment LGAs if there were no treatment. This is called the *parallel trends assumption*. In the context of this study, the assumption of parallel trends implies that the trends in relevant outcomes across project and treatment states were the same before NSHIP started and can be tested using Equation 1 on pre-NSHIP data. If the assumption of parallel trends holds, the estimates of β_3 will be insignificant. Table 5 describes this test for socio-economic indicators available from the Nigerian DHS of 2008 and 2013. Although most of the indicators change at the same rate in the absence of the project between 2008 and 2013, some key outcomes like skilled birth attendance rate do not exhibit parallel trends across treatment and control states. To summarize, out of the twenty-two key household characteristics explored, parallel trends do not hold for eight. However, it is not always the case that control states are worse off than treatment states; in only one out of these twenty-two comparisons, control states higher rates of improvement. If the control states were always worse off than the treatment states, we would worry about their validity as counterfactuals. A comparison group that is consistently worsening would suggest that we were overestimating the impact of the program by comparing it to an artificially low counterfactual.

Table 5: Pre-Intervention Trends in Key Household Characteristics

	Year 2013	Project States	Year 2013*Project States (β_3)
If any child (younger than 24 months) in the HH ever received any vaccination	0.050	-0.081**	-0.069
	(0.034)	(0.035)	(0.047)
If the mother (of at least one child younger than 24 months) received skilled assistance during child birth	-0.002	-	0.129**
	(0.041)	0.197*** (0.040)	(0.056)

² The decomposition in the variation of some key outcomes at the baseline into within and between LGA components shows that within LGA variation is the dominant component.

	Year 2013	Project States	Year 2013*Project States (β_3)
If the mother (of at least one child younger than 24 months) received 4 or more ANC visits	0.024	0.020	0.062
	(0.036)	(0.036)	(0.050)
If the child (younger than 24 months) slept under an ITN the night before the survey	0.089***	0.001	-0.018
	(0.011)	(0.006)	(0.014)
If any child's birth in the HH has been registered	-0.010	-0.035**	0.063***
	(0.016)	(0.016)	(0.023)
If child in the HH is stunted	-0.022	-0.033**	0.035*
	(0.015)	(0.014)	(0.021)
If child in the HH is wasted	0.007	0.011	0.004
	(0.006)	(0.010)	(0.012)
If child in the HH is underweight	0.028**	0.007	-0.002
	(0.013)	(0.012)	(0.018)
Household has a literate woman	-0.011	-0.030	0.062
	(0.030)	(0.035)	(0.049)
Household has a woman who can read a newspaper	-	-0.021*	0.067***
	0.034***		
	(0.011)	(0.011)	(0.017)
Household has an employed woman	-0.043*	-	0.042
		0.125***	
	(0.025)	(0.026)	(0.038)
Women in the household use modern contraceptive methods	0.026*	-0.014	-0.008
	(0.016)	(0.016)	(0.024)
Household has a man who can read	0.025	0.077***	-0.066*
	(0.023)	(0.026)	(0.038)
Proportion of women who received skilled post-natal care (PNC)	0.047	0.043	0.002
	(0.033)	(0.034)	(0.044)
Proportion of women who received tetanus before birth	-0.038	-0.067**	0.121***
	(0.029)	(0.032)	(0.041)
Proportion of child births at an institution	-0.027	-	0.110**
		0.164***	
	(0.035)	(0.031)	(0.043)
Fever prevalence rate	-0.031	-	0.053**
		0.057***	
	(0.021)	(0.016)	(0.024)
Diarrhea prevalence rate	0.031*	-0.036**	0.012
	(0.017)	(0.014)	(0.022)
Proportion of households that have at least one ITN	0.519***	0.020	0.028
	(0.029)	(0.013)	(0.036)

	Year 2013	Project States	Year 2013*Project States (β_3)
Proportion of households with improved water source	-0.030 (0.046)	-0.021 (0.046)	0.054 (0.065)
Proportion of households with improved sanitation	0.009 (0.033)	0.030 (0.034)	-0.057 (0.050)
Proportion of households that have electricity	-0.060* (0.035)	-0.010 (0.038)	0.028 (0.056)
Clustered standard errors (at LGA level) in parentheses			
* p<0.1, ** p<0.05, *** p<0.01			
Source: Authors' calculations using DHS 2008 and 2013 data			

We also compare these indicators, and some additional variables that capture quality and quantity of healthcare using the baseline data we collected for the purposes of this evaluation. However, with the matched states design, we cannot establish balance at the baseline on several important variables (Tables 6). To allay any resulting concerns about the validity of the control comparison, we use the variables initially used to compute the Euclidian distance between the project and the control states to generate a propensity that a PSU falls in either a treatment or a control state. Following Hirano, Imbens & Ridder (2003), we then use this predicted probability to reweight the estimates obtained from equation 1.³ This reweighting process thus assigns a greater weight to the observations that are in the project states but should have been in control states. Units that were expectedly placed in the project states are assigned a lower weight because a lot of information already available on such units.

Table 6: Differences in Key Household Characteristics at the Baseline

Variable	Comparison States	NSHIP States	Difference
Female employment rate	0.656 (0.017)	0.57 (0.011)	0.086***
Proportion of children who have received all basic vaccination	0.234 (0.014)	0.299 (0.012)	-0.066***
Fever prevalence rate	0.043 (0.006)	0.027 (0.002)	0.016***
Diarrhoea prevalence rate	0.015 (0.002)	0.011 (0.001)	0.004
% of children 6-23 months old who received BCG	0.64	0.657	-0.017

³ For units in project states, weight is equal to the inverse of the predicted probability and for units in comparison states, this weight is equal to the inverse of (1-predicted probability).

Variable	Comparison States	NSHIP States	Difference
	(0.017)	(0.013)	
% of children (12-23mths) who receive Penta3	0.332	0.391	-0.059**
	(0.017)	(0.014)	
Under-5 curative care expenditure	252.032	218.279	33.753
	(24.355)	(17.083)	
Average years of completed education	6.31	6.945	-0.752**
	(0.231)	(0.172)	
Proportion of households with improved water source	0.394	0.353	0.055**
	(0.020)	(0.014)	
Proportion of households with improved sanitation	0.174	0.192	-0.027
	(0.016)	(0.012)	
Proportion of households that have electricity	0.252	0.346	-0.075**
	(0.023)	(0.018)	
Female literacy	0.446	0.49	-0.060**
	(0.020)	(0.013)	
Household has a woman who can read a newspaper	0.156	0.147	0.018*
	(0.011)	(0.007)	
Average number of children in a household	3.579	3.658	-0.064
	(0.066)	(0.046)	
Proportion of women who received skilled PNC	0.214	0.216	0.016
	(0.013)	(0.010)	
Proportion of births attended by skilled personnel	0.641	0.577	0.066***
	(0.020)	(0.014)	
Proportion of institutional Delivery	0.572	0.507	0.048***
	(0.020)	(0.014)	
Proportion of women who received tetanus before child birth	0.461	0.571	-0.106***
	(0.017)	(0.012)	
Proportion of households that have at least one ITN	0.636	0.707	-0.066***
	(0.019)	(0.011)	
Proportion of children who received IPT2 vaccines	0.097	0.173	-0.064***
	(0.008)	(0.009)	
Women in the household use modern contraceptive methods	0.204	0.169	0.010*
	(0.013)	(0.009)	
Clustered standard errors (at EA level) in parentheses			
* p<0.1, ** p<0.05, *** p<0.01			

Since LGA assignment to PBF or DFF arms was randomized, and was largely successful in ensuring that the averages of key household characteristics do not differ across these project arms (Table 7), we do not reweight these estimates.

Table 7: Differences in Key Household Characteristics at the Baseline (PBF versus DFF)

	DFF LGAs	PBF LGAs	Difference
Median completed years of education	7.12 (0.358)	6.93 (0.321)	-0.190 (0.479)
Proportion of households with improved water source	0.34 (0.035)	0.35 (0.037)	0.012 (0.051)
Proportion of households with improved sanitation	0.20 (0.025)	0.18 (0.020)	-0.021 (0.032)
Proportion of households that have electricity	0.31 (0.039)	0.35 (0.047)	0.041 (0.060)
Female literacy	0.49 (0.023)	0.50 (0.021)	0.006 (0.031)
Household has a woman who can read a newspaper	0.14 (0.010)	0.14 (0.016)	0.001 (0.019)
Average number of children in the household	3.54 (0.075)	3.76 (0.074)	0.222** (0.105)
Proportion of women who received skilled PNC	0.24 (0.024)	0.19 (0.023)	-0.046 (0.033)
Proportion of women who received tetanus before child birth	0.57 (0.029)	0.56 (0.031)	-0.017 (0.043)
Proportion of child births at an institution	0.52 (0.030)	0.47 (0.034)	-0.050 (0.045)
Proportion of households that have at least one ITN	0.67 (0.022)	0.76 (0.021)	0.094*** (0.031)
Proportion of children who received IPT2 vaccines	0.17 (0.024)	0.19 (0.029)	0.013 (0.038)
Women in the household use modern contraceptive methods	0.16 (0.013)	0.18 (0.015)	0.018 (0.020)
Female employment rate	0.59 (0.021)	0.55 (0.028)	-0.035 (0.035)
Proportion of children who have received all basic vaccination	0.24 (0.035)	0.37 (0.034)	0.124** (0.049)
Fever prevalence rate	0.02 (0.006)	0.03 (0.008)	0.009 (0.010)
Diarrhea prevalence rate	0.01 (0.002)	0.01 (0.003)	0.006* (0.004)
Clustered standard errors (at EA level) in parentheses			

	DFF LGAs	PBF LGAs	Difference
* p<0.1, ** p<0.05, *** p<0.01			

3.2 Sampling

One health center in each ward of each sampled LGA was selected to be a part of the sample. In addition, three health workers were interviewed at each facility. Households were chosen in the following way. The National Population Commission of Nigeria listed all enumeration areas in the country for the 2006 census of Nigeria. In 2008, the Federal Ministry of Health used these enumeration areas to create facility catchment areas. These were our PSUs. The power calculation results determined that the number of PSUs per site is 17 for Nasarawa; 12 for Ondo; 10 for Adamawa; 17 for Benue; 12 for Ogun; and 10 for Taraba.

After determining the number of PSUs per state, the next step is to select sample PSUs. Each EA has some conspicuous natural or man-made features as boundaries. During EA delineation, states and LGAs were canvassed in geographical order, from west to east and back to the west in a serpentine fashion. The catchment areas within the LGAs are thus contiguous so that no one catchment area boundary overlaps the other and no area is omitted. The catchment areas were coded serially one after the other using map orientation in a serpentine order. From this list, PSUs were randomly selected on a probability proportional to size basis.

The last step entailed the selection of households. A household listing exercise was carried out in every PSU. The listing form was designed to obtain information about the address of buildings, number of households in the listed buildings, name of the head of each household listed, and presence of woman in the household with at least one pregnancy or birth in the two years preceding the survey. Using this list as the sampling frame, 15 households with a woman who had experienced at least one pregnancy during the last two years were selected from every PSU. If a PSU had fewer than 15 households, the listing exercise extended into the contiguous PSU.

Note that inference is somewhat limited by the representativeness of our results: given that project states were purposively chosen and that households were selected from the catchment areas of our study facilities, our results are representative of the project at the state level, and not of the entire country or even state.

.

3.3 Data

The data for this evaluation comes from two rounds of surveys commissioned by the Nigerian Federal Ministry of Health with the assistance of the World Bank. Baseline data were collected between February and April of 2014. The endline data were collected between August and October of 2017. A set of four surveys were administered at health facilities in both rounds across the six project and comparison states. These included a health facility survey, an interview with a health provider at the facility, direct observation of patient and health provider interactions for antenatal and under-5 curative care visits as well as exit interviews with the patients whose care was observed.

Data from 786 health facilities were collected at the baseline. 2,250 health providers were interviewed at these facilities. In addition, the interactions of 1,959 antenatal care patients and 1,778 caregivers of patients seeking under-5 curative care were observed, and the patients (or caregivers) were interviewed. The same 786 facilities were also surveyed at the endline. 2,640 antenatal care patients and 2,519 patients seeking under five curative care were included in the endline. In addition, 7,683 households were interviewed at baseline and 7,527 households at endline from the same 848 PSUs. The household questionnaire collected data on socio-economic and demographic conditions of these households along with detailed health histories, anthropometry, and the woman's pregnancy and delivery.

This sampling strategy limits inference to some extent: project states were purposively chosen and households were selected from the catchment areas of our study facilities. Our results are thus representative of the project at the state level and not of the entire country or even state. Finally, as noted above, while the PBF versus DFF relies on randomized assignment of LGAs to the two arms, the control comparisons are based on purposively selected states and are quasi-experimental in design.

4 Baseline Survey

4.1 Health Facility Survey

Table 8 presents summary statistics on characteristics of health facilities at baseline. Most sampled facilities provided primary-level care, and 87% were primary health centers. 71% reported offering services around the clock. On average, facilities provided ANC clinics once a week and under-five care clinics twice a week. 11% health facilities had separate hours for adolescents. The average distance between a primary health clinic and the nearest secondary hospital was 31 kilometers. On average, a facility had 13 staff members, half of whom belonged to technical cadres. Physical infrastructure can be poor: approximately 87% of the facilities did not have access to transportation for patients and 79% experienced power outage in the week that preceded the survey. The availability of water supply was better, but 16% of the sampled facilities had experienced outages in water supply in the week before the survey. These facilities were not well-connected in terms of communications: only 3% had access to a two-way radio system and 12% had access to a phone line. Only 4% of these facilities had internet connectivity, while 10% had a functioning computer. In addition to poor access to means of transportation and communication, the access to medical equipment was limited: 15% of the facilities had basic delivery equipment, 44% had basic ANC equipment, and 12% had basic immunization equipment. A little over half of facilities had an improved water source, basic sanitation capacity, and a system for disposal of medical waste. Facilities only had 34% of a list of 18 essential drugs for primary care in stock on the day of the survey.

Table 8: Baseline Health Facility Characteristics

Facility Characteristic	Mean	Std.	Obs.
Level: Primary	0.94	0.24	1,007
Level: Secondary	0.06	0.24	1007
Type: Primary Health Center	0.87	0.33	1007
Type: Health Post	0.07	0.25	1007
Type: General Hospital	0.04	0.20	1007
Type: Other Secondary	0.01	0.12	1007
Year facility began operating	1992	16.79	911
Offers services around the clock	0.71	0.46	994
Days offering antenatal care clinics	1.10	0.97	899
Days offering under-5 clinics	1.71	2.15	728
Distance to nearest hospital (km)	31.44	35.87	982
Facility has a doctor as head	0.04	0.20	909
Number of staff members	12.91	35.31	838

Facility Characteristic	Mean	Std.	Obs.
Proportion of staff: technical staff ⁴	0.52	0.25	806
Access to any kind of transportation to take patients to referral facility	0.13	0.34	1007
Power unavailable at any time in past 7 days	0.79	0.41	442
Have backup generator	0.48	0.50	352
Improved water source ⁵	0.60	0.49	1004
Water unavailable at any time in past 7 days	0.16	0.37	777
Soap and water in patient examination area	0.64	0.48	1007
Clean and functioning latrine	0.53	0.50	1007
Basic delivery equipment	0.15	0.36	1003
Basic ANC equipment	0.44	0.50	1007
Basic clinical equipment	0.05	0.22	1007
Basic EPI equipment	0.12	0.33	1006
Fully stocked Expanded Program on Immunization (EPI) vaccines ⁶	0.08	0.28	1007
Percent of essential drugs available on day of survey	0.34	0.27	1007
Waste disposal system and sharps disposal box	0.57	0.50	1007
Functioning two-way radio	0.03	0.17	1007
Phone line	0.12	0.33	1007
Radio and/or phone line	0.14	0.35	1007
Functioning computer	0.10	0.31	999
Internet connectivity	0.04	0.20	999

4.2 Household Survey

Table 9 describes the characteristics of the households surveyed. The average household had six members, two of whom were under five years of age. 73% owned the land they resided on or had other landholdings. 97% of the sampled households were headed by men. The average age of the head of the household was 38 years; 26% had at least primary education. Half of the households had access to improved water source during the dry season and 37% only during the rainy season. A quarter of the households had access to improved sanitation facilities, and 7% treated their drinking water appropriately.

Table 9: Baseline Household Characteristics

Household Characteristic	Mean	Std.	Obs.
State			
Adamawa	0.24	0.43	9567

⁴ Technical staff are defined as doctor, medical officer, (auxiliary) nurse midwife, nurse, midwife, nurse midwife, community health officer, (junior) community health extension worker.

⁵ Improved water source is defined as piped into facility or yard/plot, public tap or standpipe, protected well, or borehole/tubewell.

⁶ EPI vaccines are BCG, Penta (or Haemophilus Influenza B, diphtheria, pertussis, tetanus, and hepatitis B), oral Polio, Yellow Fever, Measles

Household Characteristic	Mean	Std.	Obs.
Benue	0.14	0.34	9567
Nasarawa	0.20	0.40	9567
Ogun	0.11	0.31	9567
Ondo	0.21	0.41	9567
Taraba	0.11	0.31	9567
Number of household members	6.15	2.98	9567
Number of children under-5	1.72	0.90	9567
Household owns land	0.73	0.44	9419
Access to improved water source: dry season	0.50	0.50	9550
Access to improved water source: rainy season	0.37	0.48	9548
Appropriate treatment of drinking water: dry season	0.07	0.26	9465
Appropriate treatment of drinking water: rainy season	0.06	0.23	9499
Access to improved sanitation facilities	0.25	0.43	7708
Access to improved shared sanitation facilities	0.24	0.43	7611
Access to non-improved sanitation facilities	0.01	0.08	5826
Head of Household Characteristics	Mean	Std.	Obs.
Male	0.97	0.18	9590
Age	38.01	11.01	9572
Education: No school	0.05	0.21	7694
Education: Primary School	0.26	0.44	7694
Education: More than primary	0.69	0.46	7694

Among the surveyed households, women who were currently pregnant or had a birth in the previous two years were interviewed using a women's questionnaire. Table 10 describes the characteristics of these respondents. 14% were Catholic, 53% other Christian, and 33% Muslim. Half of the women spoke Hausa for the interview, followed by Yoruba (27%), English (13%). 37% of the female respondents had at least some primary schooling, 53% some secondary schooling, and 10% more than secondary. For their most recent birth, 59% reported the presence of a skilled birth attendant and 52% delivered in a health facility.

Table 10: Baseline Household Characteristics

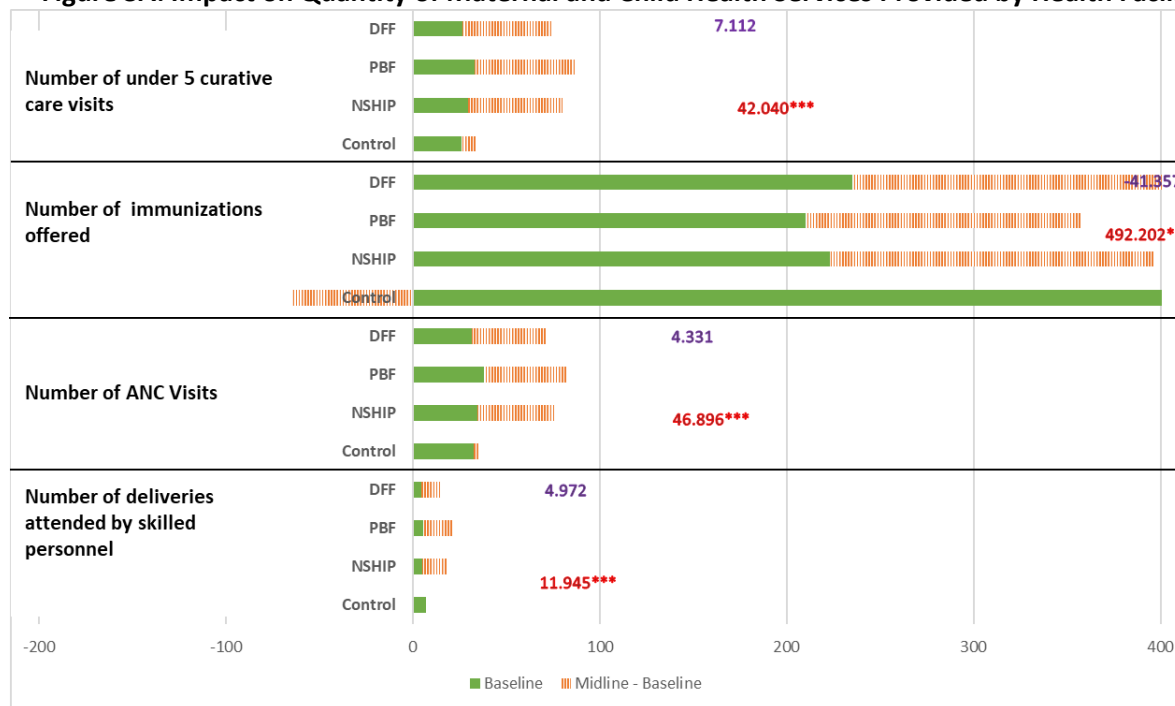
Characteristics of Women's Survey Respondents	Mean	Std.	Obs.
Religion: Catholic	0.14	0.35	9834
Religion: Other Christian	0.53	0.50	9834
Religion: Muslim	0.33	0.47	9834
Religion: Other	0.01	0.07	9834
Language: English	0.13	0.34	9756
Language: Hausa	0.50	0.50	9756
Language: Yoruba	0.27	0.44	9756
Language: Igbo	0.0002	0.01	9756
Language: Other	0.09	0.29	9756
Education: primary school	0.37	0.48	6796
Education: secondary school	0.53	0.50	6796
Education: more than secondary school	0.10	0.30	6796

Characteristics of Women's Survey Respondents	Mean	Std.	Obs.
Most recent birth: N/A	0.01	0.07	9867
Most recent birth: Skilled birth attendant	0.59	0.49	9798
Most recent birth: Institutional delivery	0.52	0.50	9656

5 Results

This section presents estimates of the impact of NSHIP on the indicators listed in Tables 3 and 4. All tables present estimates of β_3 from Equation 1. Of the six PDO indicators (four facility level, presented in Figure 3A and two household level, presented in Figure 3B), four improved as a result of NSHIP: curative care visits by children under five, the number of immunizations offered, the number of ANC visits to facilities, and the facility-level quality scores, and the proportion of children fully immunized (figure 3B). Impact estimates presented in Table 12 show that while the number of birth attended by skilled health providers showed gains in NSHIP states compared to control states, this gain was driven by PBF LGAs, and not DFF LGAs.

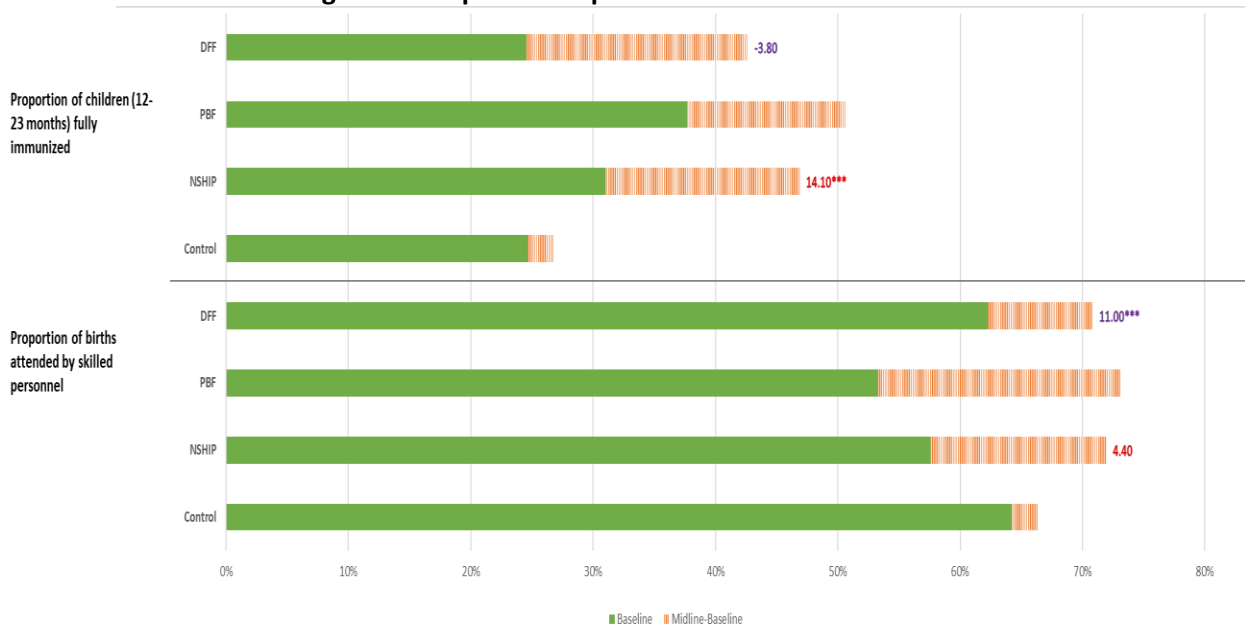
Figure 3A: Impact on Quantity of Maternal and Child Health Services Provided by Health Facilities



Note: Weighted double-difference estimates for comparisons between NSHIP and control states are in red, between PBF and DFF LGAs are in purple.

*** p<0.1, ** p<0.05, *** p<0.01**

Figure 3B: Impact on Population Level PDO Indicators



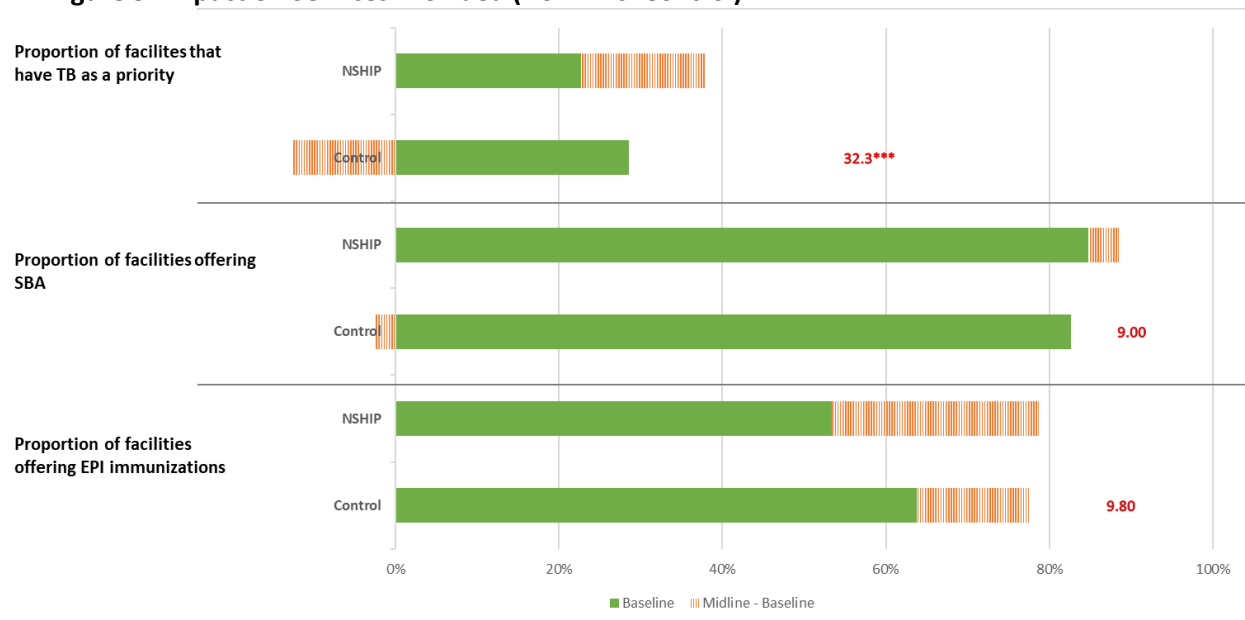
Note: Weighted double-difference estimates for comparisons between NSHIP and control states are in red, between PBF and DFF LGAs are in purple.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

5.1 Quantity of Healthcare

5.1.1 Access to Healthcare

Figure 5: Impact on Services Provided (NSHIP vs. Control)



Note: Weighted double-difference estimates for comparisons between NSHIP and control states are in red.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

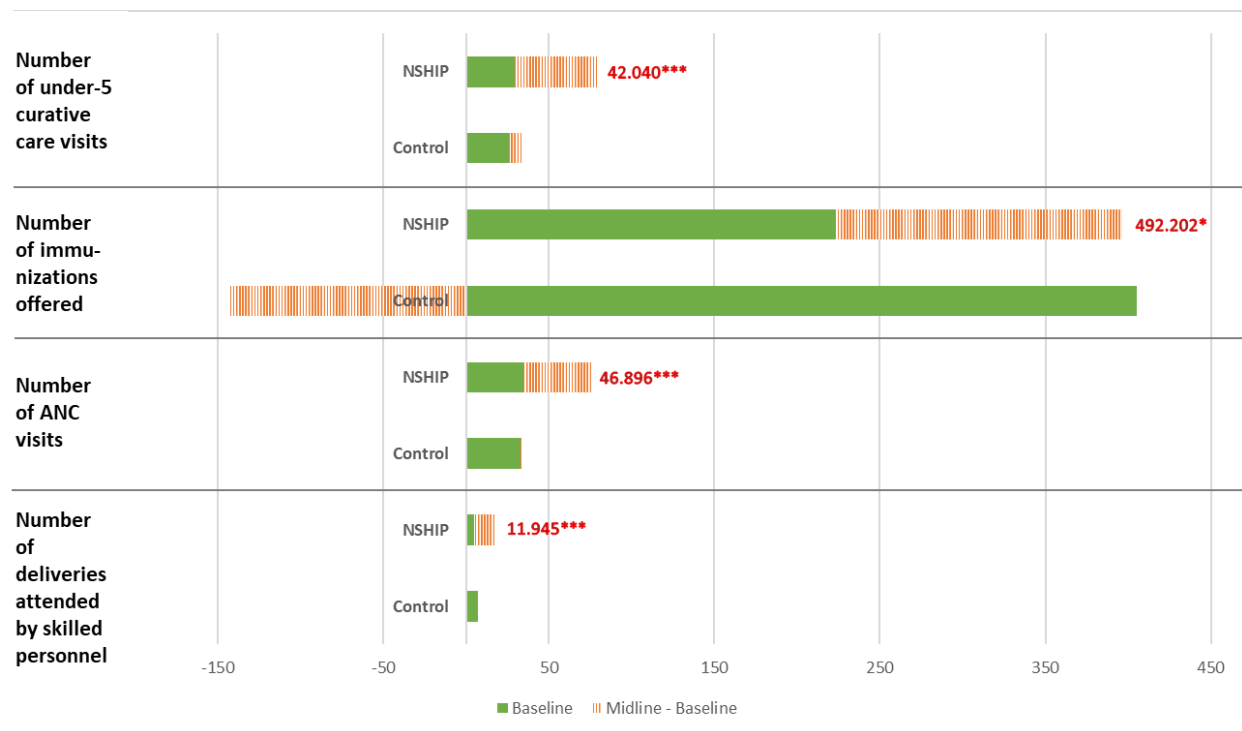
The increase in the number of facilities that offered routine immunizations during the week of the survey was not significant for the full NSHIP sample, although there were significant increases (13.2 pp, or 23% of the baseline average for facilities) in PBF facilities. Similarly, the number of facilities that offered inpatient delivery services or skilled birth attendance in the community did not increase in response to either arm of NSHIP, but 94% of the facilities were providing these services at the baseline, leaving limited scope for improvement. The proportion of health facilities that reported that tuberculosis is a priority increased by 32.2 pp, a sizeable improvement over the baseline of 24.95%. This increase was driven equally by PBF (34.2 pp) and DFF facilities (30.9 pp).

Table 11: Impact on Healthcare Service Delivery

	NSHIP vs Control	PBF vs Control	DFF vs Control	PBF vs DFF	Baseline Mean	N
% of health facilities that offered routine immunizations in the week of the survey	0.098	0.132*	0.069	0.087	0.62	1490
	(1.29)	(1.71)	(0.71)	(1.10)		
% of health facilities that offer delivery services in facility or skilled birth attendance in the community	0.09	0.079	0.101	-0.03	0.94	1515
	(1.25)	(1.06)	(1.37)	(1.10)		
% of facilities that report TB as a priority	0.323***	0.342***	0.309***	0.033	0.30	1524
	(5.59)	(5.08)	(4.48)	(0.42)		
Standard errors clustered at LGA level						
t statistics in parentheses						
* p<0.1, ** p<0.05, *** p<0.01						

5.1.2 Utilization of Healthcare

Figure 5: Impact on Utilization to Healthcare



Note: Weighted double-difference estimates for comparisons between NSHIP and control states are in red, between PBF and DFF LGAs are in purple.

*** $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$**

We also observe significant increases in the number of in-facility deliveries and births attended by skilled personnel in the community in the 30 days that preceded the survey. The impact estimate was twice the average at the baseline. The impact for PBF facilities is 14.8 deliveries and for DFF facilities was 9.4 deliveries. In addition, the number of ANC visits to NSHIP facilities increased significantly: 48.2 visits for PBF facilities, and 45.6 visits for DFF facilities, versus a baseline control average of 34.2 visits. The provision of immunization services also showed significant improvements because of NSHIP: the number of immunizations offered in a month at NSHIP facilities increased by 492.2; the control average at the baseline was only 272.6. However, this impact was not significant for PBF facilities at conventional levels of significance but was significant and higher in DFF facilities. NSHIP increased the number of curative care visits from children under five in the 30 days preceding the survey by 42 visits by 142% over the baseline control average of 29 visits. The increase for PBF facilities was 45 visits and for DFF facilities was 39 visits (Table 12).

Table 12: Impact on Healthcare Utilization

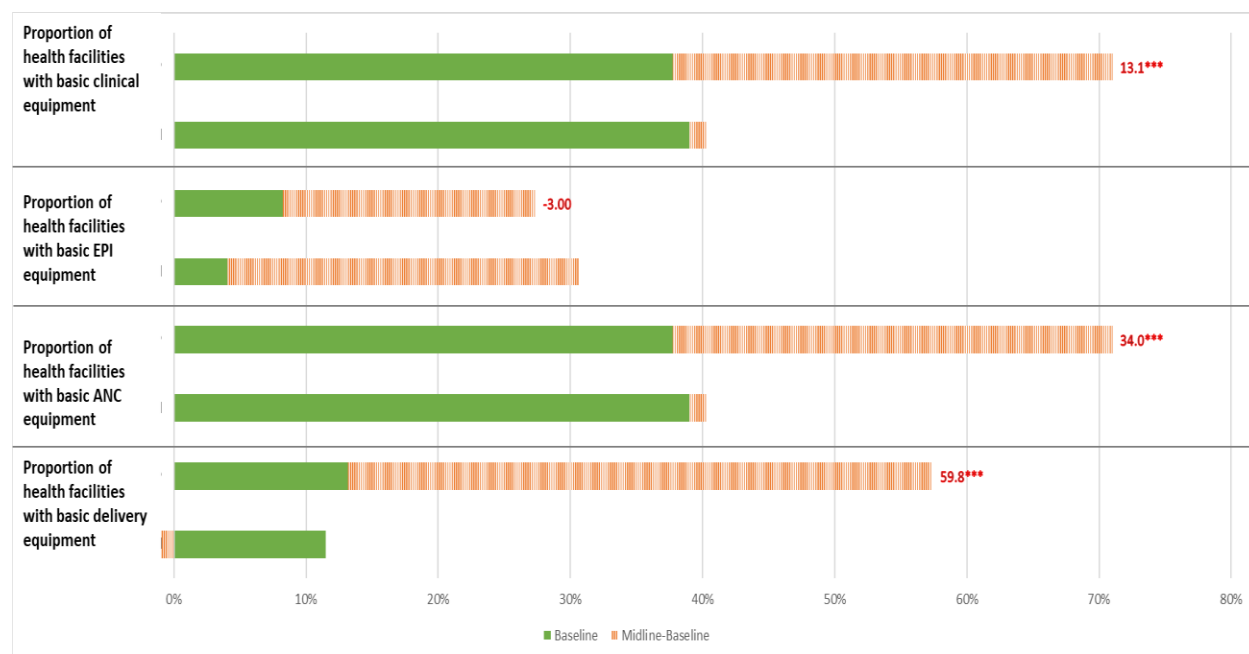
	NSHIP vs Control	PBF Vs Control	DFF Vs Control	PBF Vs DFF	Baseline Mean	N
Number of facility deliveries or deliveries in communities attended by skilled personnel in the 30 days preceding the survey	11.945***	14.813***	9.415***	4.972	22.8	1516
	(5.23)	(6.02)	(3.24)	(1.48)		
Number of ANC visits in the 30 days preceding the survey	46.896***	48.235***	45.614***	4.331	34.2	1526
	(4.06)	(3.62)	(3.71)	(0.42)		
Number of immunizations offered in the 30 days preceding the survey	492.202*	469.31	509.792*	-41.357	272.6	1490
	(1.72)	(1.64)	(1.76)	(0.61)		
Number of curative care visits from children aged under five in the 30 days preceding the survey	42.040***	45.300***	39.311***	7.112	28.6	1526
	(4.01)	(4.16)	(2.85)	(0.46)		
Standard errors clustered at LGA level						
t statistics in parentheses						
* p<0.1, ** p<0.05, *** p<0.01						

Quality of Healthcare

5.1.3 Structure

5.1.3.1 Equipment

Figure 6: Impact on Structural Quality of Healthcare, Equipment



Note: Weighted double-difference estimates for comparisons between NSHIP and control states are in red, between PBF and DFF LGAs are in purple.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

We observe a substantive improvement in the availability of basic delivery equipment at in the NSHIP facilities. To quantify this, we examine whether a facility had over 60% of the 29 basic pieces of equipment required for a safe delivery. Only 16.6% of the control facilities at the baseline had these. Significant improvements were observed in both PBF and DFF facilities: 59.8 pp and 68.8 pp, respectively. This increase in the availability of the basic delivery was observed in both PBF and DFF facilities but was significantly higher in the PBF facilities. Next, we consider the inputs to ANC provision. To quantify the availability of basic ANC equipment, we check if the facility has a fetoscope, a measuring tape, a blood pressure machine and an adult weighing scale. Although almost half of the facilities already had all ANC equipment available at the baseline, we still observe a 70% increase in the availability of these supplies for both PBF and DFF health facilities. However, there were no improvements in the availability of basic

EPI equipment⁷. Availability of basic clinical equipment improved across both the NSHIP financing modes (Table 13).

Table 13: Impact on Structural Quality of Care, Equipment

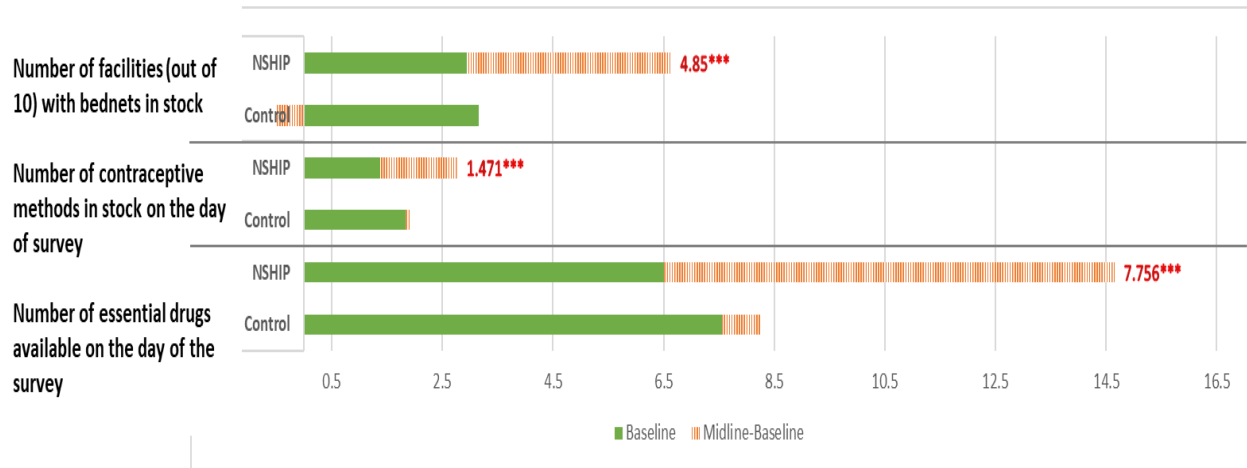
	NSHIP vs Control	PBF vs Control	DFF vs Control	PBF vs DFF	Baseline Mean	N
Proportion of health facilities with basic delivery equipment	0.598*** (11.02)	0.688*** (12.06)	0.518*** (8.22)	0.134** (2.06)	0.17	1516
Proportion of health facilities with basic ANC equipment	0.340*** (4.54)	0.340*** (3.87)	0.342*** (4.54)	-0.037 (0.70)	0.49	1526
Proportion of health facilities with basic routine immunization equipment	-0.003 (0.05)	0.074 (0.93)	-0.077 (0.94)	0.158** (2.14)	0.09	1490
Proportion of health facilities with basic clinical equipment	0.131*** (4.54)	0.129*** (3.24)	0.135*** (4.48)	0.003 (0.08)	0.06	1526
Standard errors clustered at LGA level						
t statistics in parentheses						
* p<0.1, ** p<0.05, *** p<0.01						

5.1.3.2 Drugs and Supplies

The number of essential drugs available on the day of the survey increased by 7.8 where the corresponding average for facilities in project states at the baseline was 6.5. The average number of contraceptive methods in stock on the day of survey increased by 106% over a baseline average of 1.38. The impact on the availability of bednets is 48.5 pp and was similar across PBF and DFF facilities (Table 14); only 30% of the facilities had insecticide-treated bed nets at the baseline.

⁷ BCG, DPT, Oral Polio, Tetanus, HepB, Yellow Fever and Measles

Figure 7: Impact on Structural Quality of Healthcare, Drugs and Equipment



Note: Weighted double-difference estimates for comparisons between NSHIP and control states are in red, between PBF and DFF LGAs are in purple.

*** p<0.1, ** p<0.05, *** p<0.01**

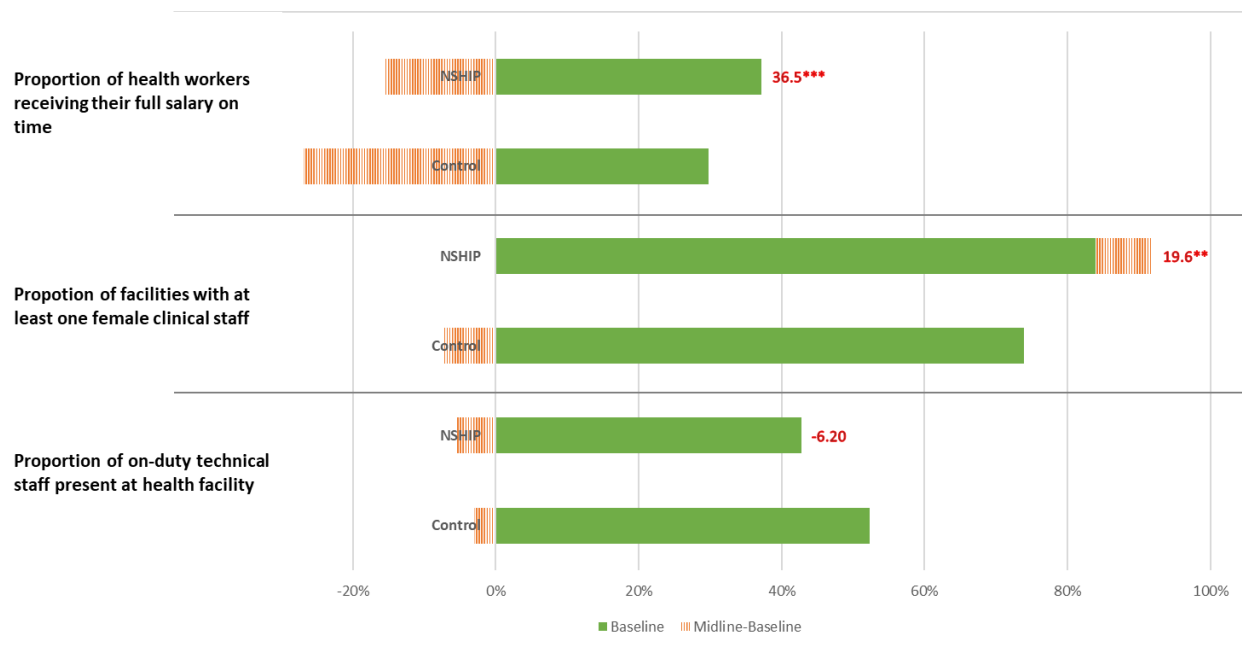
Table 14: Impact on Structural Quality of Healthcare, Drugs and Equipment

	NSHIP vs Control	PBF vs Control	DFF vs Control	PBF vs DFF	Baseline Mean	N
Number of essential drugs available on the day of the survey	7.756** * (12.43)	8.490** * (11.16)	7.114** * (10.53)	1.290* (1.73)	6.8	1526
Average number of contraceptive methods in stock on the day of survey	1.471** * (7.26)	1.729** * (8.00)	1.240** * (6.12)	0.426** * (2.91)	1.5	1526
Proportion of health facilities with bednets in stock on the day of the survey	0.485** * (5.15)	0.486** * (4.78)	0.487** * (4.37)	0.006 (0.06)	0.30	1526
Standard errors clustered at LGA level						
t statistics in parentheses						
* p<0.1, ** p<0.05, *** p<0.01						

5.1.3.3 Staffing

Next, we explore improvements in the quality of human resource practices at health facilities. While the impact on the proportion of on-duty technical staff present at the health facility on the day of the survey is not different from zero, the impact on the presence of at least one female health worker was 19.6 pp. These improvements were observed at both PBF and DFF facilities. A major innovation of NSHIP was to use the tools of decentralized financing to streamline the payment structure of health workers. We find that NSHIP was successful in achieving this goal by increasing the proportion of health workers who receive their full salary on time by 36.5 pp over the baseline average of 53.8 pp. The increase for PBF LGAs was 35.2 pp and for DFF LGAs was 37.9 pp (Table 15).

Figure 8: Impact on Structural Quality of Healthcare, Staffing



Note: Weighted double-difference estimates for comparisons between NSHIP and control states are in red, between PBF and DFF LGAs are in purple.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

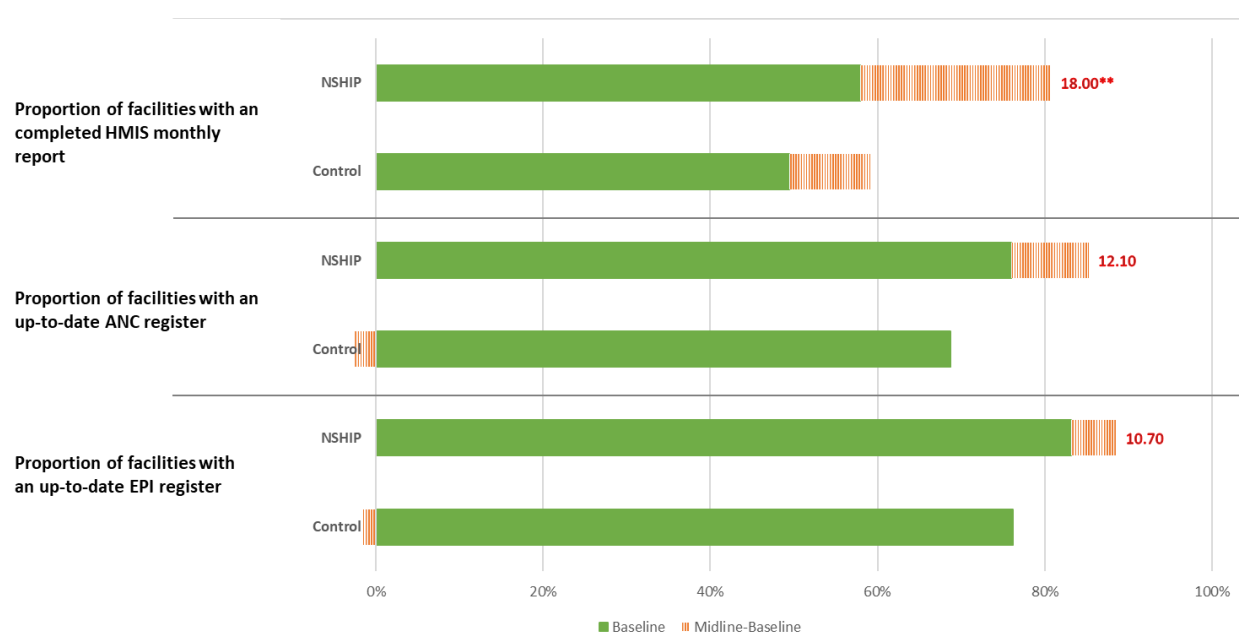
Table 15: Impact on Structural Quality of Healthcare, Staffing

	NSHIP vs Control	PBF vs Control	DFF vs Control	PBF vs DFF	Baseline Mean	N
Proportion of on-duty technical staff present at health facility on the day of survey	-0.062 (0.91)	-0.021 (0.31)	-0.100 (1.37)	0.063* (1.97)	0.50	1411
At least one female clinical staff present on the day of survey	0.196** (2.13)	0.202** (2.16)	0.190** (2.01)	0.009 (0.24)	0.84	1505
Proportion of health workers who report receiving their full salary on time	0.365*** (5.34)	0.352*** (4.11)	0.379*** (4.65)	-0.037 (0.38)	0.57	1505
Standard errors clustered at LGA level						
t statistics in parentheses						
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$						

5.1.3.4 Recordkeeping

Standard formats for record keeping are crucial to standardize patient records, which in turn improves the completeness, accessibility, and accuracy of patient information exchanged between healthcare providers and institutions. We do not observe an impact of NSHIP on recordkeeping practices for routine immunization or ANC services relative to the control group, but there was a greater improvement for DFF facilities than for PBF facilities. The proportion of facilities with complete health management information system (HMIS) reporting at the time of the survey was significantly higher at NSHIP facilities than in control facilities. This improvement was entirely driven by the improvements at DFF facilities (Table 16).

Figure 9: Impact on Structural Quality of Healthcare, Recordkeeping



Note: Weighted double-difference estimates for comparisons between NSHIP and control states are in red, between PBF and DFF LGAs are in purple.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

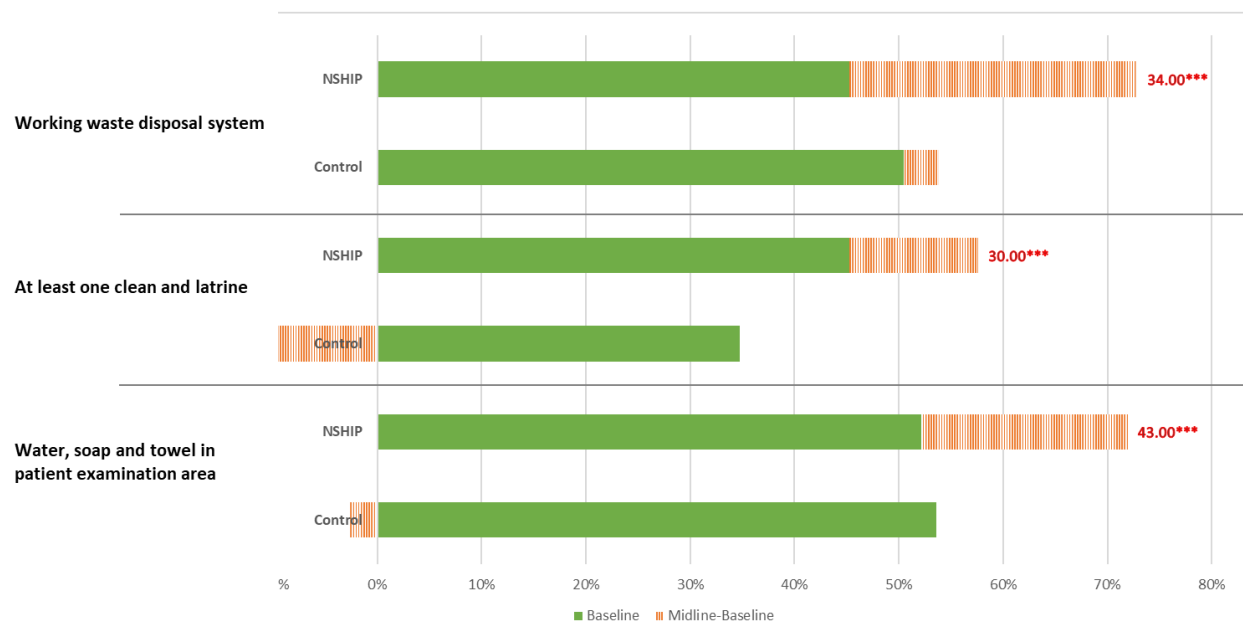
Table 16: Impact on Structural Quality of Healthcare, Recordkeeping

	NSHIP vs Control	PBF vs Control	DFF vs Control	PBF vs DFF	Baseline Mean	N
Proportion of facilities with an up-to-date routine immunization register	0.107 (1.33)	0.083 (1.01)	0.132 (1.60)	-0.050** (2.23)	0.90	1525
Proportion of facilities with an up-to-date ANC and delivery register	0.121 (1.38)	0.118 (1.30)	0.126 1.38	-0.008 (0.17)	0.82	1489
Proportion of facilities with completed HMIS monthly report	0.180** (2.05)	0.12 (1.28)	0.242** (2.57)	-0.125* (1.90)	0.61	1525
Standard errors clustered at LGA level						
t statistics in parentheses						
* p<0.1, ** p<0.05, *** p<0.01						

5.1.3.5 Sanitation, Hygiene, and Waste Management

NSHIP's impact on access to basic sanitation infrastructure was significant and large: the proportion of health facilities that had water, a soap and a towel in patient examination area increased by 38.3 pp for PBF facilities and 48.4 pp for DFF facilities. Availability of a functional latrine also improved by almost 50% from the average at the baseline. The impact on PBF facilities was 24.5 pp while that on the DFF facilities was 35.2 pp. Impact on the proportion of facilities that had a working waste disposal system was 33.6 pp for PBF facilities and 35.4 pp for DFF facilities. The corresponding average at the baseline was 60% (Table 19).

Figure 10: Impact on Structural Quality of Healthcare, Sanitation



Note: Weighted double-difference estimates for comparisons between NSHIP and control states are in red, between PBF and DFF LGAs are in purple.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

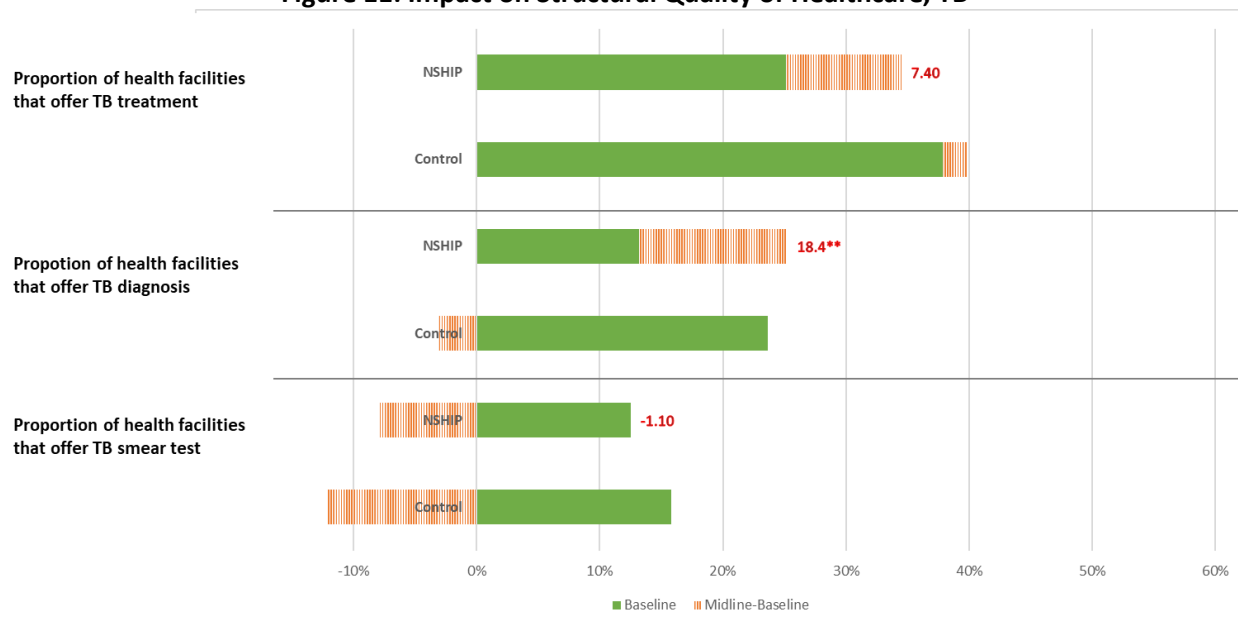
Table 17: Impact on Structural Quality of Healthcare, Sanitation

	NSHIP vs Control	PBF vs Control	DFF vs Control	PBF vs DFF	Baseline Mean	N
Proportion of health facilities with water for hand washing, soap and clean towel in patient examination area	0.434*** (5.92)	0.383*** (4.97)	0.484*** (6.08)	-0.08 (1.50)	0.69	1525
Proportion of health facilities with at least one clean and functioning latrine	0.299*** (3.56)	0.245** (2.45)	0.352*** (4.11)	-0.064 (0.88)	0.60	1525
Proportion of facilities that have a working waste disposal system (bin, pit or incinerator) in use and safety box for sharps	0.343*** (5.08)	0.336*** (4.29)	0.354*** (4.87)	-0.074 (1.20)	0.61	1525
Standard errors clustered at LGA level						
t statistics in parentheses						
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$						

5.1.3.6 Tuberculosis Detection and Care

Although the proportion of facilities that offered TB smear test on the day of the survey, and that offered TB treatment did not increase in NSHIP states, there was a 18.5 pp increase in proportion of facilities that offered TB diagnosis. Note that only 13% health facilities in project states offered TB diagnosis at the baseline. The increase for PBF facilities was 19.1 pp and for DFF facilities was 17.7 pp.

Figure 11: Impact on Structural Quality of Healthcare, TB



Note: Weighted double-difference estimates for comparisons between NSHIP and control states are in red, between PBF and DFF LGAs are in purple.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 18: Impact on Structural Quality of Healthcare, TB

	NSHIP vs Control	PBF vs Control	DFF vs Control	PBF vs DFF	Baseline Mean	N
Proportion of facilities that offer TB smear test on the day of the survey	-0.011 (0.16)	0.003 (0.05)	-0.028 (0.34)	0.017 (0.27)	0.17	948
Proportion of facilities that offer TB diagnosis	0.184** (2.48)	0.191** (2.30)	0.177** (2.27)	0.04 (0.63)	0.16	1523

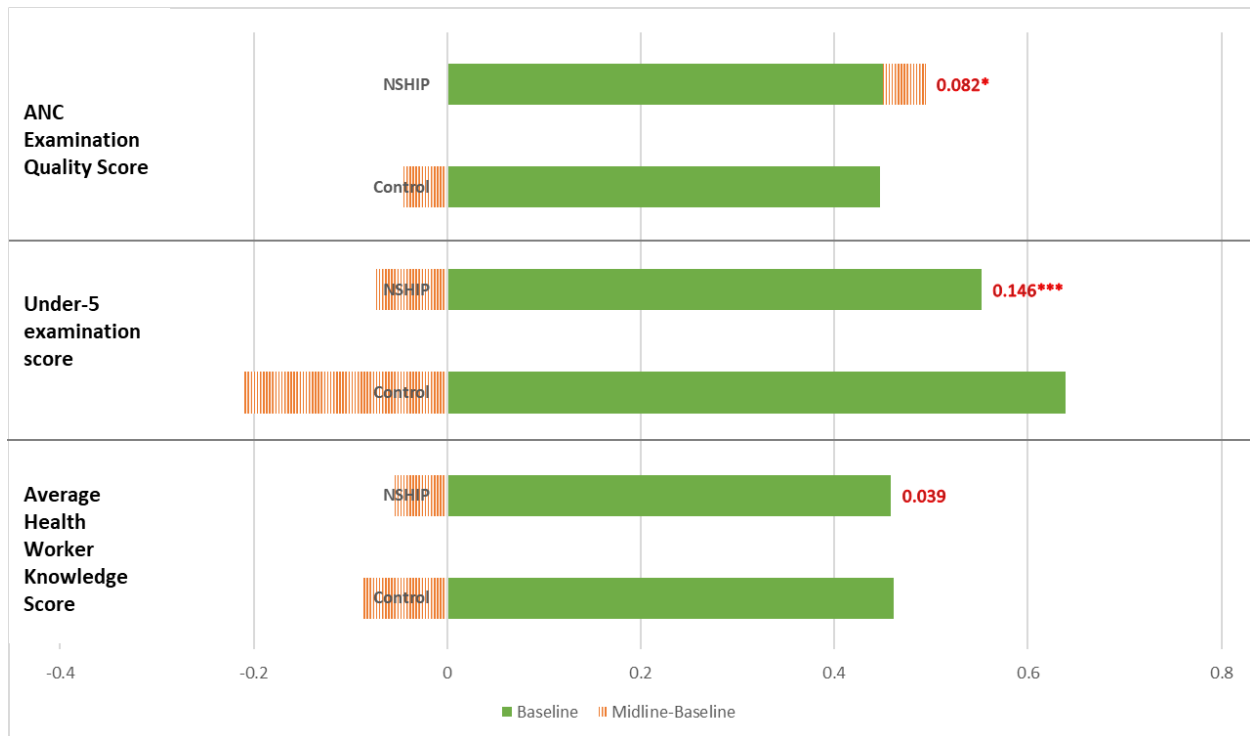
	NSHIP vs Control	PBF vs Control	DFF vs Control	PBF vs DFF	Baseline Mean	N
Proportion of facilities that offer TB treatment	0.074	0.076	0.077	-0.007	0.29	1523
	(1.27)	(1.20)	(1.17)	(0.12)		
Standard errors clustered at LGA level						
t statistics in parentheses						
* p<0.1, ** p<0.05, *** p<0.01						

5.1.4 Process

An important aspect of quality of care is the nature of the interaction between the patient and the health provider, that is the process quality of healthcare. Indicators related to the cognitive ability of the health provider and the appropriateness of the clinical care used capture the process quality of healthcare. Although we do not see any increase in the average health worker clinical knowledge score, NSHIP increased the under-five examination quality score by 45% over the average at the baseline, and the ANC examination quality of care score increased by about 15% over its baseline average. The increase in the under-five examination quality score was 17.5 pp PBF facilities. The improvements in the ANC examination quality score were almost entirely being driven by the improvements at DFF facilities. The procedural quality of care also improved due to better infrastructure that was made available by NSHIP.

Table 19 shows that NSHIP increased the proportion of facilities that had a working means of communication (radio/mobile phone/landline phone) by 57.3 pp, with the corresponding impact for PBF facilities being 66.2 pp and for DFF facilities being 49.3 pp. Similarly, the impact on the proportion of facilities that had a vehicle was more than twice the average at the baseline. The average impact for PBF and DFF facilities is not very different at 19.8 pp and 20.8 pp respectively. With an improved availability of communication and transport access at NSHIP facilities, we also see an improvement in the extent of outreach activities undertaken by these facilities. The impact on the proportion of facilities that reported conducting any outreach activities for ANC services is 22.2 pp, where the corresponding average at the baseline was 45%.

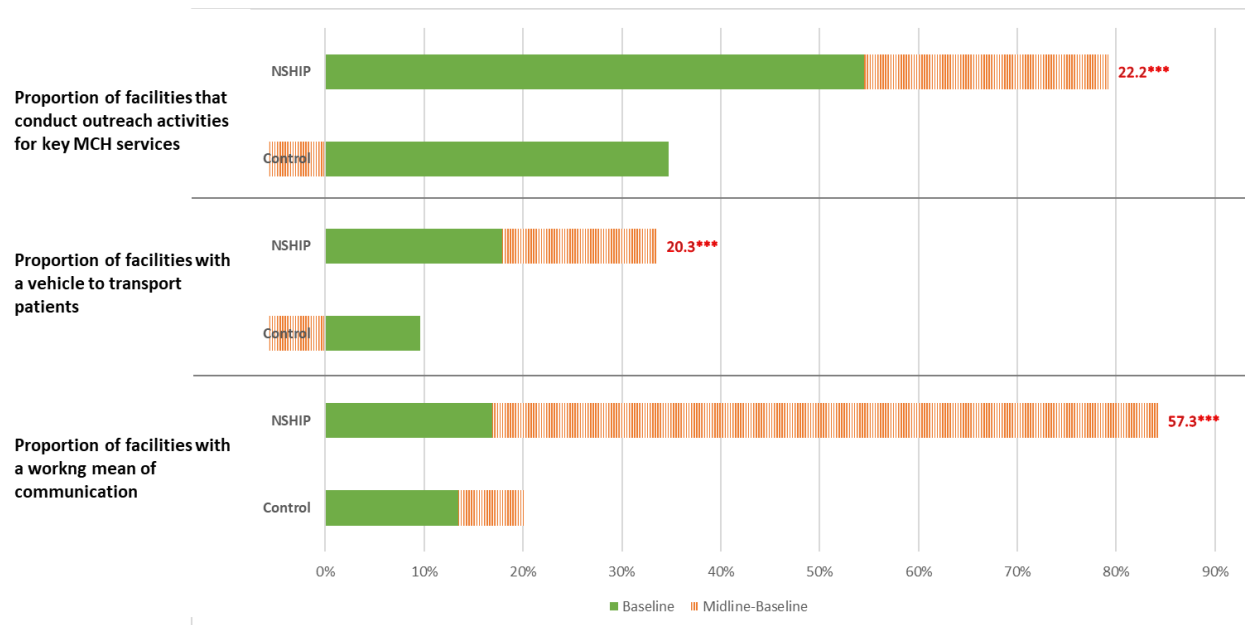
Figure 12: Impact on Procedural Quality of Healthcare, Health Worker



Note: Weighted double-difference estimates for comparisons between NSHIP and control states are in red, between PBF and DFF LGAs are in purple.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Figure 13: Impact on Procedural Quality of Healthcare, Infrastructure



Note: Weighted double-difference estimates for comparisons between NSHIP and control states are in red, between PBF and DFF LGAs are in purple.* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

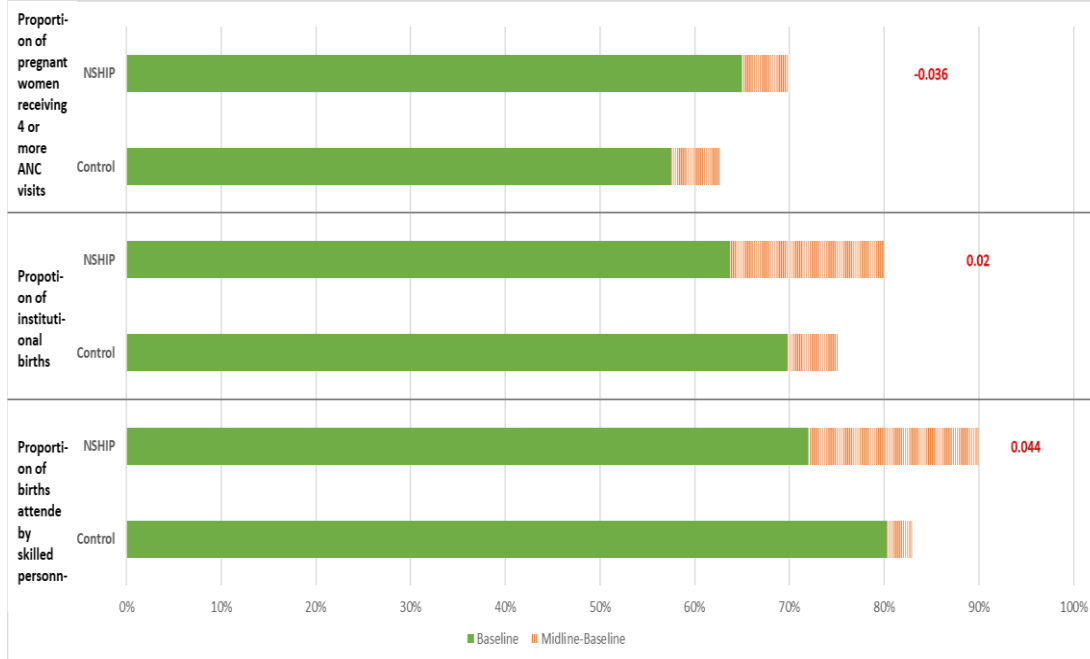
Table 19: Impact on Procedural Quality of Healthcare

	NSHIP vs Control	PBF vs Control	DFF vs Control	PBF vs DFF	Baseline Mean	N
Average health worker clinical knowledge score	0.039	0.024	0.053	-0.04	0.46	1506
	(1.00)	(0.56)	(1.27)	(1.22)		
Under-five examination quality score (based on IMCI protocols)	0.146***	0.113**	0.175***	-0.076	0.57	1374
	(3.21)	(2.16)	(3.24)	(1.34)		
ANC examination quality score (based on national ANC protocols)	0.082*	0.054	0.108**	-0.045	0.45	1285
	(1.93)	(1.16)	(2.24)	(1.05)		
Proportion of facilities with working means of communication (radio, mobile phone, landline)	0.573***	0.662***	0.493***	0.133*	0.14	1526
	(5.01)	(5.53)	(4.07)	(1.71)		
Proportion of facilities with a working vehicle to transport patients for referral	0.203***	0.198***	0.208***	-0.019	0.14	1526
	(4.16)	(3.61)	(3.20)	(0.27)		
Proportion of health facilities that conduct outreach for key MCH services	0.222***	0.252**	0.190**	0.072	0.45	1524
	(2.72)	(2.65)	(2.06)	(0.75)		
Standard errors clustered at LGA level						
t statistics in parentheses						
* p<0.1, ** p<0.05, *** p<0.01						

5.1.5 Outcomes

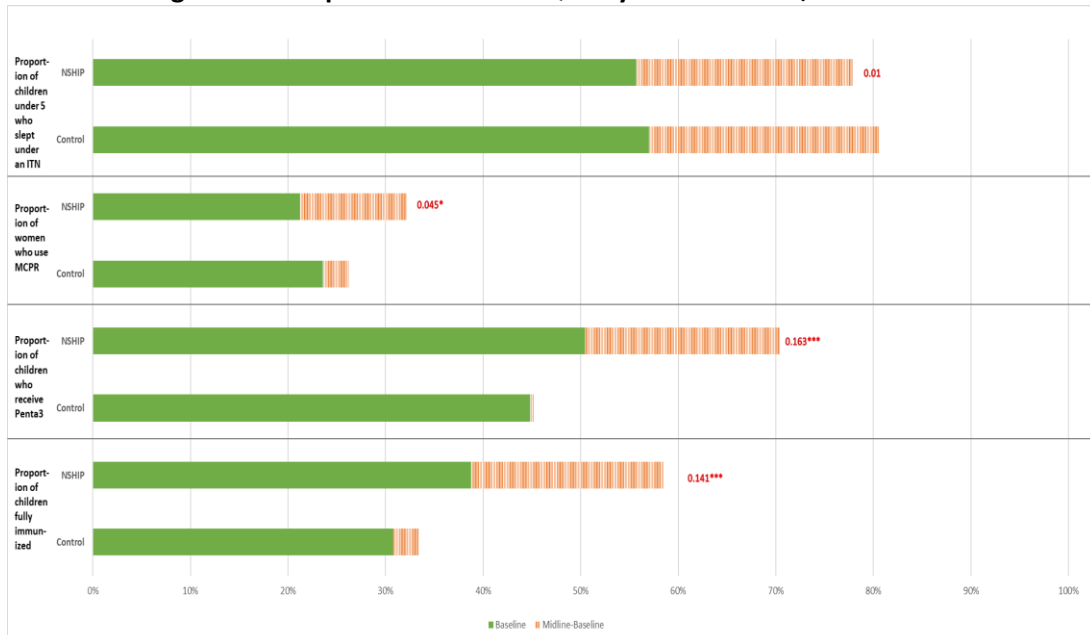
5.1.5.1 Health Facilities

Figure 14A: Impact on Outcome Quality of Healthcare, Maternal and Child Care



Note: Weighted double-difference estimates for comparisons between NSHIP and control states are in red, between PBF and DFF LGAs are in purple. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Figure 14B: Impact on Outcome Quality of Healthcare, Maternal and Child Care



Note:

Weighted double-difference estimates for comparisons between NSHIP and control states are in red, between PBF and DFF LGAs are in purple. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Patients' experiences at health centers affect their clinical outcomes, and thus the patient retention rate. Improving patient's experiences at the health facility is an important component in the strategy that emphasizes patient centric quality of care. To capture patients' experiences at health centers we looked at two indicators- client satisfaction and the convenient operational hours for a facility (Table 20). Average satisfaction for the clients who sought ANC services increased by about 5 pp in NSHIP states. This improvement was driven by improvements among health centers in DFF LGAs. An improvement in the average client satisfaction score among those who sought under five curative care was about 8.5% of the average at the baseline. These improvements were present in both the PBF and the DFF LGAs. ANC clients were 7.8 pp more likely to find the facility's opening hours to be convenient if they were visiting facilities in PBF LGAs, and 14.5 pp more likely if they were visiting facilities in DFF LGAs. It is noteworthy that the improvements among health facilities in PBF LGAs were significantly smaller than the improvements among health facilities in DFF LGAs. A similar pattern emerges when we look at the experience of clients who sought under-five curative care. The proportion of clients reporting that the hours of operation are convenient increased by 7 pp for PBF LGAs, and by 11.5 pp for DFF LGAs. Although the magnitude of improvement among DFF facilities was larger than the improvements among the PBF facilities, the difference was not statistically significant.

Table 20: Impact on Outcome Quality of Healthcare, Health Facility

	NSHIP vs Control	PBF vs Control	DFF vs Control	PBF vs DFF	Baseline Mean	N
ANC: Average client satisfaction score	0.045*	0.03	0.058*	-0.035	0.89	1374
	(1.69)	(0.99)	(1.98)	(1.25)		
U5: Average client satisfaction score	0.079**	0.077*	0.078**	-0.017	0.88	1284
	(2.33)	(1.99)	(2.16)	(0.55)		
ANC: Proportion of clients who report that facility opening hours are convenient	0.114***	0.078*	0.145***	-0.071**	0.91	1374
	(2.69)	(1.78)	(3.17)	(2.1)		
U5: Proportion of clients who report that facility opening hours are convenient	0.093***	0.070*	0.115***	-0.052	0.90	1284
	(2.68)	(1.75)	(2.93)	(1.34)		
Standard errors clustered at LGA level						
t statistics in parentheses						
* p<0.1, ** p<0.05, *** p<0.01						

5.1.5.2 Maternal and Child Care

We find sizeable increases in the rate of births attended by skilled birth personnel for households that were in PBF LGAs, but not in DFF LGAs. The impact on these households was 9.1 pp; the average at the baseline was 60%. It is worth noting that the median improvement in skilled birth attendance between 1980 and 2008 was only 0.8%, and the improvements among the top quartile of performers was also substantially smaller at 2.2% (Arur et al., 2011). This increase in skilled birth comes from an increase of 13% in institutional deliveries. The improvements in the availability of ANC services at NSHIP facilities did not lead to changes in population-level outcomes (Table 20).

There were significant improvements in the coverage of immunization facilities for these households. The impact of NSHIP on the proportion of children (12-23 months) who were fully immunized was 14.1 pp, which was about 50% of the average at the baseline. The impact was substantial for both PBF and DFF at 10.5 pp and 17.2 pp respectively. The proportion of children who receive Penta3 vaccines increased by 16.3 pp, about 42% of the average at the baseline, and was significantly higher due to DFF (20.6 pp) than due to PBF (11.1 pp) mode of financing. This indicates a substantial improvement in the immunization coverage as the median annual improvement in the DPT3 coverage rate across the globe between 1980 and 2008 was only 1% (Arur et al., 2011). Notably, curative care expenditure on children under five went up to the tune of approximately 300 naira for the households in LGAs with DFF facilities; a much smaller (and insignificant) increase of 157 naira is observed in PBF areas. We also found an increase on modern contraceptive usage of 5.7 pp for PBF areas. This was a 33.9% improvement over the average for baseline, where the median rate of increase in modern contraceptive usage was only 0.8% and 1.8% among the best 25% performers (Arur et al., 2011). The impact for DFF facilities was not statistically different from zero. Overall, the project did not affect insecticide treated net usage by young children in project states, but usage went up by 5.6 percent in DFF areas relative to PBF ones (Table 23).

Table 21: Impact on Outcome Quality of Healthcare, Maternal and Child Health Outcomes

	NSHIP vs Control	PBF vs Control	DFF vs Control	PBF vs DFF	Baseline Mean	N
Proportion of births attended by skilled personnel	0.044	0.091**	-0.006	0.110***	0.60	14,689
	(1.39)	(2.47)	(0.17)	(3.33)		
Proportion of institutional Delivery	0.02	0.067*	-0.03	0.101***	0.53	14,566
	(0.64)	(1.86)	(0.83)	(2.95)		

	NSHIP vs Control	PBF vs Control	DFF vs Control	PBF vs DFF	Baseline Mean	N
Proportion of pregnant women receiving 4 or more ANC visits	-0.036 (1.30)	-0.038 (1.10)	-0.035 (1.06)	-0.025 (0.69)	0.50	14,913
Proportion of children (12-23 months) fully immunized	0.141*** (4.70)	0.105*** (3.02)	0.172*** (4.95)	-0.038 (1.10)	0.29	7,437
Proportion of children (12-23 months) who receive Penta3	0.163*** (5.13)	0.111*** (3.01)	0.206*** (5.73)	-0.060* (1.69)	0.39	7,437
Curative care expenditure on children under-5 (in Naira)	126.791 (1.37)	105.788 (0.78)	150.824 (1.48)	-116.336 (0.96)	229	14,418
Proportion of women who use modern contraceptive methods	0.045* (1.89)	0.057** (2.15)	0.033 (1.24)	0.021 (1.00)	0.17	13,448
Proportion of children under-5 who slept under an ITN last night	0.01 (0.35)	-0.009 (0.30)	0.032 (1.07)	-0.056* (1.94)	0.45	22,441
% of children 6-23 months old who received BCG	0.094*** (2.95)	0.053 (1.52)	0.134*** (3.41)	-0.069* (1.81)	0.66	10,682
Standard errors clustered at LGA level						
t statistics in parentheses						
* p<0.1, ** p<0.05, *** p<0.01						

5.1.5.3 Public Healthcare Utilization

The effect on institutional delivery in households in PBF LGAs seems to be driven entirely by increased coverage by public health facilities (Table 22)- institutional delivery at a public health facility increased by 10.5 pp relative to a baseline average of 36.2%. Some of this effect appears to have come by displacing the demand of private facilities- in PBF areas, institutional deliveries at private health facilities fell by 3.9 percent. There were no changes in reported curative seeking behavior at a public health facility.

Table 22: Impact on Outcome Quality of Healthcare, Maternal and Child Health Outcomes

	NSHIP vs Control	PBF vs Control	DFF vs Control	PBF vs DFF	Baseline Mean	N
Proportion of institutional delivery at public health facilities	0.045	0.105***	-0.019	0.128***	0.36	14,566
	(1.58)	(3.23)	(0.54)	(3.71)		
Proportion of institutional delivery at a private health facility	-0.025	-0.039*	-0.011	-0.027	0.16	14,566
	(1.27)	(1.83)	(0.50)	(1.51)		
Proportion of HHs who visited a public health facility to seek curative care	-0.023	-0.001	-0.031	0.038	0.64	1,663
	(0.33)	(0.02)	(0.38)	(0.56)		
Standard errors clustered at LGA level						
t statistics in parentheses						
* p<0.1, ** p<0.05, *** p<0.01						

The IE CN notes that this sampling design was powered to detect a treatment effect of 10 pp with 97% certainty for facility-level outcomes and with 90% certainty for household-level characteristics. Indeed, the standard errors reported here, clustered at the LGA level, suggest that the study is reasonably well powered. For example, we detect a 7 pp increase in institutional deliveries in PBF versus control facilities with 90% confidence (95% confidence interval [-0.004, 0.14]) and a 10 pp increase in PBF versus DFF facilities with 95% confidence (confidence interval [0.03, 0.17]). On the other hand, we are unable to detect a 4.4 pp increase in skilled birth attendance in NSHIP versus control facilities (95% confidence interval [-0.02, 0.11]) or a 6.6 pp increase in Penta3 coverage in DFF versus PBF (95% confidence interval [0.10, 0.23]) at standard levels of confidence.

5.2 Corrections for Multiple Hypothesis Testing

We correct for multiple hypothesis using the average effects approach laid out in Kling and Liebman (2004). The indices examined follow the Donabedian model of quality: structural quality, process quality, and outcomes. We group structural measures of quality using the balanced scorecard that is used by NSHIP for payments. Results show that NSHIP lead to significant improvements in overall structural quality. Both PBF and DFF were associated with improvements in structural quality, although DFF improved structural quality to a greater degree. All results are significant at the 95 percent level of confidence. The indices for process quality are as presented in Table 19: the under-five examination quality score following IMCI protocol and the ANC quality score following national protocol; results show significant gains from NSHIP, but no differential impact of PBF. Indeed, if anything, DFF shows greater

gains for ANC. In addition, we group all maternal health indicators and child health indicators in Table 21 into one index for maternal health outcomes and one for child health outcomes. The p-values for the maternal health indicators are as follows: 0.00 for NSHIP versus control, 0.00 for PBF versus control, 0.055 DFF versus control, and 0.02 for PBF versus DFF. The corresponding p-values for child health indicators are: 0.079 for NSHIP versus control, 0.423 for PBF versus control, 0.065 DFF versus control, and 0.175 for PBF versus DFF. These corrected results suggest that while NSHIP improved structural and process quality, the DFF arm drove most of these gains. In terms of outcomes, PBF improved maternal health outcomes, but DFF increased child health outcomes.

6 Impact Heterogeneity by Key Household Characteristics

This section seeks to understand which socioeconomic groups responded most to the improvements in the quality and quantity of health service provision at facilities. Differential response of different groups could have consequences for the dispersion in the eventual health outcomes across these groups. A household's response to improvements at NSHIP facilities will depend on its range of healthcare options, including its ability to buy healthcare from both public and private facilities and whether it is located in a rural or an urban locality, and possibly on its religion (Ganle, 2015). We thus focus on three key aspects of a household's socioeconomic identity that may interact with their response to NSHIP: location, wealth and religion.

6.1 Wealth

Recall that the rates of skilled birth attendance and institutional delivery increased as a result of the PBF mode of financing, while the increase in DFF LGAs, and consequently the three project states was statistically insignificant. Table 25A shows that this increase in the rates of skilled birth attendance and institutional delivery was driven by higher utilization among those households that were located in the middle of the wealth distribution. The impact on skilled birth attendance was significant only for the third and the fourth quintile in the wealth distribution at 11.8 pp and 11 pp, respectively. The impact of PBF was also significantly higher than the impact of DFF for these wealth quintiles. Similarly, the impact of PBF on the rate of institutional delivery was evident only among the fourth wealth quintile (12.8 pp) and was positive but not significant for the third wealth quintile (6.9 pp and p-value of 14.7%). However, it is important to note that while the distribution of asset index in the baseline sample is similar to the distribution in DHS 2013 for the six states studied here, Table 5 of this report shows that the sample is

poorer than the national average. Therefore, the middle of the wealth distribution of this sample is poorer than the average Nigerian household.

Table 23A: Impact on Outcome Quality of Healthcare, Maternal and Child Health Outcomes, by Wealth

NSHIP					
	Q1	Q2	Q3	Q4	Q5
Proportion of births attended by skilled personnel	-0.026	-0.02	0.068	0.024	0.032
	(0.44)	(0.38)	(1.41)	(0.51)	(0.64)
Baseline Mean	0.58	0.58	0.59	0.61	0.60
Number of Observations	2,699	2,808	2,940	3,084	3,157
Proportion of institutional delivery	-0.055	-0.073	0.021	0.025	0.054
	(0.99)	(1.45)	(0.43)	(0.48)	(1.01)
Baseline Mean	0.50	0.52	0.53	0.54	0.55
Number of Observations	2,677	2,785	2,924	3,057	3,123
PBF					
	Q1	Q2	Q3	Q4	Q5
Proportion of births attended by skilled personnel	-0.025	0.045	0.118**	0.110**	0.061
	(0.36)	(0.73)	(2.12)	(2.07)	(1.15)
Proportion of institutional delivery	-0.038	-0.036	0.082	0.128**	0.069
	(0.61)	(0.60)	(1.45)	(2.22)	(1.24)
DFF					
	Q1	Q2	Q3	Q4	Q5
Proportion of births attended by skilled personnel	-0.012	-0.092	0.007	-0.065	0.002
	(0.19)	(1.49)	(0.13)	(1.20)	(0.03)
Proportion of institutional delivery	-0.062	-0.115*	-0.049	-0.083	0.038
	(0.95)	(1.80)	(0.84)	(1.42)	(0.62)
PBF vs DFF					
	Q1	Q2	Q3	Q4	Q5
Proportion of births attended by skilled personnel	0.007	0.111*	0.142** *	0.185** *	0.082*
	(0.12)	(1.83)	(2.78)	(3.98)	(1.87)
Proportion of institutional delivery	0.023	0.079	0.156** *	0.209** *	0.027
	(0.38)	(1.26)	(3.00)	(4.25)	(0.59)
Standard errors clustered at LGA level					
t statistics in parentheses					
* p<0.1, ** p<0.05, *** p<0.01					

While the impact on the utilization of ANC services was insignificant for the whole sample, examining it for different wealth quintile shows that while the impact of NSHIP on ANC visits among the lower wealth quintiles was negative for the first (-10.7 pp), second (-16.3 pp) and third (-9 pp) quintiles, it was positive but not significant for the highest wealth quintile (7.9 pp, p-value- 0.11). Similar patterns were observed when we separate the samples into PBF LGAs and DFF LGAs.

The increase in the utilization of immunization services was driven by households in the two richest wealth quintiles. The impact of NSHIP on the immunization rate of children in the fourth and fifth quantile were 18.3 pp and 22.8 pp, while that on Penta3 coverage were 13.6 pp and 25.6 pp, respectively. In DFF LGAs, the coverage of Penta3 increased among all the households along the wealth distribution, but the increase was the largest among those in the fifth wealth quintile at 27.5 pp.

Table 23B: Impact on Outcome Quality of Healthcare, Maternal and Child Health Outcomes, by Wealth

NSHIP					
	Q1	Q2	Q3	Q4	Q5
Proportion of pregnant women receiving 4 or more ANC visits	-0.107**	-	-0.090*	-0.053	0.079
	(2.08)	(3.55)	(1.75)	(1.11)	(1.62)
Baseline Mean	0.50	0.52	0.53	0.54	0.55
Number of Observations	2,743	2,852	2,996	3,127	3,195
Proportion of children (12-23 months) fully immunized	0.066	0.062	0.036	0.183**	0.228**
	(1.11)	(1.12)	(0.59)	*	*
Baseline Mean	0.27	0.28	0.27	0.30	0.32
Number of Observations	1,422	1,397	1,474	1,567	1,577
Proportion of children (12-23 months) who receive Penta3	0.134**	0.095	0.05	0.136*	0.256**
	(2.00)	(1.60)	(0.79)	(1.94)	*
Baseline Mean	0.38	0.38	0.38	0.40	0.40
Number of Observations	1,422	1,397	1,474	1,567	1,577
PBF					
	Q1	Q2	Q3	Q4	Q5
Proportion of pregnant women receiving 4 or more ANC visits	-0.152**	-	-0.082	-0.029	0.073
	(2.41)	(2.92)	(1.42)	(0.53)	(1.32)
Proportion of children (12-23 months) fully immunized	0.035	0.064	0.003	0.138*	0.181**
	(0.46)	(1.01)	(0.05)	(1.81)	(2.21)

Proportion of children (12-23 months) who receive Penta3	0.083	0.075	-0.021	0.099	0.207** *
	(0.98)	(1.11)	(0.29)	(1.28)	(2.61)
DFF					
	Q1	Q2	Q3	Q4	Q5
Proportion of pregnant women receiving 4 or more ANC visits	-0.058	- 0.17***	-0.097*	-0.079	0.086
	(0.98)	(2.97)	(1.67)	(1.44)	(1.59)
Proportion of children (12-23 months) fully immunized	0.105	0.074	0.08	0.204** *	0.249** *
	(1.49)	(1.16)	(1.10)	(2.71)	(2.96)
Proportion of children (12-23 months) who receive Penta3	0.199** *	0.123*	0.126*	0.148*	0.275** *
	(2.74)	(1.74)	(1.69)	(1.90)	(3.29)
PBF vs DFF					
	Q1	Q2	Q3	Q4	Q5
Proportion of pregnant women receiving 4 or more ANC visits	-0.097	-0.049	0.004	0.016	-0.015
	(1.61)	(0.81)	(0.07)	(0.33)	(0.31)
Proportion of children (12-23 months) fully immunized	-0.041	0	-0.034	-0.024	-0.064
	(0.54)	(0.00)	(0.52)	(0.40)	(1.00)
Proportion of children (12-23 months) who receive Penta3	-0.101	-0.035	-0.086	-0.027	-0.049
	(1.21)	(0.50)	(1.23)	(0.42)	(0.76)
Standard errors clustered at LGA level					
t statistics in parentheses					
* p<0.1, ** p<0.05, *** p<0.01					

The key takeaway is that the increases in the utilization of health services were driven by higher usage among households in the middle of the wealth distribution. These findings may suggest that the constraints in the health sector were not only situated in the limited capacity of the health facilities (that NSHIP seeks to improve), but also in the constraints on the financial capacity of the potential patients to avail of these health facilities.

6.2 Location

NSHIP increased institutional deliveries by 14.1 pp in urban households in PBF areas. Perhaps unsurprisingly, these gains stemmed from a 20 pp increase in births at public facilities in urban areas, concomitant with an 11 pp decline in deliveries at private facilities in these areas as well as insignificant

changes in rural areas. Such a shift may arise if households substitute away from private health centers in urban areas, but not in rural areas.

Table 24A: Impact on Outcome Quality of Healthcare, Maternal and Child Health Outcomes, by Location

	NSHIP vs Control	PBF vs Control	DFF vs Control	PBF Vs DFF	Baselin e Mean	N
Rural: Proportion of births attended by skilled personnel	0.003 (0.09)	0.055 (1.26)	-0.057 (1.28)	0.131** * (3.07)	0.63	10,424
Urban: Proportion of births attended by skilled personnel	0.079 (1.30)	0.141** (2.14)	0.028 (0.42)	0.105** (2.02)	0.63	4,265
Rural: Proportion of institutional delivery	-0.013 (0.35)	0.042 (0.98)	-0.077* (1.76)	0.121** * (2.77)	0.53	10,344
Urban: Proportion of institutional delivery	0.04 (0.64)	0.09 (1.34)	-0.001 (0.02)	0.083 (1.46)	0.51	4,222
Rural: Proportion of institutional delivery at public health facilities	0.01 (0.32)	0.062 (1.65)	-0.05 (1.19)	0.108** (2.44)	0.35	10,344
Urban: Proportion of institutional delivery at public health facilities	0.103* (1.68)	0.198** * (3.07)	0.023 (0.33)	0.155** * (2.73)	0.40	4,222
Rural: Proportion of institutional delivery at a private health facility	-0.023 (1.00)	-0.021 (0.85)	-0.027 (0.97)	0.013 (0.55)	0.18	10,344
Urban: Proportion of institutional delivery at a private health facility	-0.063 (1.58)	-0.109** (2.59)	-0.024 (0.58)	-0.07*** (2.65)	0.11	4,222
Standard errors clustered at LGA level						
t statistics in parentheses						
* p<0.1, ** p<0.05, *** p<0.01						

Surprisingly, not only the increase in the rate of ANC visits was driven by higher utilization among household in urban areas, the utilization of ANC services decreased in rural areas in both PBF (-11.5 pp) and DFF LGAs (-10.9 pp). In PBF LGAs, the rate of ANC visits in urban areas by 14.3 pp. The rate of ANC visits also increased DFF LGAs by 7.7 pp, but this increase was not statistically significant.

The improvements in immunization rates did not depend on the location of households- the improvements were substantial among households in rural as well as urban areas, and in PBF and DFF LGAs. NSHIP increased the proportion of children who were fully immunized by 12.6 pp in rural areas and 15.6 pp in urban areas. The increase in the utilization of Penta3 vaccines were limited to rural areas.

Proportion of children in rural households who received Penta3 vaccination increased by 17.3 pp; this increase was higher among rural households residing in DFF LGAs. Similarly, the improvements in the usage of modern contraceptive methods for households in the PBF LGAs were entirely being driven by the improvements among rural households. Finally, the impact of DFF on curative care expenditure for children under five was driven by the decreases for households in rural areas. No significant patterns were observed by location in out-of-pocket expenditures at public health facilities.

While these patterns do not suggest that all the observed impacts of NSHIP stemmed entirely from impacts in easier-to-reach urban communities, impacts on two vital outcomes, skilled birth attendance and institutional delivery, were indeed induced by the response by urban households who may be better able to respond to perceived improvements in quality.

Table 24B: Impact on Outcome Quality of Healthcare, Maternal and Child Health Outcomes, by Location

	NSHIP vs Control	PBF vs Control	DFF vs Control	PBF Vs DFF	Baseline Mean	N
Rural: Proportion of pregnant women receiving 4 or more ANC visits	-0.11***	-0.16***	-0.109***	0.000	0.48	10,576
	(3.49)	(2.97)	(2.73)	(0.00)		
Urban: Proportion of pregnant women receiving 4 or more ANC visits	0.11**	0.143**	0.077	0.056	0.50	4,337
	(2.09)	(2.33)	(1.36)	(0.94)		
Rural: Proportion of children (12-23 months) fully immunized	0.13***	0.088**	0.149***	-0.033	0.26	5,271
	(3.66)	(2.12)	(3.65)	(0.72)		
Urban: Proportion of children (12-23 months) fully immunized	0.156**	0.126*	0.183**	-0.027	0.34	2,166
	(2.26)	(1.76)	(2.46)	(0.55)		
Rural: Proportion of children (12-23 months) who receive Penta3	0.173***	0.106**	0.222***	-0.089*	0.36	5,271
	(4.78)	(2.38)	(5.41)	(1.92)		
Urban: Proportion of children (12-23 months) who receive Penta3	0.107	0.091	0.123	0.016	0.45	2,166
	(1.47)	(1.20)	(1.54)	(0.30)		
Rural: Curative care expenditure on children under-5 (in Naira)	176.006	137.677	225.244**	-195.26	249.9	10,179
	(1.54)	(0.79)	(1.99)	(1.29)		
Urban: Curative care expenditure on children under-5 (in Naira)	333.773	191.685	448.574	-281.715	177.6	4,239

	(1.29)	(1.00)	(1.07)	(0.74)		
Rural: Proportion of women who use modern contraceptive methods	0.060**	0.074**	0.044	0.025	0.18	3,920
	(2.18)	(2.41)	(1.39)	(0.93)		
Urban: Proportion of women who use modern contraceptive methods	-0.031	-0.02	-0.042	0.028	0.18	9,528
	(0.70)	(0.39)	(0.87)	(0.70)		
Standard errors clustered at LGA level						
t statistics in parentheses						
* p<0.1, ** p<0.05, *** p<0.01						

6.3 Religion

The increases observed in skilled birth attendance and institutional delivery in PBF area were driven by gains among non-Muslim households, especially through the gains in institutional delivery at public health centers. However, we also find a near-significant 9.4 pp increase in institutional deliveries at private facilities for Muslim households.

Table 25A: Impact on Outcome Quality of Healthcare, Maternal and Child Health Outcomes, by Religion

	NSHIP vs Control	PBF vs Control	DFF vs Control	PBF Vs DFF	Baseline Mean	N
Non-Muslim: Proportion of births attended by skilled personnel	0.04	0.116***	-0.029	0.137***	0.62	10,097
	(1.15)	(2.94)	(0.67)	(3.30)		
Muslim: Proportion of births attended by skilled personnel	0.053	0.052	0.052	0.045	0.52	4,592
	(0.97)	(0.82)	(0.91)	(0.94)		
Non-Muslim: Proportion of institutional Delivery	0.032	0.103**	-0.033	0.130***	0.57	10,010
	(0.91)	(2.56)	(0.75)	(3.08)		
Muslim: Proportion of institutional Delivery	-0.013	-0.008	-0.021	0.031	0.45	4,556
	(0.25)	(0.14)	(0.35)	(0.58)		
Non-Muslim: Proportion of institutional delivery at public health facilities	0.031	0.119***	-0.048	0.158***	0.35	10,010
	(0.94)	(3.31)	(1.18)	(3.78)		
Muslim: Proportion of institutional delivery at public health facilities	0.076	0.094	0.048	0.066	0.37	4,556
	(1.51)	(1.63)	(0.81)	(1.27)		

Standard errors clustered at LGA level
t statistics in parentheses
* p<0.1, ** p<0.05, *** p<0.01

Similarly, parsing the NSHIP impact on ANC care-seeking into Muslim and non-Muslim households shows a close-to-significant increase in the PBF arm on non-Muslim households. NSHIP was also coincident with large declines in ANC utilization by Muslim households, perhaps indicating the presence of secular trends faced by these households. The gains in modern contraception usage also stemmed from impacts among Non-Muslim households. In contrast, full immunization of children increased in both treatment arms among non-Muslim (PBF: 8.8 pp and DFF: 17.3 pp) and Muslim households (PBF: 15 pp and DFF: 18.1 pp). Similarly, the impact on the utilization of Penta3 vaccinations was positive for DFF among the Non-Muslim households (22.4 pp) and for both PBF (19.8 pp) and DFF among Muslim households (18.1 pp). These results highlight that the healthcare innovations designed to improve the coverage of maternal and child care for non-Muslim families in Nigeria might lack the flexibility required to respond to the unique needs of women in Muslim households.

Table 25B: Impact on Outcome Quality of Healthcare, Maternal and Child Health Outcomes, by Religion

	NSHIP vs Control	PBF vs Control	DFF vs Control	PBF vs DFF	Baseline Mean	N
Non-Muslim: Proportion of pregnant women receiving 4 or more ANC visits	0.03	0.061	-0.003	0.026	0.50	10,254
	(0.97)	(1.60)	(0.07)	(0.62)		
Muslim: Proportion of pregnant women receiving 4 or more ANC visits	-0.18***	-0.21***	-0.127**	-0.099**	0.49	4,659
	(3.23)	(3.60)	(2.05)	(2.05)		
Non-Muslim: Proportion of children (12-23 months) fully immunized	0.140***	0.088**	0.173***	-0.058	0.31	5,164
	(4.19)	(2.12)	(4.53)	(1.35)		
Muslim: Proportion of children (12-23 months) fully immunized	0.160***	0.150**	0.181***	-0.004	0.23	2,273
	(2.80)	(2.47)	(2.73)	(0.07)		
Non-Muslim: Proportion of children (12-23 months) who receive Penta3	0.159***	0.066	0.224***	-0.12***	0.42	5,164
	(4.38)	(1.54)	(5.51)	(2.78)		
Muslim: Proportion of children (12-23 months) who receive Penta3	0.188***	0.198***	0.181***	0.038	0.31	2,273
	(3.24)	(3.12)	(2.69)	(0.65)		
Non-Muslim: Proportion of women who use modern contraceptive methods	0.055**	0.081***	0.032	0.045*	0.20	9,354
	(2.04)	(2.63)	(1.07)	(1.71)		
Muslim: Proportion of women who use modern contraceptive methods	0.026	0.009	0.045	-0.036	0.11	4,103
	(0.62)	(0.20)	(0.99)	(1.22)		
Standard errors clustered at LGA level						
t statistics in parentheses						
* p<0.1, ** p<0.05, *** p<0.01						

7 Health Production

The underlying production of quality and quantity of care determines where the largest potential gains might be located, and has implications for the distributive characteristics of a health intervention like NSHIP. For instance, if the better performing facilities at the baseline are the ones that show maximal progress, the gaps in the quality and the quantity of care offered across facilities might be exacerbated. In this section, we use to score on the checklist used for the quarterly audits⁸ to measure the composite quality and quantity of care and explore the underlying production function using the following methodology. First, we calculate the improvements in the checklist score from the baseline to the midline. Next, we divide the facilities into two halves of better and worse performers, where performance is denoted by their improvement in the checklist score. Finally, we look at the average values of some key indicators that denote the basic human resource and management infrastructure at these facilities to ascertain how better performing facilities differed from the worse performing facilities before they were covered under NSHIP.

Results presented in Table 28 show that the facilities that performed the best during the period of the impact evaluation started at a lower baseline on a range of crucial human resources related outcomes: at baseline, average health worker satisfaction, the WHO wellbeing index, the proportion of health workers satisfied with their compensation schedule and who believed that their salary is fair were all lower among the best performing facilities. These facilities also had a lower management index and a lower proportion of full time employees at the facility. These findings indicate the underlying production function of health may yield diminishing returns to human resource inputs and the injections into poorly performing facilities might lead to the largest gains.

Table 26: Baseline Characteristics of HFs by their performance in improving health care provision

	Checklist Score: Bottom 50% Performers	Checklist Score: Top 50% Performers	Difference
PCA: Overall supervision	-0.200	-0.0650	-0.135
PCA: Internal supervision	-0.268	-0.164	-0.104
PCA: External supervision	0.173	0.168	0.00500
PCA: Health worker satisfaction	0.0370	-1.073	1.110***
PCA: Health worker motivation	-0.866	-1.095	0.229
PCA: Health worker wellness	0.488	0.152	0.337**

⁸ Although in this instance measured at baseline and endline as part of the IE surveys

	Checklist Score: Bottom 50% Performers	Checklist Score: Top 50% Performers	Difference
Proportion of workers satisfied with compensation	0.399	0.271	0.128***
Proportion of workers satisfied with benefits	0.187	0.117	0.069***
Proportion of workers receiving their salary in a timely manner in the last one year	0.640	0.489	0.151***
Proportion of workers who believe their salary is fair	0.566	0.490	0.077**
PCA: Community engagement	-0.318	-0.378	0.0600
Days absent in the past 30 days	2.316	1.979	0.337
Proportion of staff positions filled	0.454	0.483	-0.0300
Management index	21.57	15.35	6.219***
Proportion of full time employees at the facility	0.804	0.739	0.066**
* p<0.1, ** p<0.05, *** p<0.01			

In Table 30, we explore the profile of the patients who attended better performing and badly performing facilities. The top performing facilities were visited by relatively poor patients at the baseline, who were less likely to speak English and were likely to be able to read suggesting that the facilities that catered to poorer patients and were presumably in poorer areas showed the greatest improvements as a result of the investments made by NSHIP⁹.

Table 27: Baseline Characteristics of Patients by Facility Performance in on Quality Checklist

	Checklist Score: Bottom 50% Performers	Checklist Score: Top 50% Performers	Difference
Average wealth index of the patients at the facility	1.693	1.174	0.519***
Proportion of patients at the facility who speak English	0.216	0.131	0.085***
Proportion of patients at the facility who speak Yoruba	0.371	0.210	0.161***
Proportion of patients at the facility who speak Hausa	0.409	0.647	-0.238***
Proportion of patients at the facility who can read	0.611	0.474	0.137***
* p<0.1, ** p<0.05, *** p<0.01			

⁹ We also repeat this exploratory exercise for levels of performance on quantity (number of deliveries in the month preceding the survey) and quality (ANC quality of care index) and find similar patterns with better performing facilities having lower baseline values of human resource inputs, and catering to poorer patients.

8 Potential Mechanisms

To extend the lessons from this evaluation to other contexts, it is crucial to understand which component of NSHIP was most effective in initiating the cycle of change, and thus improving health outcomes. NSHIP's design is based on a multipronged approach that aims to target multiple interrelated outcomes. There are five important components in NSHIP: (1) additional financing to the states, LGAs and health centers, (2) decentralization of fund disbursement, (3) strengthening institutional performance by providing better expenditure tracking systems, supervision and monitoring, (4) regular monitoring of the facility activities through quarterly audits, and (5) incentives to facilities and health workers. While the design of this evaluation does not allow us to explore how relevant each of these components were in producing large gains witnessed with NSHIP, we explore three mutually non-exclusive mechanisms.

8.1 Incentives and Monitoring

There are typically two key components of an PBF program: an increase in resources available to the health centers, and an increase in incentives and monitoring for priority activities. Indeed, the PBF mode of financing under NSHIP follows this model. It provides cash injection through incentive-based payments along with quarterly audits. In contrast, the additional funds provided to the DFF facilities are not contingent on performance. DFF funds cannot be used for worker bonuses, shutting off the incentive channel for DFF facilities. However, the fixed payment that DFF facilities receive, to finance the same package of services, is based on the same contracting, supervision, and funds flow modality. Thus, a comparison of the performance of PBF and DFF facilities tests the efficacy of additional incentives to health facilities and health workers.

Recall that there was no difference in the improvements in the quantity of care- access and utilization- between PBF and DFF facilities. Among the indicators for structural quality of care, PBF facilities showed greater improvements in the availability of drugs and equipment, but improvements in recordkeeping were higher for DFF facilities. The story is even more mixed when we look at quality of care outcomes. Average client satisfaction increased more for DFF facilities, although the increase is not statistically significant. Clients at DFF facilities were more likely to report that the clinic hours are convenient. Impact on the rates of institutional delivery and skilled birth attendance was higher for PBF LGAs, but impact on outcomes related to immunization coverage was higher in DFF LGAs. The comparison between PBF and DFF facilities, thus, does not lead to an unambiguous verdict on the relative success of PBF and DFF in

generating improvements in quality and quantity of care. These results do suggest that incentives and monitoring provided under the PBF played a limited role.

In addition to looking at these indicators of quantity and quality of care, we explore the results on health care workers' experiences. Since PBF and DFF were randomly assigned within project states, we calculate intent-to-treat estimates to compare health care workers' responses in the endline interview. Their gender, age and experience at current facility are included as additional controls, and robust standard errors clustered at LGA level are calculated.

The first set of indicators relate to health workers' innovativeness in case of an adverse situation related to post-natal care, transportation, health workers and delivery. Indexes capturing the innovativeness of health workers' responses to any of these situations do not differ across PBF and DFF (Table 30).

Table 28: Innovativeness

	Innovation: PNC	Innovation: Transport	Innovation: Health Workers	Innovation: Delivery
PBFvDFF	-0.004	0.233	-0.064	0.109
	(0.01)	(0.88)	(0.28)	(0.41)
DFF Mean	0.04	-0.04	0.16	0.05
N	1,695	1,695	1,695	1,695
Standard errors clustered at LGA level				
t statistics in parentheses				
* p<0.1, ** p<0.05, *** p<0.01				

PBF did not have an additional effect over DFF for other aspects of a worker's performance- number of hours worked, and the number of patients attended to. Workers at PBF facilities did not experience a higher increase in their work hours over the past one year when compared to the workers at DFF facilities. This is not because the PBF did not work as it designed. More workers at PBF facilities experienced salary increases, and more workers' salaries increased because of their individual performance.

Table 29: Health Worker Performance

	Hours worked in a week	Proportion of health workers who work hours increased in the last 12 months	Number of patients attended the day before the survey	Proportion of health workers whose salary increased	Proportion of health workers whose salary increased due to their individual performance
PBFvDFF	1.406 (0.41)	0.054 (0.85)	0.171 (0.21)	0.075** (2.21)	0.045** (2.59)
DFF Mean	47.6	0.56	9.5	0.29	0.03
N	1,691	1,682	1,694	1,695	1,695
Standard errors clustered at LGA level					
t statistics in parentheses					
* p<0.1, ** p<0.05, *** p<0.01					

Despite increasing the salaries of health workers, PBF did not have an additional impact on health care outcomes, or on indicators that relate to productivity of health workers. Perhaps these changes take time, but PBF could improve the motivation of the workforce. Indeed, their score on WHO wellbeing index was significantly higher. The increase in their score on indexes capturing satisfaction and motivation was not statistically significant. Workers at PBF facilities were also not more likely to be satisfied with their salary, or feel that it is fair (Table 32).

Table 30: Health Worker Experiences

	Satisfaction	Motivation	Wellness	Salary is fair	Satisfied with salary	Timely salary	Received additional benefits
PBFvDFF	0.376 (1.23)	0.510 (0.92)	0.476* (1.87)	-0.017 (0.12)	0.022 (0.88)	0.005 (0.11)	0.142*** (3.81)
DFF Mean	0.25	-0.09	-0.18	0.46	0.16	0.31	0.04
N	1,600	1,341	1,694	1,694	1,694	1,676	1,694
Standard errors clustered at LGA level							
t statistics in parentheses							
* p<0.1, ** p<0.05, *** p<0.01							

There is a group of workers at PBF facilities who were more likely to be satisfied with their emoluments than the workers at DFF facilities – full time employees. They were more likely to think that their salary is fair and feel satisfied with their salary. They were more likely to receive additional emoluments. However,

the differential impact on indicators of their productivity – innovativeness, work hours and number of patients attended – is not significantly different from zero (Table 33).

Table 31: Health Worker Experiences for Full-time Employees

	Salary is fair	Satisfied with salary	Timely salary	Received additional benefits
PBFvDFF	-0.086	-0.029	-0.002	0.095***
	(1.37)	(1.18)	(0.03)	(2.84)
PBFvDFF*Full time employee	0.106**	0.069***	0.008	0.065
	(2.26)	(3.01)	(0.13)	(1.63)
DFF Mean	0.46	0.16	0.31	0.04
N	1,694	1,694	1,676	1,694
Standard errors clustered at LGA level				
t statistics in parentheses				
* p<0.1, ** p<0.05, *** p<0.01				

Were the fulltime employees at these facilities different from their temporary counterparts? Indeed, they were. They were more likely to be older men, less likely to hold another job and importantly were more likely to be aware of NSHIP. This is crucial as only 47% of the interviewed staff knew that a program such as NSHIP existed. Awareness about NSHIP varies considerably across PBF and DFF facilities- while only 44.8% of the staff knew about NSHIP in DFF facilities, 82.1% of the staff at PBF facilities was aware of NSHIP.

Table 32: Differences Across Full-time and Temporary Employees

	Temporary Staff	Fulltime Staff	Difference
Female	0.717	0.676	0.040***
	(0.027)	(0.028)	
Age	39.213	40.494	-1.281***
	(0.436)	(0.377)	
Years of experience at this facility	4.831	4.487	0.344
	(0.286)	(0.209)	
Held another job	0.193	0.138	0.055**
	(0.016)	(0.018)	
Salary (in Naira)	63482.76	102000	-38700
	(4858.816)	(25504.578)	
Knows about NSHIP	0.444	0.508	-0.064*

	Temporary Staff	Fulltime Staff	Difference
	(0.070)	(0.057)	
Standard errors clustered at LGA level			
t statistics in parentheses			
* p<0.1, ** p<0.05, *** p<0.01			

Taking a hint from the variation in the awareness about NSHIP across fulltime and temporary employees, we explore if there is an additional effect of PBF when compared to DFF on employees who were aware of NSHIP. Those workers at PBF facilities who knew about NSHIP, did attend to significantly higher number of patients on the day before the survey as compared to their counterparts in DFF facilities. The numbers of hours worked were also higher, but not significant with a p-value of 14%.

Table 33: Time Use by Health Workers

	Hours worked in a week	Proportion of health workers who work hours increased in the last 12 months	Number of patients attended the day before the survey
PBFvDFF	-2.659	0.137*	-2.207**
	(0.73)	(1.96)	(2.13)
PBFvDFFXaware	4.987	-0.102	2.919***
	(1.49)	(1.32)	(3.06)
DFF Mean	47.6	0.56	9.5
N	1,691	1,682	1,694
Standard errors clustered at LGA level			
t statistics in parentheses			
* p<0.1, ** p<0.05, *** p<0.01			

The additional effect of PBF on aware workers' satisfaction and motivation was also substantially and significantly higher. Moreover, fulltime employees at PBF facilities were more likely to believe that their salary was fair, satisfactory and that they receive it on time. They were also more likely to have received additional benefits.

Table 34: Health Worker Experiences for Employees Aware of NSHIP

	Satisfaction	Motivation	Wellness	Salary is fair	Satisfied with salary	Timely salary	Received additional benefits
PBFvDFF	-0.41 (0.65)	-1.417*** (3.49)	0.525* (1.99)	-0.169** (2.36)	-0.076** (2.59)	-0.105* (1.84)	-0.02 (0.67)
PBFvDFFXaware	1.109*** (2.94)	2.211*** (5.69)	-0.06 (0.31)	0.198*** (3.18)	0.120*** (4.17)	0.134** (2.38)	0.200*** (5.82)
DFF Mean	0.25	-0.09	-0.18	0.46	0.16	0.31	0.04
N	1,600	1,341	1,694	1,694	1,694	1,676	1,694
Standard errors clustered at LGA level							
t statistics in parentheses							
* p<0.1, ** p<0.05, *** p<0.01							

8.1 More Money

To understand if the impacts of NSHIP are at least partly routed through a simple injection of funds into the health care system, we look at the correlation between key facility-level outputs at the endline and the amount of funds received by these facilities as assessed by interviews with health facility staff at endline. While we don't expect these reported values of funds received under NSHIP to be accurate to the dollar, the correlations with facility level outputs are still suggestive that the hypothesized mechanism cannot be discarded. The correlations between the NSHIP payment and the quantity of health services (number of deliveries, ANC visits and immunizations) provided by these facilities is significant in several cases. On the other hand, correlations between NSHIP payments and clinical quality clear are less clear.

	PBF			DFF		
	Correlation Coefficient	Lower Limit	Upper Limit	Correlation Coefficient	Lower Limit	Upper Limit
Number of facility deliveries or deliveries in communities attended by skilled personnel in the 30 days preceding the survey	0.130*	-0.002	0.258	0.081	-0.037	0.196
Number of ANC visits in the 30 days preceding the survey	0.265***	0.139	0.382	0.055	-0.063	0.171
Number of immunizations offered in the 30 days preceding the survey	0.019	-0.114	0.151	0.007	-0.111	0.124
Proportion of clients (ANC) who report that facility opening hours are convenient	-0.039	-0.183	0.107	-0.012	-0.136	0.113
Proportion of clients (under-5 care) who report that facility opening hours are convenient	0.033	-0.112	0.177	-0.019	-0.156	0.118
Average health worker clinical knowledge score	-0.103	-0.233	0.030	0.012	-0.107	0.131
Average client (ANC) satisfaction score	-0.152***	-0.292	-0.007	0.039	-0.085	0.163
Average client (under-5 care) satisfaction score	-0.024	-0.169	0.121	-0.061	-0.196	0.076
ANC examination quality score (based on national ANC protocols)	-0.064	-0.208	0.082	0.096	-0.028	0.218
Under-five examination quality score (based on IMCI protocols)	-0.037	-0.181	0.109	0.039	-0.099	0.174

8.2 Better Expenditure Tracking

Providing additional resources through NSHIP may have also provided an exogenous shock to the bureaucracy. Consider the case of a health center in rural Nigeria that is resource strapped and has not received a lot of support from higher levels of administration. Now this center receives additional funds from the state through NSHIP, which in turn “unclogs” the expenditure flow systems that deliver resources in both the PBF and DFF arms. To explore this mechanism, we look at the correlation between the timeliness of NSHIP payments and facility outputs. It is worth mentioning that there was some variation in the timeliness of payments received by both PBF and DFF facilities.

Table 36: Timeliness of Payment for NSHIP Employees (in %)

	DFF	PBF	NSHIP
All of the time	26.42	20.32	23.73
Most of the time	33.65	40.64	36.73
Some of the time	28.62	30.28	29.35
Rarely	4.09	3.98	4.04
Never	7.23	4.78	6.15

As compared to the facilities that always received NSHIP payments on time, the average quantity of services provided by facilities (ANC visits and immunization) that often received NSHIP payments with delay were lower (Table 36). As with the NSHIP amounts, this pattern of correlation between the timeliness of NSHIP payments and quality of care indicators are not very clear. One concern with this exploratory analysis may be that the facilities that receive NSHIP funds in a timely fashion are likely to be facilities that receive larger amounts of money. We check this correlation and find that the timeliness of receiving NSHIP funds is not correlated with the amounts they receive.

Table 37: Correlation of timeliness of NSHIP payments and HF Outcomes (difference from the average for "Always Paid on Time")

	PBF				DFF			
	Most of the time	Some of the time	Rarely	Never	Most of the time	Some of the time	Rarely	Never
Number of facility deliveries or deliveries in communities attended by skilled personnel in the 30 days preceding the survey	-4.876	-6.205	-8.464	-7.607	-0.662	3.113	-1.692	38.808
	(0.97)	(1.01)	(1.37)	(1.04)	(0.23)	(1.01)	(0.53)	(1.17)
Number of ANC visits in the 30 days preceding the survey	-25.368	-30.825	-40.196	-85.096***	-10.484	8.415	-23.82**	-7.359
	(1.51)	(1.59)	(1.45)	(4.14)	(0.94)	(0.65)	(2.22)	(0.35)
Number of immunizations offered in the 30 days preceding the survey	-39.815	-121.64**	14.575	-246.58***	-56.978	175.739	-93.910*	-72.857
	(0.89)	(2.41)	(0.09)	(4.96)	(1.34)	(0.88)	(1.93)	(0.87)
Proportion of clients (ANC) who report that facility opening hours are convenient	-0.004	-0.013	0.013	-0.02	-0.02	-0.031	0.021	-0.070*
	(0.21)	(0.41)	(0.35)	(0.26)	(0.93)	(1.04)	(1.18)	(1.81)
Proportion of clients (under-5 care) who report that facility opening hours are convenient	-0.055*	-0.026*	0.018	-0.107	0.006	-0.023	0.035	-0.016
	(1.82)	(1.78)	(1.33)	(1.12)	(0.26)	(0.71)	(1.45)	(0.40)
Average client (ANC) satisfaction score	0.013	0.024	0.00	0.067**	-0.003	0.008	0.098	0.091**
	(1.01)	(0.88)	(0.00)	(2.56)	(0.11)	(0.30)	(1.35)	(2.57)
Average client (under-5 care) satisfaction score	-0.028	-0.026	-0.018	-0.016	-0.005	-0.01	0.002	-0.049
	(1.60)	(1.39)	(0.53)	(0.36)	(0.26)	(0.47)	(0.09)	(1.41)
Average health worker clinical knowledge score	-0.029	-0.006	-0.05	-0.063	-0.013	-0.025	-0.01	-0.031

	PBF				DFF			
	(1.44)	(0.40)	(1.67)	(0.86)	(0.64)	(1.29)	(0.34)	(1.02)
ANC examination quality score (based on national ANC protocols)	0.022	-0.015	0.100**	0.039	0.043	0.041	0.013	-0.022
	(0.71)	(0.29)	(2.12)	(0.42)	(0.60)	(0.60)	(0.17)	(0.28)
Under-five examination quality score (based on IMCI protocols)	-0.008	-0.032	-0.037	0.002	0.06	0.069	0.099*	0.005
	(0.36)	(1.03)	(1.64)	(0.05)	(1.64)	(1.70)	(1.99)	(0.08)

Next, we examine the correlation between the improvement in the timeliness of salary payments of the health facility staff and facility level outcomes as more timely salaries may have led to better motivated staff who performed at higher levels of effort. While the timeliness of NSHIP payments was correlated with increases in the quantity of services provided, greater improvements in the timeliness of worker salary payments are correlated with better quality of antenatal and under five curative care (Table 38).

Table 38: Correlation of timeliness of HW salary payments and HF Outcomes (difference from the average for "Always Paid on Time")

	PBF			DFF		
	Correlation Coefficient	Lower Limit	Upper Limit	Correlation Coefficient	Lower Limit	Upper Limit
Number of facility deliveries or deliveries in communities attended by skilled personnel in the 30 days preceding the survey	0.051	-0.075	0.175	-0.017	-0.134	0.100
Number of ANC visits in the 30 days preceding the survey	0.019	-0.103	0.141	-0.002	-0.117	0.113
Number of immunizations offered in the 30 days preceding the survey	-0.019	-0.140	0.103	-0.079	-0.192	0.037
Proportion of clients (ANC) who report that facility opening hours are convenient	-0.091	-0.213	0.033	0.092	-0.024	0.206
Proportion of clients (under-5 care) who report that facility opening hours are convenient	-0.074	-0.196	0.051	0.010	-0.106	0.125

		PBF			DFF	
Average client (ANC) satisfaction score	-0.024	-0.145	0.098	0.165*	0.051	0.275
Average client (under-5 care) satisfaction score	-0.064	-0.186	0.061	0.105	-0.011	0.218
Average health worker clinical knowledge score	-0.049	-0.172	0.076	0.094	-0.022	0.208
ANC examination quality score (based on national ANC protocols)	0.247**	0.127	0.360	0.132**	0.016	0.244
Under-five examination quality score (based on IMCI protocols)	0.185**	0.063	0.302	0.067	-0.049	0.182

Finally, it may also be the case that the three states responded differently to NSHIP. While, we don't have data explicitly on this aspect of NSHIP, we can compare the intra-cluster correlation (ICC) for two key process related variables in project states: timeliness of NSHIP payments and timeliness of worker salary payments. ICCs measure the "relatedness" of a variable within a cluster, states for our purpose. A higher ICC would indicate that more of the variation in the timeliness comes from within a state rather than across states, which would in turn be indicative of a greater role of a state's administrative capacity. The ICC of timeliness of NSHIP payments within a state for PBF facilities was 3.8% and for DFF facilities was 13.9%, while the ICC of the timeliness of health worker payments within a state for PBF was 15.1% and for DFF facilities was 10.5%. These patterns on timeliness of NSHIP and health worker payments suggest that while the state administrative capacity did not play a significant role in improving the timeliness of NSHIP payments to PBF facilities, it was still important in ensuring that workers are paid. On the other hand, state administrative capacity was an important correlate of timeliness of payments to both the facility and the workers in DFF LGAs.

Further, LGAs would have only one kind of disbursement- PBF or DFF. It might be instructive to explore the ICC for timeliness of NSHIP payments and timeliness of worker salary payments within LGAs. The ICC of timeliness of NSHIP payments within an LGA for PBF facilities was 17.7% and for DFF facilities was 36.7%, while the ICC of the timeliness of health worker payments within an LGA for PBF was 31.8% and for DFF facilities was 30.4%. This indicates that the LGA administration played a significant role to play in improving the timeliness of payments, and perhaps "unclogging" the expenditure tracking systems under DFF.

9 Key Results from the Fiscal Audit

A financial review of NSHIP was a part of the mid-term review (MTR) of NSHIP. This financial review was designed to be able to (a) determine how much the health facilities and LGA PHC departments earned during the period January 2015 to June 2017 and how the funds received were utilized, (b) utilization of DLI funds at the SPHCDA and LGA, (c) the presence of phantom health facilities and the duplication of health facilities, and (d) verify the existence and adequacy of financial records at the SPHCDA, LGA PHC Departments and selected health facilities.

Ernst and Young was contracted by the World Bank in Nigeria to conduct a review of the flow of funds for the NSHIP with a focus on the PBF mode of financing at the health facilities and LGA Primary Health Care (LGA PHC) Departments, DLI at state and LGA levels and DFF mode of financing at the health facilities and LGA PHC departments in the chosen project and comparison states. The scope of work included review of the following at the SPHCDA and at a selected sample of health facilities:

- 1) Receipt and transfer of funds at SPHCDA;
- 2) Timeliness of funds transfer;
- 3) Receipt of funds at the LGA and health facilities;
- 4) Utilization of funds at the SPHCDA, LGA PHCs and health facilities;
- 5) Presence of phantom facilities and the duplication of facilities;
- 6) Existence and accuracy of financial records; and
- 7) Compliance with financial agreements

Key results from a comparison of project and control states shows that a greater availability of funds in the project states helped ensuring that health facilities provide basic amenities and equipments. This is evident from the renovations at the health facilities and the LGA PHC departments using NSHIP funds. This improved the efficacy of health care delivery in project states when compared to the control states. Health facilities and LGA PHC departments in control states did not have access to sufficient funds, and relied only on internally generated revenue from service charge levied on patients. Infrastructure and equipments were in a delapidated state, and the health facilities did not offer a conducive environment for delivering high quality care. Moreover, health facilities in the project states maintained a stock of drugs and supplies to meet their patients' needs. These were dispensed at affordable rates and were readily available to community. To the contrary, in the comparison states, drugs were supplied centrally

and were always not available, and when available were dispensed at unaffordable prices.

In order to ensure transparency and accountability, every health facility and LGA PHC department in project states is expected to maintain financial records of their transactions. Even though the health facilities and LGA departments had a standard cash book, these were not adequately maintained. In comparison states, however, health facilities and the LGA PHC departments were not mandated to keep financial records. Consequently record keeping was poor at these levels and records were maintained at the LGA.

An important component of NSHIP's design was the *guiding* document (Project Implementation Manual) for implementing the project seamlessly. Having a set of clear guidelines was clearly important in ensuring that all the facilities and LGA PHC departments had improved financial management. Since such a manual is not available in the comparison states, the health facility managers are solely responsible to deciding how internally generated revenues are spent.

Examining the flow of funds in the facilities in project states more carefully, right amounts of funds based on the invoices received and approved (with no objections raised by the World Bank) into the designated accounts of all the LGA PHC departments and health facilities across all the three states by the State Project Finance Management Units (SPMFUs). However, funds transferred to most of the health facilities and LGA PHC departments took an average of 46 to 56 days as against the 45 days recommended by the PBF manual. Adamawa took an average of 48 -53 days, Nasarawa took an average of 46 to 56 days while Ondo took an average of 48 to 50 days. There was a significant delay at Ondo for 111 days in 2016 due to sanctions imposed by World Bank for issues of non performance and irregularities. The operations and maintenance of the health facilities depends on availability of this funds, and any delay in funding would make it difficult for the health facilities to meet up with their daily financial obligations like purchase of fuel for generators, purchase of consumables and other running costs.

The LGA PHC departments and health facilities had a standard cash book even they are not adequately maintained across the three states. Some facilities use exercise books instead of the standard cash book ledgers while some facilities in Ondo state use pencil to write in their cash books. Other financial records such as fixed asset registers, bank reconciliation statements and inventory registers were not standardised and were either not in existence or were not adequately maintained where they existed. Some of the assets procured with NSHIP funds were not recorded. Performance bonuses were computed manually at all

the PBF health facilities across all the three states. While they were computed using excel at the PBF LGA PHC departments. At the three states, the SPHCDA maintained adequate financial records of cash books, asset registers and bank reconciliation statements.

Generally, in all the facilities visited across all the three states, there were weaknesses in capability to prepare financial records. All the primary health care facilities and LGA PHC departments did not have officers with the capacity to prepare financial records as most of them do not have the required financial record keeping trainings. Even though secondary health facilities had designated account officers, these officers lacked adequate capacity to prepare a standard financial record. At the three states, the SPHCDA had accountants who were capable of preparing the financial records.

Funds were being used for operational expenses like contract staff salary, performance bonus payments at PBF facilities, purchase of drugs, maintenance of the facilities, etc. However, in all the three states, cheques were being raised in favour of officers at the health facilities and LGA PHC departments who encashed the cheque and used the fund to pay vendors and make other settlements. Cash payments were preferred over online and cheque transfers. This can pave way to missappropriation of funds as most of the expenditures were neither receipted nor captured adequately in the financial records.

There were cases of contraventions of the financial agreements: non compliance with procurement process, two instances of drugs purchased from unapproved vendors, unapproved business plans by the SPHCDA among other things. These contraventions were a result of inadequate understanding of the NSHIP PIM, and the PBF manual on part of the health facility managers and LGA PHC department secretaries/coordinators.

10 Cost Effectiveness Analysis

10.1 Methods

Incremental cost assessment

To provide practical recommendations for decision making by key stakeholders, this cost-effectiveness analysis (CEA) was conducted from a health system perspective and examined financial costs rather than economic costs. In this CEA, we focused on the incremental cost (additional cost) incurred in PBF and DFF groups, compared to control group. Thus, for the cost analysis, the additional cost that we included in the analysis were: (1) PBF implementation costs in the Nigeria, and (2) the World Bank headquarters' cost for designing, implementing and monitoring the PBF program¹⁰.

Program costs, primarily for administration of the PBF project (e.g. costs of operations, capacity building, verification, and monitoring and evaluation) and incentive payments, were obtained from the World Bank Nigeria office. The World Bank headquarters' costs were obtained from the World Bank headquarters and were allocated to the PBF and DFF groups in proportion to the program costs. Given the difference in population size among the three groups, the program costs and World Bank headquarters' costs were rescaled by population size and calculated as costs per capita.

All costs were measured in US 2015 dollars, and a discount rate of 3% was applied.

Incremental effectiveness assessment

Estimate improvement of utilization of health services

Results from the impact evaluation (described in Section 5) were used to estimate the improvements in health service utilization. More specifically, impacts on coverage of services for antenatal care, iron supplementation, postnatal care, skill birth attendance, immunization, modern contraceptive use, and children slept under insecticide-treated bed nets were used to compute the effectiveness of the intervention. Data on curative services and HIV/AIDS services, though available at the health facilities level, were not disaggregated in the way that could be converted to coverage at the household level for

¹⁰ Costs of consumables (drugs and supplies) due to increased services were likely completely paid through PBF or DFF. We excluded them from the analysis to avoid double counting.

LiST to model their impact. These two services were therefore excluded from the analysis, but all other incentivized services were included.

Estimate improvement of quality of health services

For computing effectiveness of NSHIP, impact evaluation results on following services were used: immunization, family planning, skill birth attendance, and ANC, and curative care for under-5. The health facility survey had a wide range of questions to assess the quality of care for difference maternal and health services. For each specific service, a quality score was generated, and rescaled to between 0-1. Quality of care may not have a linear relationship with health benefit gained from the care (e.g. 80% quality of care does not necessarily mean the care will gain 80% of its full potential of health benefits). To ascertain the impact of quality of care on potential health benefits from the care, we convened an expert panel in September 2018 to generate a health-effect index of quality of care using a quadratic function. The quadratic function was used because of its flexibility to accommodate concave up, concave down, and linear relationships.

Modelling outcomes by combining utilization and quality improvement

An effective coverage was generated by multiplying the health-effect index by the coverage of corresponding services. The result was treated as quality-adjusted coverage to feed into the Lives Saved Tool (LiST), which converts the coverage of health services to the number of lives saved (Avenir Health, 2015; Stenberg, 2014; Boschi-Pinto et al., 2010; Singh et al., 2014). LiST was developed through a joint consortium and is widely used to estimate MCH outcomes (mortality) with good validity (Keats et al., 2014). However, it only handles a limited number of interventions. It cannot deal with morbidity, nor can it implement probabilistic sensitivity analyses. We have included a sensitivity analysis on the link between quality of care and health outcomes in LiST as there are some assumptions made between utilization and mortality that merit some additional scrutiny. The default values in LiST for percentage of skilled delivery that has BEmOC (basic emergency obstetric and newborn care) and CEmOC (comprehensive emergency obstetric and newborn care) in Nigeria are estimated at 15% and 60%, respectively, of the coverage of skilled birth attendance. Since this level of quality of care for delivery might be an overestimate for the country, we decreased it by 50% in a sensitivity analysis to 7.5% and 30% respectively.

We used key parameters from the Nigeria data preloaded in LiST, and adjusted the population size to that covered by PBF. LiST produced the number of lives saved from improved interventions. We converted this

into quality-adjusted life years (QALYs) applying the formula for fatal cases (Sassi, 2006) using Nigeria's life table and disease burden. (WHO, 2015; Hay, et 2016). Years gained in early life play a more important role in determining QALYs. We estimated total QALYs gained by multiplying the number of cases saved by QALYs gained per case. Additionally, as the result from the LiST was estimated by years, we estimated lives saved and QALYs gained for two years, and then extrapolated to 2.8 years, the duration of PBF implementation before the mid-term evaluation.

To estimate the QALYs gained from NSHIP, we combined the QALYs gained from PBF and DFF groups while adjusting the population difference. Additionally, both costs and QALYs gained were rescaled by the population size to estimate cost and QALYs gained per capita.

Cost-effectiveness analysis and sensitivity analysis

Based on incremental costs and effectiveness per capita, we then estimated the incremental cost-effectiveness ratio (ICER) under two scenarios: one without quality improvement and the other with it. The sensitivity analysis was focused on the relative impact of quality of care over that of the coverage of services. In this study, when we adjusted quality of care to generate an index of quality-adjusted coverage, the underline assumption is that the impact from additional 1% increase in quality of care is equivalent to that from 1% increase in coverage of the health service. In the sensitivity analysis, we generated a scenario where the impact from 2% increase in quality of care was equivalent to that from 1% increase in coverage.

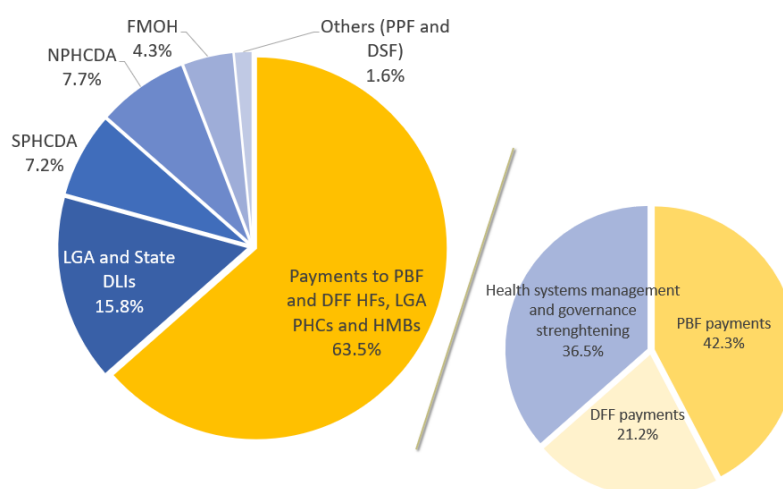
While there has been much debate over thresholds of cost-effectiveness analysis (Marseille et al., 2018; Bertram et al., 2016, Neumann et al., 2014), a study evaluating returns on investment specific to MCH valued a healthy life year as 1.5 times GDP per capita (Stenberg et al., 2013) which we used as the threshold to interpret the results. In 2015, GDP/capita was \$2,655 in Nigeria, (World Bank, 2015) and thus the threshold was estimated at \$3,983. Interventions with ICERs lower than the threshold value were regarded as cost-effective.

10.2 Findings

As of June 2017, a total of USD132.9 million (129.4 million in 2015 USD) were spent on operating the PBF and DFF programs, among which USD 56.3 million (54.8 million in 2015 USD) were used for incentive

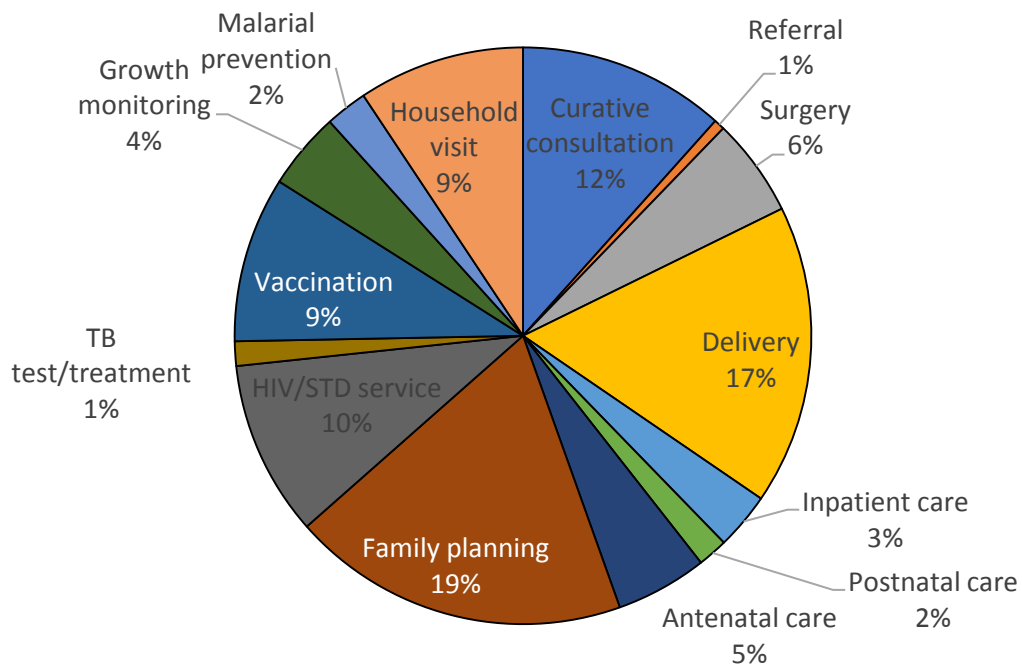
payments to PBF facilities while USD 28.1 million (27.4 million in 2015 USD) were for DFF facilities (50% of incentive payment for PBF), and the rest of USD 48.5 million (47.2 million in 2015 USD) were for the health systems management and governance strengthening at the LGA, State and Federal levels such as for technical assistance, payments to local government agency (LGA) and state disbursement linked indicators (DLIs), monitoring and evaluation and so on (Figure 16). Annual overall per capita expenditure was \$2.6.

Figure 16: NSHIP disbursements by component



Among the incentive payments to PBF health facilities, family planning, institutional delivery, curative consultation, HIV/AIDS services, household visits and vaccination shared the greatest portion of incentives, representing 19%, 17%, 12%, 10%, 9%, and 9%, respectively (Figure 17). It is expected that family planning and institutional delivery may be improved in the PBF facilities.

Figure 17: Distribution of Incentives Payments



The World Bank headquarters costs for supervising and planning the program was estimated to be USD 3.03 million (USD 3.06 million in 2015 US dollars).

Among 5.67 million population in the PBF group, without including quality of care in the analysis, the PBF program resulted in saving 2,584 lives compared to DFF, and 4,700 lives compared to the control group (Table 41). The majority of lives saved were for children under-5. Adjusting for quality, PBF saved 12,488 more lives as compared to the control group—the quality-adjusted impact of PBF is almost triple the impact without quality. Adjusting DFF impact for quality increases lives saved by 52.8%. Quality-adjusted impact of DFF was estimated at 3,951 lives saved as compared to control. Overall, NSHIP averted 6,740 maternal and child deaths in the project states as compared to the control states—the quality-adjusted impact of NSHIP is higher at 20,721 lives saved.

Table 41: Number of Deaths and Lives Saved

	Number of deaths					Number of lives saved							
	Control	RBF	DFF	RBF-quality adjusted	DFF-quality adjusted	RBF vs control	DFF vs control	RBF vs DFF	NSHIP vs control	RBF vs control (QA)	DFF vs control (QA)	RBF vs DFF (QA)	NSHIP vs control (QA)
Children under-5													
2015	18,983	18,074	18,593	16,623	17,244	909	390	519	1,285	2,360	1,739	621	4,037
2016	18,066	16,033	17,089	12,670	14,390	2,033	977	1,056	2,975	5,396	3,676	1,720	8,941
Subtotal	37,049	34,107	35,682	29,293	31,634	2,942	1,367	1,575	4,260	7,756	5,415	2,341	12,978
Maternal deaths													
2015	1,744	1,604	1,698	1,407	1,507	140	46	94	184	337	237	100	566
2016	1,670	1,395	1,572	843	1,224	275	98	177	370	827	446	381	1,257
subtotal	3,414	2,999	3,270	2,250	2,731	415	144	271	554	1,164	683	481	1,823
Total (2 years)	40,463	37,106	38,952	31,543	34,365	3,357	1,511	1,846	4,814	8,920	6,098	2,822	14,801
Total (2.8 years)	56,648	51,948	54,533	44,160	48,111	4,700	2,115	2,584	6,740	12,488	8,537	3,951	20,721

When converting health benefits to QALYs gained, it shows that without quality adjustment PBF, compared to DFF and control facilities, saved 110,896 QALYs and 60,837 QALYs, respectively (Table 42). These were equivalent to gaining 0.0107 QALYs and 0.0195 QALYs per capita. The number of QALYs gained was larger when the improvement of quality of care was considered.

Table 42: QALYs Saved from Programs

	RBF vs control	DFF vs control	RBF vs DFF	NSHIP vs control	RBF vs control (QA)	DFF vs control (QA)	RBF vs DFF (QA)	NSHIP vs control (QA)
Children under five	98,410	45,726	52,684	142,506	259,439	181,132	78,307	434,113
Pregnant women	12,486	4,332	8,153	16,664	35,021	20,549	14,472	54,837
All	110,896	50,059	60,837	159,170	294,459	201,681	92,778	488,950

Table 43 shows ICERs of the PBF program in comparison with DFF and control groups, and of NSHIP overall. The ICERs of PBF compared to DFF and control were \$698 and \$796/QALY gained respectively, without quality of care adjustment. These ratios fell to \$458 and \$300/QALY gained if the quality of care was added. The ICER for NSHIP was estimated to be \$831 and \$271/QALY gained, without and with quality adjustment, respectively. Table 44 shows that even with the assumption of lower quality of care in skilled delivery as modelled in LiST, all scenarios are highly cost-effective.

Table 43: Incremental Cost Effectiveness Ratio

	Cost/life saved	Cost/QALY gained	Cost/life saved (QA)	Cost/QALY gained (QA)
PBF vs Control	18,777	796	7,067	300
DFF vs Control	21,630	914	5,360	227
PBF vs DFF	16,442	698	10,756	458
NSHIP vs Control	19,641	831	6,388	271

Table 44: LiST Sensitivity Analysis—Incremental Cost Effectiveness Ratios

	Cost/life saved	Cost/QALY gained	Cost/life saved (QA)	Cost/QALY gained (QA)
RBF vs Control	19,754	836	8,091	342
DFF vs Control	21,278	900	5,647	239
RBF vs DFF	18,340	777	15,153	642
NSHIP vs Control	20,237	856	7,071	299

The sensitivity analysis of quality of care results linked through the Delphi survey showed that if the effect of quality of care was only half of the impact of the coverage of care, then the ICER of PBF increased to \$990 and \$595/QALY gained, compared to DFF and control groups, respectively. This showed that the ICER of PBF was quite sensitive to the quality adjustment. The results should be interpreted with caution.

To summarize, the cost effectiveness analysis suggests that PBF is cost-effective as compared to the control group regardless of the inclusion of the impact of the improvements in quality of care. Improvements in skilled birth attendance and modern contraceptive use are the key components that drive the effectiveness of NSHIP. DFF is also cost effective, albeit through improvements in vaccination coverage. PBF is marginally more cost-effective than DFF in averting maternal and child deaths as the LiST model indicates that per percentage point improvement in SBA saves more maternal and child lives than per percentage point increase in immunization services¹¹. Effectiveness of both PBF and DFF is driven by the improvements in the quality of care, which lead to a significant reduction in incremental cost effectiveness ratio. Not including the improvements in quality of care severely underestimate the effectiveness of NSHIP.

¹¹ Particularly as even with the increased immunization, the NSHIP states are nowhere close to the rates at which herd immunity benefits materialize.

11 Conclusion

This impact evaluation demonstrates that both PBF and DFF components of the Nigeria State Health Investment Project had important effects on the coverage and structural quality of MCH services while the control arm, like the rest of Nigeria, made only modest progress. Under real-world conditions and at large scale PBF and DFF appear to be practical and scalable interventions in the Nigerian context. Further, the improvements seen under NSHIP were accomplished at a cost that is affordable using domestic resources, particularly if the BHCPF is implemented and funded as envisaged in the National Health Act. PBF and DFF are cost-effective compared to Nigeria's per capita GDP.

The similar results achieved by PBF and DFF suggest that providing operating budgets to health facilities, allowing them to spend the funds on their perceived priorities, systematic supervision using a QSC, and strengthened management and governance at LGA, state and federal levels were the main reasons for the success of NSHIP. Of course, DFF may have benefited from the concurrent implementation of PBF. However, it's possible that if supervision is strengthened, management improved, and monitoring and evaluation robustly implemented, the DFF model can achieve similar results at lower cost.

The financial review of NSHIP found better, albeit imperfect bookkeeping at treatment facilities relative to control. Nonetheless, not all assets procured with NSHIP funds were recorded. In general, however, funds were being used for operational expenses like contract staff salary, performance bonus payments at PBF facilities, purchase of drugs, maintenance of the facilities, etc. However, in all the three states, cheques were being raised in favour of officers at the health facilities. Cash payments were preferred over online and cheque transfers, which leaves the door open to misappropriating funds as most of the expenditures were neither receipted nor captured adequately in the financial records.

The key conclusions from cost-effectiveness analysis suggest that when compared to the control group, PBF is cost-effective, no matter whether the improvements in quality of care are included in the analysis or not. Improvements in skilled birth attendance and modern contraceptive usage are the key components that drive the effectiveness of NSHIP. DFF is also cost effective, albeit through improvements in vaccination coverage. The effectiveness of both PBF and DFF arms is driven by improvements in quality of care, which lead to a significant reduction in incremental cost effectiveness ratio. Not including the improvements in quality of care severely underestimates the effectiveness of NSHIP.

References

- Ashir, G. M., Doctor, H. V., & Afenyadu, G. Y. (2013). Performance based financing and uptake of maternal and child health services in yobe state, northern Nigeria. *Global journal of health science*, 5(3), 34.
- Adeyi, O. (2016). Health System in Nigeria: From Underperformance to Measured Optimism. *Health Systems & Reform*, 2(4), 285-289.
- Arur, A., Mohammed-Roberts, R., Bos, E. (2011). SETTING TARGETS IN HEALTH, NUTRITION AND POPULATION PROJECTS. HNP Discussion Paper. Washington D. C.: World Bank Avenir Health. (2015). *Spectrum Manual: Spectrum System of Policy Models*. Glastonbury, CT: Avenir Health.
- Basinga, P., Gertler, P. J., Binagwaho, A., Soucat, A. L., Sturdy, J., & Vermeersch, C. M. (2011). Effect on maternal and child health services in Rwanda of payment to primary health-care providers for performance: an impact evaluation. *The Lancet*, 377(9775), 1421-1428.
- Bertram, M. Y., Lauer, J. A., De Joncheere, K., Edejer, T., Hutubessy, R., Kieny, M. P., & Hill, S. R. (2016). Cost-effectiveness thresholds: pros and cons. *Bulletin of the World Health Organization*, 94(12), 925.
- Bhatnagar, A., & George, A. S. (2016). Motivating health workers up to a limit: partial effects of performance-based financing on working environments in Nigeria. *Health policy and planning*, 31(7), 868-877.
- Boschi-Pinto, C., Young, M., & Black, R. E. (2010). The Child Health Epidemiology Reference Group reviews of the effectiveness of interventions to reduce maternal, neonatal and child mortality. *International Journal of Epidemiology*, 39(suppl_1), i3-i6.
- Campbell, S. M., Reeves, D., Kontopantelis, E., Sibbald, B., & Roland, M. (2009). Effects of pay for performance on the quality of primary care in England. *New England Journal of Medicine*, 361(4), 368-378.
- Central Intelligence Agency. (2018). *The World Factbook*.
- De Walque, D., Robyn, P. J., Saidou, H., Sorgho, G., & Steenland, M. (2017). *Looking into the*

performance-based financing black box: evidence from an impact evaluation in the health sector in Cameroon. The World Bank. Fritsche, G. B., Soeters, R., & Meessen, B. (2014). Performance-based financing toolkit. world bank publications.

Ganle, J. K. (2015). Why Muslim women in Northern Ghana do not use skilled maternal healthcare services at health facilities: a qualitative study. *BMC international health and human rights*, 15(1), 10.

Glickman, S. W., Ou, F. S., DeLong, E. R., Roe, M. T., Lytle, B. L., Mulgund, J., Rumsfeld, J. S., Gibler, W. B., Ohman, E. M., Schulman, K. A., & Peterson, E. D. (2007). Pay for performance, quality of care, and outcomes in acute myocardial infarction. *Jama*, 297(21), 2373-2380.

Hay, S. I., Abajobir, A. A., Abate, K. H., Abbafati, C., Abbas, K. M., Abd-Allah, F., et al. (2017). Global, regional, and national disability-adjusted life-years (DALYs) for 333 diseases and injuries and healthy life expectancy (HALE) for 195 countries and territories, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. *The Lancet*, 390(10100), 1260-1344.

Hirano, K., Imbens, G. W., & Ridder, G. (2003). Efficient estimation of average treatment effects using the estimated propensity score. *Econometrica*, 71(4), 1161-1189.

Hongoro, C., & Normand, C. (2018). Health Workers: Building and Motivating the Workforce In Dean T. Jamison, Rachel Nugent, Hellen Gelband, Susan Horton, Prabhat Jha, Ramanan Laxminarayan, Charles N. Mock. (Eds.), *DISEASE CONTROL PRIORITIES* (pp. 1309-1322). Washington D. C.: World Bank.

Ireland, M., Paul, E., & Dujardin, B. (2011). Can performance-based financing be used to reform health systems in developing countries?. *Bulletin of the World Health Organization*, 89(9), 695-698.

Keats, E. C., Ngugi, A., Macharia, W., Akseer, N., Khaemba, E. N., Bhatti, Z., et al. (2017). Progress and priorities for reproductive, maternal, newborn, and child health in Kenya: a Countdown to 2015 country case study. *The Lancet Global Health*, 5(8), e782-e795.

Kombe, Gilbert; Lisa Fleisher; Eddie Kariisa; Aneesa Arur, Parsa Sanjana (Abt Associates Inc. Health Systems 20/20); Ligia Paina (USAID); Lola Dare, Ahmed Ubok-Udom, Sam Unom. April 2009. *Nigeria Health System Assessment 2008*. Abt Associates Inc.

Marseille, E., Larson, B., Kazi, D. S., Kahn, J. G., & Rosen, S. (2014). Thresholds for the cost–

effectiveness of interventions: alternative approaches. *Bulletin of the World Health Organization*, 93, 118-124.

National Population Commission (NPC) [Nigeria] and ICF International. (2014). *Nigeria Demographic and Health Survey 2013*. Abuja, Nigeria, and Rockville, Maryland, USA: NPC and ICF International.

National Population Commission (NPC) [Nigeria] and ICF Macro. (2009). *Nigeria Demographic and Health Survey 2008*. Abuja, Nigeria: National Population Commission and ICF Macro.

Neumann, P. J., Cohen, J. T., & Weinstein, M. C. (2014). Updating cost-effectiveness—the curious resilience of the \$50,000-per-QALY threshold. *New England Journal of Medicine*, 371(9), 796-797.

Nguyen, H. T. H., Gopalan, S., Mutasa, R., Friedman, J., Das, A. K., Sisimayi, C., ... & Kane, S. (2015). *IMPACT OF RESULTS-BASED FINANCING ON HEALTH WORKER SATISFACTION AND MOTIVATION IN ZIMBABWE*.

Peabody, J., Shimkhada, R., Adeyi, O., Wang, H., Broughton, E., & Kruk, M. E. (2018). Quality of Care In Dean T. Jamison, Rachel Nugent, Hellen Gelband, Susan Horton, Prabhat Jha, Ramanan Laxminarayan, Charles N. Mock. (Eds.), *DISEASE CONTROL PRIORITIES* (pp. 185-213). Washington D. C.: World Bank.

Renmans, D., Holvoet, N., Orach, C. G., & Criel, B. (2016). Opening the ‘black box’ of performance-based financing in low-and lower middle-income countries: a review of the literature. *Health policy and planning*, 31(9), 1297-1309.

Sassi, F. (2006). Calculating QALYs, comparing QALY and DALY calculations. *Health policy and planning*, 21(5), 402-408.

Shen, G. C., Nguyen, H. T. H., Das, A., Sachingongu, N., Chansa, C., Qamruddin, J., & Friedman, J. (2017). Incentives to change: effects of performance-based financing on health workers in Zambia. *Human resources for health*, 15(1), 20.

Singh, S., Darroch, J. E., & Ashford, L. S. (2014). *Adding it up: The costs and benefits of investing in sexual and reproductive health 2014*. New York, NY: Guttmacher Institute.

Stenberg, K., Axelson, H., Sheehan, P., Anderson, I., Gülmezoglu, A. M., Temmerman, M., ... & Sweeny,

K. (2014). Advancing social and economic development by investing in women's and children's health: a new Global Investment Framework. *The Lancet*, 383(9925), 1333-1354.

Stenberg, K., Axelson, H., Sheehan, P., Anderson, I., Gülmezoglu, A. M., Temmerman, M., et al. (2014). Advancing social and economic development by investing in women's and children's health: a new Global Investment Framework. *The Lancet*, 383(9925), 1333-1354.

The World Bank (2015). GDP per capita (current US\$). Washington, DC: The World Bank; [cited 2015 Sept 30]; Available from: <http://data.worldbank.org/indicator/NY.GDP.PCAP.CD>.

Werner, R. M., Kolstad, J. T., Stuart, E. A., & Polsky, D. (2011). The effect of pay-for-performance in hospitals: lessons for quality improvement. *Health Affairs*, 30(4), 690-698.

World Bank. (2012). Nigeria State Health Investment Project. Washington D. C.: World Bank.

World Health Organization (2015). Zambia: Global health observatory data repository. Geneva, Switzerland: WHO; 2015 [cited 2015 Sept 30]; Available from: <http://apps.who.int/gho/data/view.main.61850?lang=en>.

World Health Organization. (2006). The world health report 2006: working together for health. World Health Organization.