Final Results of the SSDI-PBI Process and Impact Evaluation

Prepared by Research Teams From

Heidelberg University (Germany)

and

The College of Medicine (Malawi)

in collaboration with

Bocconi University (Italy)

July 2016

Prepared by: Shannon McMahon, Manuela De Allegri (PI)

Edited by: Adamson Muula (Co-PI), Stephan Brenner, Rachel Chase, Julia Lohmann, Christopher Makwero, Aleksandra Torbica

This study is made possible by the support of the American People through the United States Agency for International Development (USAID) via Translating Research into Action (TRAction). The results of this study do not necessarily reflect the views of USAID or the United States Government. Translating Research into Action, TRAction, is funded by United States Agency for International Development (USAID) under cooperative agreement number GHS-A-00-09-00015-00. The project team includes prime recipient, University Research Co., LLC (URC), Harvard University School of Public Health (HSPH), and sub-recipient research organizations.

Executive Summary

This report is the product of a 1-year study led by Heidelberg University in Germany and the College of Medicine in Malawi in collaboration with Bocconi University in Italy, which were commissioned to evaluate a health-financing program called "SSDI-PBI" in Malawi. Implemented by the Ministry of Health with funding from USAID and technical support from Jhpiego and Abt associates, SSDI-PBI operates in 17 facilities across three districts. Program implementation began in September 2014. The evaluation of the program began in July 2015 and included four study components: fidelity of implementation, an assessment of the program's effects on health service utilization, an assessment on provider motivation, and an economic evaluation.

The fidelity of implementation component found that, after an initial (pre-implementation) overhaul of the program design, the program was later executed in a manner that was faithful to its most critical elements. Factors mentioned most often as underscoring success within the SSDI-PBI program are largely rooted in the **successful execution of essential PBI program ingredients**. In this sense, the program fostered changes in attitudes and behaviors across multiple levels and then benefitted from those changes. Nevertheless, **a persistent challenge that undermines program success entails delays in procurement**. Measures to mitigate, reduce or remove penalties against facilities that have ordered though not received goods or cash, or that cannot meet a given indicator due to structural limitations (such as national stock-outs of a drug) warrants consideration. Several respondents also mentioned a need to foster a sense of ownership across stakeholders; at present the program is often viewed as "owned by" the implementer alone.

The economic evaluation found that the **intervention was valued at over $3 million** (USD), that **personnel costs were the most costly component** (1.4 times more costly than the costs of financial incentives), and that the **design phase absorbed one third of all costs**. Incentives paid to facilities differed substantially across facilities, but as time passed incentive payments gradually rose across facilities, which we view as an indication of health facilities' ever-increasing responsiveness to the intervention.

The service utilization component of the study found that the SSDI-PBI program had **overall positive effects on services related to maternal and newborn health** (antenatal care, prevention of mother-to-child transmission, and to some degree postnatal care), that the program positively affected service quality related to HIV counseling and testing and Vitamin A distribution, but that the program had **no effect on counseling services** related to family planning or skilled attendance at birth, and that child **immunization services were negatively affected**. In an overarching sense, the program seemed to bolster stronger changes in health centers compared to hospitals, suggesting that **it may be useful to pursue a more tailored approach to service provision by health facility cadre**. Female clients generally expressed no knowledge of the SSDI-PBI program and struggled to describe differences observed within facilities since SSDI-PBI. **Community leaders were highly conversant on the program**, were generally positive about the program and shared several suggestions on how the program could be improved in the future.

The provider motivation component found that **providers positively perceive the intervention** in all respects *except* procurement, thereby triangulating information gleaned in fidelity interviews. Implementation challenges as well as critical shortages in human resources and drugs prevented the intervention from unleashing its full motivational effect. Finally, **providers report that their motivation to make an effort and perform well has increased with PBI**.

Looking ahead, this report identified several strengths as well as opportunities to refine the program. Recommendations include: engaging and sensitizing a broader network of stakeholders and beneficiaries, refining indicators to align with on-the-ground realities, and streamlining procurement. Roughly half of all respondents in this study said it would be beneficial to **financially compensate providers directly**. Ministry staff clarified that such an adjustment would not be sustainable in the long-term.

Table of Contents

List of Abbreviations

AIDS	Acquired immunodeficiency syndrome
ANC	Antenatal care
ART	Antiretroviral therapy
BCG	Bacillus Calmette–Guérin
CoM	College of Medicine (of Malawi)
DC	Districts Commissioners
DHIS2	District Health Information Software version 2
DHO	District Health Officer
DHMT	District Health Management Team
DHS	Demographic and Health Survey
DiD	Difference-in-differences
DMO	District Medical Officer
DNO	District Nursing Officer
EHP	Essential health package
FeFo	Iron-folate
FEFOL	Ferrous sulfate with Folic acid
FGD	Focus group discussion
FOI	Fidelity of implementation
FP	Family planning
HAC	Health advisory committee
HIV	Human Immunodeficiency Virus
HMIS	Health Management Information System
HSA	Health surveillance assistant
HTC	HIV testing and counseling
ICER	Incremental cost effectiveness ratio
IDI	In-Depth Interview
IP	Infection Prevention
IPT	Intermittent preventive treatment
ITS	Interrupted time series
LMIC	Low and middle income countries
MDG	Millennium Development Goal
MKW	Malawian Kwacha
MoH	Ministry of Health
NGOs	Non-Government Organizations
PBF	Performance-based financing
PBI	Performance-based incentives
PI	Principal Investigator
PNC	Postnatal care
PEPFAR	Presidents Emergency Plan for AIDS Relief
PMTCT	Prevention of mother-to-child transmission (of HIV)
PQI	Performance Quality Improvement
QoC	Quality of care
RBF4MNH	Results-Based Financing for Maternal Neonatal Health
SBA	Skilled birth attendant
SBM-R	Standards Based Management and Recognition
SP	Sulfadoxine-Pyrimethamine
SPA	Service Provision Assessment
SSA	Sub-Saharan Africa
SSDI	Support for Service Delivery Integration
SSDI-PBI	Support for Service Delivery Integration Performance-Based Incentives
TB	Tuberculosis
TOI	Theory of Intervention
TRAction	Translating Research into Action
USAID	United States Agency for International Development
USD	US Dollar

List of Figures

List of Tables

Background

The intervention under evaluation in this report is entitled the Support for Service Delivery Integration Performance-Based Incentives (SSDI-PBI) Project[1]. The Ministry of Health (MoH) of Malawi launched the intervention in 2014 with financial support from the United States Agency for International Development (USAID). Performance-based incentives (PBI) refer to a range of health system interventions that provide financial rewards based on the attainment and verification of predefined quantity and/or quality outputs[2-4]. Performance-based approaches link incentives to desired outputs, in an attempt to spark an entrepreneurial, autonomous spirit among providers. In the past 10 years, performance-based programs have proliferated across sub-Saharan Africa (SSA). A landscape analysis of programs that were initiated between 2008 and 2015 identified 32 programs across low and middle income countries (LMICs) - including 25 programs across 21 countries in SSA, of which two programs were situated in Malawi (one of which is the focus of this report)[5].

Health in Malawi

Table 1. Quantity indicators used in SSDI-PBI

1.	Number of pregnant women starting antenatal care during the first trimester
2.	Number of women completing the four antenatal care visits
3.	Number of pregnant women receiving at least two doses of intermittent preventive therapy
4.	Number of births attended by skilled birth attendants (doctor, nurse or midwife)
5.	Number of 1-year-old children fully immunized
6.	Number of HIV-positive pregnant women who were initiated on antiretroviral therapy
7.	Number of HIV/AIDS cases screened for Tuberculosis
8.	Number of children receiving Vitamin A supplementation
9.	Number of clients counseled for family planning
10.	Number of couples tested for HIV during HIV testing and counseling services
11.	Number of infants born by HIV positive mothers tested for HIV
12.	Number of women who receive postnatal care after delivery by skilled health workers within seven days
13.	Number of pregnant women attending antenatal care receiving iron supplementation

Malawi's MoH considers PBI a potential solution to the longstanding problem of inadequate service provision[6]. Malawi suffers from a heavy burden of HIV and communicable diseases (especially tuberculosis and Malaria) and, more recently, increases in non-communicable diseases (hypertension, diabetes and cancer)[7]. While the country has met several of its Millennium Development Goals (MDG) targets including those related to child mortality (MDG 4) and HIV and AIDS (MDG 6), other targets were not met including those related to maternal mortality (MDG 5)[8]. High rates of morbidity and mortality – particularly maternal mortality – have been linked to shortages in human resources for health, and inadequacies within facilities related to basic and essential infrastructure, management, support and services[9-11]. Healthcare delivery is largely centered around provision of an essential healthcare package (EHP) (including reproductive health services, child health services, as well as services related to the prevention, detection and management of infectious and non-communicable health problems) which is intended to be provided free of charge at point of use either in public facilities or in private not-for-profit facilities contracted by the Ministry of Health (MoH)[12]. Evidence indicates, however, that services included in the EHP are not as effectively available as they should be, thereby subjecting clients to substantial out-of-pocket expenditures[13-16].

The SSDI-PBI Intervention

The objective of the SSDI-PBI initiative is to combine efforts of local and foreign organizations in order to strengthen the provision of EHP services[12]. Two SSDI sectors, SSDI-Systems and SSDI-Services, designed a PBI intervention that has been implemented in 17 facilities across three SSDI target districts (Chitipa, Nkhotakota, Mangochi) since 2014. SSDI-Systems, led by Abt Associates, focused primarily on the design of

the SSDI-PBI program. SSDI-Services, led by Jhpiego, focused primarily on the implementation of the SSDI-PBI program. SSDI-PBI facilities were selected (non-randomly) based on minimum quotas related to equipment, infrastructure and personnel as deemed necessary to guarantee adequate EHP service delivery. Most facilities were also chosen based on their inclusion in a larger quality improvement program implemented by Jhpiego called "PQI" (for Performance Quality Improvement). PQI is based on the Standards-Based Management and Recognition (SBM-R) approach to quality improvement, which urges providers and staff to consider the root causes and attainable solutions to address poor performance [17]. In this respect, the SSDI-PBI program builds upon an existing intervention that aims to improve provider performance and service delivery [17].

Table 2. Quality dimensions assessed in SSDI-PBI

1.	General activities
2.	Follow-up assessment and HMIS
3.	Hygiene, environment, and sterilization
4.	Outpatient and inpatient consultation
5.	Maternity ward
6.	Antenatal consultation
7.	Family planning
8.	Vaccination and monitoring of newborns
9.	HIV/AIDS control
10.	Tuberculosis
11.	Laboratory
12.	Minor surgery
13.	Drug and commodity management

The SSDI-PBI program aims to increase access, utilization, and quality of EHP services by linking rewards to service utilization and quality indicators across a range of conditions and services. Utilization, or quantity, indicators focus on increasing total counts in terms of services across the maternal health continuum of care (during antenatal, delivery and postpartum periods), newborn and child health, and HIV and AIDS care and treatment (see Table 1 for a list of quantity indicators). Quality indicators emphasize improvements in the broader facility environment and in the nature of how care is provided across 13 service areas (see Table 2 for a list of quality service areas). Quality assessments are complemented by a community component wherein focus groups (also called "community scorecards") and exit interviews are conducted with clients to gauge satisfaction.

SSDI-PBI rewards are comprised primarily of quantity and quality scores, with community scores serving as a source of potential bonus payments. Rewards are paid to facilities upon achievement of set targets, but the rewards can only be used toward facility improvements and cannot be partially redistributed in the form of performance bonuses to individual health workers, which is common under other performance-based schemes. Another defining characteristic of the rewarding system is that procurement at the facility level is managed through existing SSDI finance and procurement structures rather than through facility-based personnel. Rewarded funds are invested in previously determined service improvement activities or strategies outlined in annual business plans, which are developed by facility staff in collaboration with SSDI staff. These plans outline the activities, procurements or technical support that facility staffs intend to prioritize as a means to improve care.

The Evaluation Study Design

The evaluation was launched in 2015 and led by Heidelberg University in Germany, the College of Medicine in Malawi, and Bocconi University in Italy. The evaluation relied on mixed-methods, applying both quantitative and qualitative methods of data collection and analysis. The study was partially prospective, collecting and analyzing primary and secondary data during the year of study implementation, and partially retrospective, relying on secondary data existing at the time leading up to and including the study launch.

The study team undertook a combined impact and process evaluation, with an aim to assess the effect that the SSDI-PBI intervention produced on the work environment, on provider behavior (including service outputs) and on providers' sense of knowledge, motivation and self-efficacy. Via a fidelity of implementation (FOI) assessment, the study also looked at intervention effects in light of contextual

factors that shaped program implementation. Finally, the study examined the costs of implementing the intervention in relation to outcomes produced. The research questions were as follows:

1. Focusing on service provision, to what extent did the SSDI-PBI intervention produce changes in the quantity and quality of services provided? Which work environment changes could be attributed to PBI (i.e. availability of equipment, drugs, staff, training, supervision in respect to clinical performance)? What heterogeneity in effects could be observed across districts and facilities? To what extent have changes affected incentivized vs. non-incentivized services?
2. Focusing on providers specifically, how has the SSDI-PBI intervention changed motivation of health workers? Are changes in motivation reflected in changed attitudes or behavior at work?
3. Focusing on fidelity, to what extent was the SSDI-PBI intervention implemented according to its original implementation plan? Which contextual factors affected implementation, as defined in relation to acceptance and adoption of the intervention, at the various levels of service provision, including at health facility levels? Which contextual factors explained heterogeneity in implementation processes across districts and facilities?
4. Focusing on efficiency, what were the costs of implementing the SSDI-PBI intervention in relation to the outcomes produced? Is the economic burden of designing, implementing and managing the SSDI-PBI system worthwhile considering results achieved?

An outline of methods employed across study components is detailed in Table 3.

Table 3. Research questions, methodological approach, data collection activity

Research Question Divided by Study Component	**Approach**	**Data Collection (a) Data Collection Instrument (b) Sampling Unit (c) Data Content**	
		Quantitative	**Qualitative^**
Study Component 1. Fidelity of Implementation. How has the intervention aligned with intended design, and what factors have affected this?	Qual	--	a. IDI^ & Document Review
			b. IDIs with MoH, Funders, PBI desk officers, DHMTs, SSDI employees, health workers in Intervention Facilities; Document Review of implementation planning and monitoring material
			c. Primary and Secondary data
Study Component 2. Service Provision. How has the intervention affected quality of service provided and why?	Quant & Qual	a. Structured Checklist; Data extraction lists for (i) routine surveillance databases (HMIS) and (ii) Service Provision Assessment (at baseline)	a. IDI & FGD^
		b. Health facility (intervention and control)	b. Clients, Community members
		c. Primary and Secondary data	c. Primary data
Study Component 3. Provider Motivation. How has the intervention affected health worker motivation?	Qual & Quant	a. Health worker survey	a. IDI
		b. Health workers in intervention and control	b. Health workers in intervention facilities
		c. Primary and secondary data	c. Primary data
Study Component 4. Economic Evaluation. What are the costs of implementing the intervention in relation to outcomes produced?	Quant	a. Data Extraction	--
		b. Implementer materials (SSDI costing data)	
		c. Secondary data	
^Methods acronyms: IDI is in-depth interview, FGD is focus group discussion			

Accomplishments

This evaluation achieved all stated objectives. The team adjusted to emerging field conditions, adapted study tools to contend with existing realities, identified complementary data sets (for secondary data

collection) and in the end, delivered a rich product that aligns with the original research proposal. Furthermore, the team built on local capacity by providing a travel stipend to a Malawian masters student who was completing her public health degree at Heidelberg University. This student collected qualitative, in-depth interview data related to service utilization (detailed within Study Component 2b below). Accomplishments are summarized below.

Successful completion of all data collection activities

During the one-year study period, the research team collected the following surveys and interviews:

- 25 IDIs with stakeholders related to program fidelity
- 34 facility inventories
- 30 FGDs with women and community leaders
- 29 IDIs with facility-based providers
- 76 structured interviews with facility-based providers
- Collection of secondary data
    - For Service Utilization
        - Health Management Information Systems (HMIS) data, which was crosschecked with data from the Presidents Emergency Plan for AIDS Relief (PEPFAR) data
        - Service Provision Assessment (SPA) data, which was used as a proxy for baseline data and complemented with primary data collection on facility inventories
    - For Economic Evaluation
        - Financial statements from implementers (Jhpiego and Abt Associates)
        - Estimated personnel time dedicated to the program (Ministry of Health and USAID)

Successful establishment of supportive exchange and collaboration with the SSDI-implementation team

From the outset of the program, the evaluation team sought to open collaborative and constructive communication channels with the implementation and funder teams. In August 2015 the evaluation team met with implementers and funders (from USAID and TRAction) in Lilongwe to present the study, seek consensus and insights on study components (and the feasibility of attendant research approaches) and to coordinate agreements on the nature of secondary data available and the timing of its delivery to the evaluation team. Given the absence of key stakeholders at this meeting, the scientific coordinator from the evaluation team returned to Lilongwe in November to again present the study and seek clarification regarding the timing and availability of secondary data (from funders, implementers and the Ministry). In March, 2016 the scientific coordinator again met with key stakeholders to present progress related to data collection (and in some cases to conduct IDIs). In May 2016, the PI, co-PI and scientific coordinator presented preliminary findings in two meetings with the funding and implementing teams, respectively.

Successful supervision of the one Masters candidate resulting in an (expected) award of Master title

As part of an effort to foster local capacity, the evaluation team extended a stipend to one Malawian student who was enrolled in the Masters of Public Health program at the Heidelberg University. This arrangement was not initially envisioned as a component of the evaluation, but ultimately proved beneficial to both the evaluation team's efforts and the student's efforts to establish her research career. The student collected IDI data from clients living far from intervention and control facilities as a means to explore facets of service utilization. Her thesis is in progress, and she is poised to complete her work successfully before the close of 2016. The student also advised the evaluation team in terms of refining the phrasing of data collection tools.

Study Component 1. Fidelity of Implementation

The Challenge

The implementation of complex health interventions presents challenges that are multifaceted, and tied to acceptability, feasibility, adaptability, affordability and ownership. Programs must gain and sustain "buy-in" across a spectrum of stakeholders- from officials at ministry and regional levels, to district health staff, facility-based providers and community leaders. An added layer of complexity is inherent to programs that are new or novel; under these circumstances an extra dose of sensitization and knowledge dissemination is needed.

The design of the Support for Service Delivery Integration- Performance Based Incentives (SSDI-PBI) program began in 2012, and implementation started in September 2014.

Methods

Focusing on the implementation period, this study component relied most heavily on a document review (of implementing materials) conducted in November and December 2015, followed by a series of 25 in-depth interviews with SSDI-PBI staff, Ministry of Health representatives, USAID representative, health providers, community Leaders (namely Health Advisory Committee members) conducted in March 2016 (see Table 4). The interview guide was semi-structured in format and covered the following topics: program design, program implementation, program modification, program adoption and recommendations looking ahead. All interviews were transcribed and coded using NVivo[18]. Codes were deductive and based on a codebook informed by Hasson 2010[19] which emphasizes several components inherent to a fidelity of implementation study including adherence (the degree to which an implementation aligns with its design), and moderating factors (such as program complexity, context of a program, implementation facilitators). While not a tenet of fidelity studies, interviews (and thus analysis) also probed on program sustainability and recommendations moving forward.

Table 4. In-depth interview sample characteristics

In-Depth Interview Respondent	n
Female	7
Male	18
SSDI-PBI staff (incl. Abt, Jhpiego)	8
Ministry of Health	5
USAID	1
Health professional at facility or district level (incl. PBI coordinator)	9
Community leaders	2
TOTAL	**25**

Results

Execution and Timing of Key Intervention Components

The fidelity of implementation component found that after an initial (pre-implementation) overhaul of the program design (to eliminate the provision of salary top-ups to providers), the program was later executed in a manner that was faithful to its most critical elements which include: program sensitization (within

facilities and among communities), business plan development, supportive supervision, verification and assessment, documenting data (introducing or expanding monitoring and evaluation skills within facilities), tracking data and rewarding progress.

In terms of timing, the design overhaul forced a delay in the start of the program (from an initial start date of April 2013 to September 2014), and necessitated further sensitizations at several levels (from facilities up to the Ministry of Health) regarding the model of the new scheme and the rationale behind the elimination of individual bonuses. This process of altering the design and then re-sensitizing stakeholders on the redesign, forced a late start to program implementation. This late start then compressed the remaining period of time within which the program could be executed, which in turn meant that the implementation team had less time to allow concepts and components to gel among providers and fellow stakeholders. To address this, implementers undertook several sensitization tasks themselves rather than outsourcing this work and conducting it in a more individualized manner as initially intended. For example, rather than being able to conduct a training of trainers – wherein SSDI-PBI program staff would train trainers who would then provide on-the-job training to facility-base providers on the concept and progress of PBI – the condensed timeline compelled SSDI-PBI staff to collapse this activity into district-wide trainings and to lead all trainings themselves. As one implementer said, "Time was simply not on our side with this project. We had to do so much ourselves in the interest of time." Some timing delays were not the result of delayed program start, but rather reflected misestimating the amount of time necessary to undertake a new (oftentimes complex) task. For instance, the implementation team expected that supportive supervision activities (wherein each facility is visited in order to ensure that it is functional in key respects and that it has a functioning health advisory committee) would last four hours. Each supportive supervision visit ultimately lasted a full day (approximately seven hours). Implementers expected to receive all performance declarations and verifications by early March of 2015, but the declarations and verifications trickled in about week later than slated. These delays at the program outset were often attributed to "realities of a new initiative" and challenges among facilities of "trying something new." At later periods, such activities were typically conducted on time and in a more fluid manner.

The most substantive delay that endured throughout program implementation according to several respondents in this fidelity study (and is triangulated with data from the provider motivation component as well as within several community-based FGDs) was the delay of delivery of procured goods to facilities. This is detailed below in the section entitled "Moderating Factors- Program Complexity" below.

Despite delays, interviews within this study component underscore the finding that **the SSDI-PBI program undertook and successfully executed essential PBI program ingredients. In this sense, the program fostered changes in attitudes and behaviors across multiple levels and then benefitted from those changes** during later periods.

Program Adaptations and Modifications

As a pilot program being implemented in a new setting, SSDI-PBI designers and implementers discussed many adaptations across the design and implementation phases. The most substantive design change was a shift in the reward package away from a format that would entail salary bonuses to providers in favor of a package that exclusively entailed rewards in the form of facility-based improvements (namely, equipment and infrastructure). The rationale for this and other changes during the design phase were not the focus of interviews, where instead changes during implementation were emphasized. In that vein, the following key changes occurred:

- Change in verification routines
    - Increases in resources (personnel and vehicles) to conduct quality verifications in order to reduce facilities abilities to warn one another of the approximate timing of a quality verification route ("We need it to be a true surprise")

    - Removal of peer-to-peer verification in favor of verifiers who are part of a facility improvement program but not part of the SSDI-PBI Program. Peer-to-peer verification proved to elicit an unhealthy level of competition across facilities wherein verifiers were intent on punishing one another (**see quote box "On Peer Verification"**).
    - Cancellation of external verification given duplicative findings between facility reports and external verifiers' reports.

> **On Peer Verification**
> It was rubbish. This is something that I can't think about in my life any more. It frustrated me. It hurt me. Peer reviewers from Nkhotakota penalized Chitipa in the verification. And then there was the revenge. Revenging! The independence of verification and assessment teams has to be perfect.- Program Implementer

- Expansion of Implementing Team Staff
    - Hiring of program assistants in each district to assist in management of resources/finances at district level.
    - Hiring of a procurement officer (to facilitate in procurement of equipment and supplies)
    - Hiring of a civil engineer to support infrastructure and related improvements in facilities
- Changes in Scoring System
    - Switch to community scorecards from community-based one-on-one interviews due to the complexity of procuring ethical approval for one-on-one interviews.
- Creation of a Results Declaration and Reward Meeting where representative from facilities came together and data on facility performance was entered in software and projected on a wall. This was done to "enhance transparency" and to create "a bit of healthy competition and assist facilities in learning from one another." In the most recent round of data compilation, this was done via letters instead of meetings.
- Change in Standard Operating Procedure of Implementing Organization
    - Increase expenditure limit for purchases that do not require headquarter approval (from $5,000 to $25,000).

While the implementers adapted the program in several ways in an effort to foster implementation, all respondents mentioned that several adaptations that may have proved beneficial could nevertheless not be made (namely related to procurement). These implementation barriers largely center on a sense that the program cannot be modified or adapted to be more sensitive to modifying in a manner that would bolster autonomy at the facility level and be more sensitive to on-the-ground priorities. This sentiment was perhaps best encapsulated by one district-level provider who said, "… let me tell you, I wanted them to bring us a skeleton. A skeleton and then together we would put on some flesh. Build something together. But they came from Lilongwe and brought their prince. He could not be touched, nothing could be changed or altered". Implementers described how their hands were often tied due to regulations and standard-operating procedures of their organization(s) or of the funder, USAID. Providers described how the nature of indicators and measurement of performance could be unrealistic or inappropriate when essential medical supplies (such as pregnancy tests) were out of stock, or when the indicators themselves were devised without updated input or using "old, bad data". Several providers also described how quality indicators were too rigidly interpreted or enforced. A provider described how the maternity ward was missing bed sheets on one bed during an inspection (the sheets were drying on a clothes line), but because each bed was not covered, the facility was penalized: "It feels like… there is no flexibility... No understanding. If I'm at home, and I want to make a meal I need water, a pot, some fire and some food. These are the major things. This program is penalizing us because we don't have salt. The meal is there but that salt is missing. … I don't need salt to eat a meal."

Moderating Factors- Program Complexity

Across respondents, PBI was described as a well-designed, yet highly complex program. The program was situated within two implementing Non-Governmental Organizations (NGO). Abt Associates largely oversaw

the design of the program, but also worked on issues related to finances and business plans during implementation. Jhpiego largely oversaw the implementation process and worked largely on more clinical and medical aspects, but also oversaw procurement throughout implementation. Both agencies engaged in supervisory visits and provided complementary technical support to one another. Nevertheless, the arrangement of having one NGO oversee design while another oversaw implementation, and having both organizations try to adapt to one another's operating procedures could incite "a bit of awkwardness" as one implementer highlighted.

Along with the SSDI-PBI program extending across two NGOs, the intervention also transcended several dimensions of clinical or medical care. While other health interventions are typically focused on a given domain of health (HIV or childhood immunizations or community-based health promotion), SSDI-PBI extended across these (and other) categories. It also placed demands on other units within implementing organizations such as those devoted to financing, supply chain management, construction and procurement.

In terms of the operationalization of SSDI-PBI in the field, this was described as challenging by most respondents. Several providers and district-level staff mentioned that they had never encountered a program like SSDI-PBI before. Providers and implementers agreed that the concept of drafting business plans and being attuned to data points and trends was conceptually difficult for clinicians to grasp. Implementers described how, "Providers really struggled. Planning is not passed onto lower levels. This initiative gave them a feel of what it means to plan. We had to really encourage them to start planning … to not be waiting for higher level facilities to choose something and to champion something and then for the lower level to just be on the receiving end." One provider described the situation as, "Providers are more inclined to do patient care than to do data management. (When SSDI-PBI asked us to show them our data) … They were asking people to give them something that they literally didn't have, didn't know how to give … People left these (introductory informational meetings) with more questions than answers and didn't really 'get it' until things started happening."

On Meeting Indicators

We've been testing everyone (for HIV). I mean like a lot of people. But none of them are positive. They're all negative! I mean you can't force people to be positive. You don't want to force people to have HIV. But it's also bad when you look through a day of testing and have found nobody. Like NOT ONE. You know what we did to find people the other day? We started mass testing a prison. That's how we finally found people. We will continue to take this approach. However, at some point we are going to run out of prisons and bars and brothels and rest houses.- Health Professional

Along with grasping the essentials of the program (gathering and assessing data, devising business plans and procuring structural improvements), the ability to meet indicators related to quantity and quality was described as conceptually straightforward yet nevertheless difficult to operationalize. In other words, providers understood that they were being scored based on their performance along a series of targets, but actually meeting some of the benchmarks proved untenable. Several respondents described the frustration facilities felt in terms of identifying HIV-positive populations (**see quote box "On Meeting Indicators"**), of encouraging community members to come to antenatal care (ANC) and bring their partners, and of contending with stock-outs of medical supplies that were necessary to meet targets (pregnancy tests, and ferrous sulfate with folic acid (FEFOL) tablets to be used during ANC). Despite these challenges, several respondents also described the benefit of having targets. One community leader described how it gave her a sense of how the facility was progressing and where it needed to improve. A District Health Officer (DHO) described how it gave providers a sense of purpose and piqued facility-wide interest in how data could be used to learn and improve.

The program element that sparked the most vociferous discussion in interviews centered on **procurement**. Challenges related to procuring items that had been earned by facilities, and the delayed delivery of goods

was mentioned in every interview. Implementers highlighted that they underestimated the complexity of procuring goods: "We're not procurement specialists. We didn't even have a procurement officer dedicated to PBI, not even a program *officer*, we thought we would use the systems of Jhpiego. Nobody had seen how big this would turn out -- that we would need a special unit. Now we have an engineer and procurement person *just* for PBI." At the facility level, respondents described how goods (uniforms, generators, cloth wrappers, dustbins) would arrive late, be of substandard quality or not meet required specifications, which was frustrating in itself but was more so problematic because facilities were penalized for goods that were not present – yet had been included in their earliest business plan. Similarly, cash for outreach and meetings as stipulated in the business plans were described as arriving late at the facility, preventing providers from proceeding with planned activities. **The experience of being penalized because an ordered item or cash had not been delivered (and facilities could do nothing to expedite the process) is among the most problematic facets of the SSDI-PBI program according to respondents**. The situation was described as undermining facility and community motivation and autonomy, and breeding resentment and distrust. Implementers and Ministry officials are intimately aware of this problem, but consistently maintained that facilities could not be rewarded for goods that were not present during inspection.

Another program element that is a tenet of PBI programming, but sparked unease among respondents, was the verification process wherein metrics stated in monthly reporting is verified and quality checks are conducted. Respondents described how these activities were initially confusing and, at times, hurtful to those working within facilities being evaluated. One facility in-charge said his staff felt disrespected and unduly criticized by verifiers (this is further detailed in the Provider Motivation section below). Several respondents described how peer-to-peer verification devolved into a vindictive endeavor where teams that felt slighted in a previous round of peer review tried to exact punishment in later rounds.

Moderating Factors- Program Context

Context refers to larger social, political, economic and/or environmental factors that may have influenced implementation. In the case of SSDI-PBI, each interview mentioned at least one facet of context that affected the timing or nature of implementation. Examples of contextual factors include factors inherent to Malawi's health system (high staff turnover, critical shortages in human resources for health, and a largely centralized health system). Social or political factors include sporadic (and sometimes acute) fuel shortages, currency inflation and the fact that SSDI-PBI was implemented immediately after the emergence of "cashgate," a systematic theft of public money by government officials that ultimately upended the sitting government. A final contextual consideration was the existence of another health financing program, called Results Based Financing for Maternal and Newborn Health, that preceded the onset of SSDI-PBI and which (unlike SSDI-PBI) entailed cash bonuses to providers. Results-Based Financing for Maternal Neonatal Health (RBF4MNH) was ongoing and well known by those engaged with the health system, leading many respondents to draw comparisons between SSDI-PBI and RBF4MNH and to highlight that providers in SSDI-PBI facilities could feel slighted or envious of those in RBF4MNH facilities who were perceived as being handsomely compensated for their efforts related to the implementation of the intervention.

Moderating Factors- Program Engagement – Bridging Facilities and Communities

Among the most successful elements of the SSDI-PBI program – in terms of its ability to facilitate implementation – was its emphasis on engagement, communication and participation across stakeholders. Respondents described how the implementation of the program fostered substantive changes in the ways that community members and providers interacted with one another. Via orientations, sensitizations, community outreach (by facility staff to remote villages (**see quote box "On Community-Facility Engagement"**)), collective volunteering

On Community-Facility Engagement

I have lived in this community for 76 years, and it wasn't until 2015 that the people from the facility came to my community. And they showed displays and they talked to us about working together. I am proud. I am so proud of this that the community and the facility are working together. – Community leader

(by community members working within facilities to build fences, clear fields or fundraise for facilities) and community score cards (where communities assess facilities), communities and providers began to more regularly interact. In this process of continual interaction, providers were described as growing more conscientious of the importance of community satisfaction and participation, and communities began to feel a stronger sense of ownership of health facilities (see quote box). As one community leader said, "We learned… the hospital is there because of the people, without them, there is no hospital." As one District Medical Officer (DMO) said, "The interaction with the community is an advantage because if we want things to go well, we can't stay isolated."

In particular, respondents described how a powerful (but initially vexing) program component that strengthened facility-community relationship was the implementation of "score cards," wherein communities were convened to discuss their attitudes and opinions related to care received in facilities, and at a later time point both contingents (facility staff and community members) were brought together to discuss concerns and solutions. Respondents (including facility staff, implementers and community members) described how the creation of a forum for communities to express their views was novel. According to several respondents, the initial experience of bridging communities and facilities, and asking communities to talk about their experiences in facilities led some staff to feel "attacked" and, in an extreme case, to level threats toward community members such as "I will not continue to provide you any service". As time progressed however, score-carding forced providers to more thoughtfully consider community attitudes. As one DHO said, "That (community scorecard) brought in issues. Things which we never knew, (and) hadn't really thought about. … Honestly, there was fighting at first but … as we went on, the collaboration started to make sense to people. Like (community members) would say things like, 'You make us wait a whole day.' That was a thing that many providers had not really thought about." A District Nursing Officer (DNO) described how the score card allowed a facility and community to settle a protracted land dispute, to arrange for the building of a fence and to demarcate space for the construction of a maternity wing.

Communities were also interviewed in the form of focus group discussions (FGD) that centered on program perceptions in relation to community engagement and SSDI's effect on quality of care. These findings are highlighted in the Service Utilization section of the report.

On Inter-Facility Engagement

This program brought knowledge. When (providers across facilities in my district) go out for review meetings, they have been able to learn how best to improve what they're doing at facilities. What's happening in a different facility? What did they do to improve on a particular indicator? How could I learn from that to improve my own facility? Historically, there are really very few forums to talk. What I noted was as time went on, people realized it wasn't a competition. There wasn't one pot of money that we were fighting for. Once that became clear, providers wanted to tell each other "You can improve by doing this, this and that."– DHO

Moderating Factors- Program Engagement – Bridging Within and Across Facilities

Along with emphasizing relationships between communities and facilities, the program also facilitated stronger intra-and inter-facility relationships. Respondents described how the process of creating (and negotiating) business plans strengthened intra-facility cohesion. This was triangulated in interviews with providers (see Section 3 on Provider Motivation). In terms of inter-facility relationships, respondents described how routine meetings, performance declaration meetings (and the attendant performance comparisons across facilities) forced providers to interact across facilities within a given district **(see quote box "On Inter-Facility Engagement")** and ultimately to help one another troubleshoot in the event of addressing difficult indicators or garnering more community support. In two instances, the data collection team encountered facilities that were in the midst of inter-facility knowledge exchanges wherein providers (often accompanied by community leaders) would visit other facilities to share and compare PBI experiences. These ad hoc exchanges were not an explicit component or target of the SSDI-PBI program and were not initiated, coordinated or led by the

implementing agency. Nevertheless, providers viewed the exchanges as a means to galvanize buy-in and to support struggling facilities to better understand the purpose and process of PBI.

Sustainability

A majority of respondents who discussed sustainability mentioned that the program was viewed as prohibitively expensive in the sense that the Ministry would not be capable of continuing such a program without external support. As one ministry official said, "The government is not ready to support these 17 facilities. Or we could support them but then we would need support ourselves in terms of resources." Other ministry staff discussed how the program has compelled the ministry to consider how they could reshape aspects of the health system - including making health centers cost centers. In terms of program design and sustainability, several respondents drew comparisons with the RBF4MNH program and noted that SSDI-PBI was more sustainable in the sense that infrastructure improvements are more enduring than salary top-ups (**see quote box on "Enduring Facets of SSDI-PBI").**

On Enduring Facets of SSDI-PBI

But what do you think will happen when that program goes away? ... The End! ... Have you seen Malawi? Do you think our ministry can keep that kind of money continuing? Now compare the Balaka (RBF4MNH) program with this. Let us say this program ends tomorrow. ... Our activities can continue. The program may end but nobody is going to come and carry away the BP machine, our digital data entry screens. No no no. And nobody will take down our standards guidelines. They are glued to our walls. - PBI coordinator

Key recommendations made across respondents in order to strengthen potential sustainability of the program include:

- Foster a strengthened relationship with the Ministry of Health, and engage more officials (ministry of health and elected government officials) in program sensitization. At the national ministry level appeal to officials who are highly placed ("who know how to sway support, move funds"), and consider nesting an SSDI-PBI employee within the ministry. At the zonal level, draw in and sensitize ministry officials such as Zonal Health Supervisors. In terms of elected officials, reach out to Districts Commissioners (DC) ("who would like to tout the progress made in their district's facilities ... provided they know about it")
- Enhance autonomy at the facility level by allowing facilities to manage funds or be more intimately engaged in procurement
- Sensitize a larger audience within facilities and communities ("we relied too much on talking with directors and leaders and thinking there would be trickle down. Trickle down did not happen")

Discussion

In an overarching sense, a majority of respondents within each respondent group are conversant with the program, and could critically reflect on its strengths and weaknesses. On the whole, views of the program are positive. The sharpest criticism of the program is that it is too heavily concentrated within the implementing team. District level health staff, and representatives within the Ministry of Health described how the program is not instilling a sense of meaningful ownership. Yet in the same conversations where this critique is leveled, respondents appear equally inclined to say that their dissatisfaction is linked to high expectations, intimate knowledge of (and sensitization on) the program, a desire to see the program continue and a longing to feel more engaged in that continuation. In several respects, the program was able to adapt and adjust amid challenges. Moving forward, most respondents agree that it would be beneficial for the program to **bolster autonomy for facilities**, namely via an improved procurement process. If the ability to procure cannot be directly granted to facilities, the amount of time between procurement of a good and arrival of the good to a facility needs to be reduced. Other measures to foster autonomy could also be considered and reductions (or removal) of penalties against facilities that have ordered though not received goods or cash merits consideration.

Implementers at the district level suggested that the PBI program do more to **engage local and district level government officials**, who would not only appreciate the tangible benefits of the program but could

also provide added credence to the program and potentially bolster its profile. Ministry members described a need for **higher-ranking staff within the ministry to be better sensitized on the program** as a means to reinforce program sustainability and better navigate how, where and how much to "push for the program to continue". Roughly half of respondents at facility, district, implementer and ministry levels highlighted that it would be beneficial to **financially compensate providers directly**. Ministry staff clarified that this particular adjustment would not be sustainable in the long-term.

Challenges or Modifications in Undertaking This Study Component

The main challenge in implementing this component was the difficulty of scheduling interviews at a time that was convenient for both the interviewer and respondent. In the case of the counter-verification team and SSDI-PBI coordinators in two regions, a convenient time could not be arranged and therefore while it was initially envisioned to collect data from these respondent types it was ultimately not possible.

Study Component 2a. Service Utilization – Quantitative Findings

Performance-based incentives (PBI) are commonly used as a financial measure to improve health service provision, either by targeting service outputs (e.g. number of patients seen) or service content or quality (e.g. number of patients treated according to protocol, availability of essential service inputs). Effects of PBI programs on service provision differ depending on initial service performance, health worker and facility capacities, PBI design and reward size, and other contextual factors.

Methods

Design and sampling

The SSDI Performance-Based Incentive (PBI) program officially started in August 2014 and included 17 facilities (12 health centers, five hospitals) across three districts (Chitipa, Nkhotakota, Mangochi) to improve the quality of service provision related to reproductive and child health services. To assess the effect of the program on incentivized facilities, we used a pre-test-post-test design with independent controls. We selected 17 control facilities (12 health centers, 5 hospitals) across 8 districts (Chitipa, Nkhotakota, Mangochi, Karonga, Kasungu, Salima, Lilongwe, and Zomba). Control facilities were selected to reflect similar characteristics as intervention facilities, all being public, selected from the same district when possible and from similar geographic areas (near or far from a main road).

To assess different quality dimensions of service provision, we used routine data as well as primary data sources. Where monthly data were available, we used interrupted time series analysis; where one baseline and one endline measurement were available, we used difference-in-differences analysis to estimate the effect of the intervention. Eleven of 13 incentivized indicators were routinely collected on a monthly basis and available through the District Health Information Software 2 (DHIS2) database. These eleven indicators were:

- Number of pregnant women starting antenatal care during the 1st trimester
- Number of women completing the 4 ANC visits
- Number of pregnant women receiving at least 2 doses of IPT
- Number of births attended by skilled birth attendants
- Number of 1 year old children who are fully immunized
- Number of HIV-positive pregnant women who were initiated on ART
- Number of children receiving Vitamin A supplementation
- Number of clients counseled for FP
- Number of couples tested for HIV during HTC services
- Number of women who receive PNC by skilled HCWs within 2 weeks
- Number of pregnant women attending ANC receiving iron supplementation

The two indicators for which data were not collected or too scant, respectively, were:

- Number of HIV/AIDS cases screened for Tuberculosis (TB)
- Number of infants born by HIV positive mothers tested for HIV

Additionally, 11 additional indicators that were not directly incentivized were selected as quantity indicators from the DHIS2 database. Six of these were available for analysis, namely:

- BCG vaccine coverage rate
- Pentavalent III vaccination coverage rate
- Polio-III vaccination coverage rate
- Total number of HIV-tested pregnant females
- Total number of HIV-tested males
- Total number of HIV-tested non-pregnant females

The other five of the 11 additional indicators were missing too much data to be analyzed, and included:

- Measles vaccination coverage rate
- Age under-1 year old fully immunized children – outreach
- Age over-1 year old fully immunized children – outreach
- Number of postnatal mothers supplemented – outreach
- Proportion of all deliveries that are facility-based

Where one baseline and one endline measurement were available, we used difference-in-differences analysis to estimate the effect of the intervention. All 10 quality indicators were of this type, and included:

- Proportion of facilities that reported receiving external supervision within past 6 months
- Proportion of facilities that reported having management meetings at least every 6 months
- Proportion of facilities with client feedback system in place
- Proportion of facilities with SP available at facility
- Proportion of facilities with iron supplements available at ANC service site
- Proportion of facilities with injectable FP methods available
- Proportion of facilities with oxytocin available in maternity unit
- Proportion of facilities with delivery packs available at maternity unit
- Proportion of facilities with partograph forms available at maternity unit
- Proportion of facilities with rapid HIV tests available

Unfortunately, these quality of care indicators had limited utility in revealing any changes that might have occurred in care quality due to high performance rates at baseline and endline among intervention and control facilities.

Training

Data enumerators were trained for five days (including pilot) in March 2016. Data collectors possessed clinical backgrounds and had experience collecting quantitative data for health programming.

Data Collection

Table 5. Quantitative data sources for service utilization study component

Type of Analysis and Data Source	Number of Indicators
Time Series Analysis Data Sources	
HMIS	11
PEPFAR	6
Difference-in-Differences Analysis Data Sources	
SPA (at baseline)	10
Primary data collection	10

Primary data were collected in March 2016. Secondary data were collected beginning in the Fall of 2015 through May 2016. Given the heavy reliance on secondary data, the research team sought to draw from multiple sources of secondary data in order to confirm or refute patterns across sources. This was possible, to a large extent, in relation to the service utilization data in the sense that Health Management Information Systems (HMIS) data could be crosschecked with data stemming from the Presidents Emergency Plan for AIDS Relief (PEPFAR) data. For the quality-related data, the research team intended to draw from three main sources: Service Provision Assessment (SPA) data collected in 2014, primary data that could replicate the SPA and serve as a follow-up assessment in 2016 and Jhpiego's own (SBM-R) data. Ultimately, due to data quality concerns, the SBM-R data could not be used for analysis. See Table 5 for a list of data sources.

Data Analysis

Where monthly data were available, we used interrupted time series analysis (detailed below); where one baseline and one endline measurement were available, we used difference-in-differences analysis to estimate the effect of the intervention (detailed below). Time series analysis allowed us to compare how individual facilities performed in the intervention period relative to their own baseline performance, and revealed when PBI facilities experienced a change in level (a leap in performance at program outset) and/or a change in trend (gradual improvement over time) compared to control facilities. Difference-in-differences analysis allowed us to compare what proportion of facilities were performing to a certain standard at baseline and post-intervention, and to compare whether changes in PBI and control facilities were measurably different.

Interrupted Time Series Analysis (ITS): ITS analyses with controls were used to assess the effects of PBI on intervention facilities. Facilities were assessed relative to their own baseline non-zero performances.

Baseline period: 12 months of data prior to the intervention were used as baseline performance data on each indicator. For each facility, we took the mean value across those 12 months with nonzero values as the average baseline performance estimate. We excluded any zero values from baseline performance calculations for two primary reasons: 1) it seemed likely that some zeroes in the data should have been missing values, and 2) excluding months with zero values from the baseline allowed us to compare monthly performance to typical baseline performance when the facility was actively providing the evaluated services, making the assessment of effects more conservative.

Intervention period: The PBI intervention started in August 2014 and is since ongoing. Our study assessed the effect of the intervention over an 18-month period until January 2016. Using ITS we were able to estimate not only the overall effect of the PBI on service provision, but were also able to differentiate between immediate effects vs. longer-term trends across the 18-month study period.

Monthly performance estimation: For every month in the baseline and post-intervention period, the given month's value was divided by the average baseline performance to render a ratio indicating how well the facility was performing that month against its nonzero baseline average. A value of 1 indicates that the month's performance is on par with the baseline average; a value above 1 means the month's performance is relatively improved compared to the baseline average;

and a value below 1 means the month's performance is relatively decreased compared to the baseline average.

Assessment of performance change: This method of assessing performance compares each facility against its own prior performance. Doubling of performance is given equal credit regardless of whether doubling means an increase in performance from 10 to 20 per month or from 100 to 200 per month. In both scenarios, we report an improvement of 100 %-points ("percentage points") in our results.

Effect estimation: The ITS method also allows a comparison of intervention facilities against a counterfactual, (an alternate future expected to have occurred if the intervention had not been in place). This counterfactual was modeled as a continuation of observed baseline trends in intervention facilities plus the changes observed among controls once the intervention was in place. To estimate the effect of the intervention, we calculated positive or negative changes in the performance of intervention facilities after intervention start compared to performance changes in controls. Significant differences resulting from this comparison were then statistically attributed as intervention effect.

To allow for a more reader-friendly presentation of our findings, we frequently converted percentage points to absolute counts in the following results sections. The reader, however, should be aware that the underlying statistical models assessed the effect based on proportional changes at the facility level, not in absolute counts.

Difference in Difference Analysis (DID): DID analysis was used to assess the difference between how intervention and control facilities changed from one baseline measurement to one follow-up measurement. The resulting differences between intervention and control facilities over time allowed an estimation of the overall intervention effect.

Results

The SSDI-PBI program incentivized performance indicators across a range of health services. Presented below are results arranged by service area including: antenatal care (ANC), labor and delivery, postnatal care (PNC) services, child health and HIV services (including PMTCT). Thirteen services were directly incentivized by performance-specific SSDI-PBI indicators. In an overarching sense, this evaluation found that most of these incentivized indicators had a significant positive impact in health centers' performance, but few performance changes were detected in hospitals (see Table 6). In relation to indicators not directly incentivized, we found positive effects in hospitals' and health centers' performance related to the number of people (males, non-pregnant females, and pregnant females) tested for HIV but few significant effects otherwise (see Table 7). In general, we see overall positive effects on MNH service performance, generally positive effect on HIV services performance, and neutral/no/negative effects on performance related to family planning and child immunization. Because we noticed differences in the performance toward indicators comparing health centers with hospitals (with the former generally posting more sustained or sizable improvements compared to the latter), we have sought to distinguish findings along this dimension as much as possible.

Table 6. PBI effects on Performance Indicators incentivized by the intervention

For each indicator, health center results are presented in white rows, hospital results in gray rows. Indicators 7 and 11 could not be analyzed due to Indicator 7 data not being collected and Indicator 11 data being missing for the majority of facilities and time points.
Stars (*) mark significant changes in intervention group compared to control from the baseline to the intervention period. Daggers (†) mark significant difference in slopes during the intervention period.
* or † p < 0.05, ** or †† p < 0.01, *** or ††† p < 0.001

Incentivized Indicator	Baseline non-zero monthly average	Immediate effect attributable to the intervention (percentage points)	Change in long-term monthly trend attributable to the intervention (percentage points)
1. Number of pregnant women starting antenatal care (ANC) during the 1st trimester	PBI: 8 Con: 6 p-value: 0.002	21	16 * †††
	PBI: 22 Con: 22 p-value: 0.973	11	3
2. Number of women completing the 4 ANC visits	PBI: 19 Con: 12 p-value: <0.001	53 ***	6 *
	PBI: 38 Con: 44 p-value: 0.209	Aug 2014: 75*** Feb 2015: -87**	13 -8
3. Number of pregnant women receiving at least 2 doses of IPT	PBI: 63 Con: 45 p-value: <0.001	30 *	2 *
	PBI: 142 Con: 174 p-value: 0.095	20	5*** †††
4. Number of births attended by skilled birth attendants	PBI: 63 Con: 40 p-value: <0.001	Aug '14: 14 * Feb '15: 1	- 4 ** ††† 5 **
	PBI: 299 Con: 353 p-value: 0.155	-2	0
5. Number of 1 year old children who are fully immunized	PBI: 84 Con: 54 p-value: <0.001	28	2
	PBI: 146 Con: 211 p-value: <0.001	-30*	-3
6. Number of HIV-positive pregnant women who were initiated on ART	PBI: 3 Con: 3 p-value: 0.628	49*	0
	PBI: 8 Con: 9 p-value: 0.581	Aug '14: 37 Aug '15: 41	7 -13**††
7. Number of HIV/AIDS cases screened for TB			
8. Number of children receiving Vitamin A supplementation (reported using non-zero average of 3-month periods and slopes over 3-month periods rather than monthly periods)	PBI: 75 Con: 226 p-value: 0.526	90 *	47
	PBI: 140 Con: 1835 p-value: 0.014	3	53 **††
9. Number of clients counseled for FP	PBI: 386	- 40	- 1 †

Incentivized Indicator	Baseline non-zero monthly average	Immediate effect attributable to the intervention (percentage points)	Change in long-term monthly trend attributable to the intervention (percentage points)
	Con: 146 p-value: 0.052		
	PBI: 560 Con: 619 p-value: 0.500	Aug '14: -30 Feb '15: -15 Aug '15: 102	3 -15 15
10. Number of couples tested for HIV during HTC services	PBI: 30 Con: 28 p-value: 0.810	6	23 * †††
	PBI: 57 Con: 176 p-value: <0.001	Aug '14: -21 Feb '15: 193**	-18 † 36** †
11. Number of infants born by HIV positive mothers tested for HIV			
12. Number of women who receive PNC by skilled HCWs within 2 weeks	PBI: 54 Con: 31 p-value: <0.001	28	-3 †††
	PBI: 82 Con: 150 p-value: 0.013	-24	8* ††††
13. Number of pregnant women attending ANC receiving iron supplementation	PBI: 34 Con: 25 p-value: 0.017	Feb '15: 19 Aug '15: -46	61 *** ††† -109 *** †††
	PBI: 78 Con: 93 p-value: 0.172	Feb'15: -9 Aug '15: 57	-9 6

Table 7. PBI effects on additional indicators not directly incentivized by the intervention

For each indicator, health center results are presented in white rows, hospital results in gray rows.
Indicators 2, 8, 9, 10, and 11 could not be analyzed due to missing data.
Stars (*) mark significant changes in intervention group compared to control from the baseline to the intervention period. Daggers (†) mark significant difference in slopes during the intervention period.
* or † p < 0.05, ** or †† p < 0.01, *** or ††† p < 0.001

Non-incentivized Indicator	Baseline non-zero monthly average	Immediate effect attributable to the intervention (percentage points)	Change in long-term monthly trend attributable to the intervention (percentage points)
1. BCG vaccine coverage rate	PBI: 94 Con: 55 p-value: <0.001	19	2
	PBI: 278 Con: 347 p-value: 0.031	-30*	0†
2. Measles vaccination coverage rate			
3. Pentavalent III vaccination coverage rate	PBI: 90 Con: 60 p-value: <0.001	24 *	1
	PBI: 162 Con: 203 p-value: 0.018	-40**	-1
4. Polio-III vaccination coverage rate	PBI: 89 Con: 58 p-value: <0.001	31	0
	PBI: 159 Con: 191 p-value: 0.058	-42	-1
5. Total number of HIV-tested pregnant females	PBI: 85 Con: 64 p-value: 0.011	32	-2
	PBI: 196 Con: 366 p-value: <0.001	40	13** †
6. Total number of HIV-tested males	PBI: 34 Con: 38 p-value: 0.543	-16	10 ** †††
	PBI: 109 Con: 437 p-value: <0.001	-55	15*** †††
7. Total number of HIV-tested non-pregnant females	PBI: 35 Con: 42 p-value: 0.215	-26	8 †
	PBI: 129 Con: 362 p-value: <0.001	-28	12*** †††
8. Age under-1 year old fully immunized children - outreach			
9. Age over-1 year old fully immunized children - outreach			
10. Number of postnatal mothers supplemented - outreach			
11. Proportion of all deliveries that are facility-based			

PBI effects on antenatal care (ANC) service provision

In respect to ANC service performance, the PBI incentivized facilities on improving both service outputs and clinical content. In terms of **ANC service outputs**, the following performance indicators are used: a) the *number of pregnant women starting ANC during their first trimester of pregnancy* (Indicator 1 in Table 6), and b) the *number of women completing at least 4 ANC visits during their pregnancy* (Indicator 2 in Table 6). In terms **ANC service quality** the following performance indicators are used: a) the *number of pregnant women attending ANC services receiving at least 2 doses of IPT during the course of their pregnancy* (Indicator 3 in Table 6), and b) the *number of pregnant women attending ANC services receiving iron supplementation* (Indicator 13 in Table 6).

ANC service outputs

For health centers, we found significant positive effects of the PBI on both the number of first trimester ANC visits and the number of total ANC visits during a pregnancy. For hospitals, significant effects in were only observed for indicator 2.

Indicator 1: Number of pregnant women starting ANC during their first trimester of pregnancy

Health centers: Although PBI and control facilities showed similar counts for the number of women starting ANC during their first pregnancy trimester (on average 8 vs. 6 women per months), this difference between intervention and control facilities prior to PBI introduction was statistically significant (p=0.002) due to opposite slope directions. Comparing slopes between baseline and intervention periods across study arms, we estimated an intervention effect of the PBI of 21%-points (percentage points) that was statistically not significant (i.e. this finding may have likely occurred by chance given the underlying data). This positive effect size is mainly due to a stronger decline in the number of first trimester women in controls compared to PBI at the interaction point. However, comparing changes in trends between PBI and control facilities during the intervention period, we found a significant stronger positive upward trend (i.e. improvement) of 16 %-points among PBI sites. For graphical representation of estimated time trends, see Figure 1.

Hospitals: Baseline estimates were identical with an average of 22 women per month starting ANC during their first trimester. Intervention effects at both interaction point and trend changes were positive, but statistically non-significant. For graphical representation of estimated time trends, see Figure 2.

Figure 1. Time trends for health centers based on ITS for Indicator 1: Number of pregnant women starting ANC during their first trimester of pregnancy

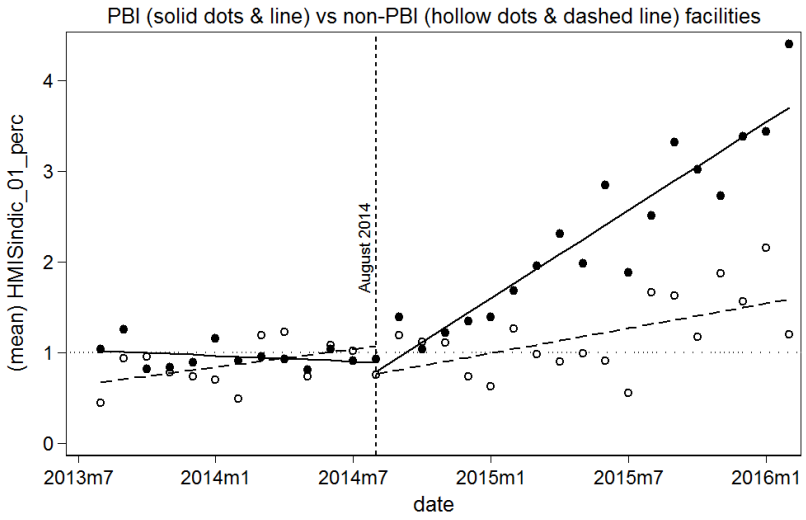
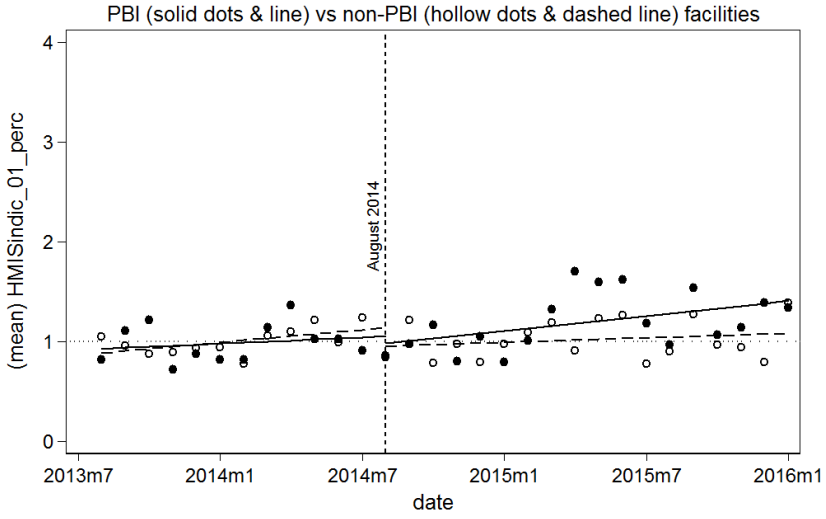


Figure 2. Time trends for hospitals based on ITS for Indicator 1: Number of pregnant women starting ANC during their first trimester of pregnancy



Interpretation: Overall, the PBI positively affected the performance of indicator 1. This effect was statistically stronger among health centers throughout the observed intervention period. However, hospitals had already at baseline a much higher service output of women starting ANC during early pregnancy, which might explain the less pronounced effects of this performance incentive across hospitals. Albeit non-significant, compared to control hospitals, the PBI estimated effect in hospitals is still positive.

Indicator 2: Number of women completing at least 4 ANC visits during their pregnancy

Health centers: PBI and control facilities showed significantly different counts for the number of women completing four or more ANC visits during their pregnancy (on average 19 vs. 12 women per month, p= <0.001) due to opposite slope directions during the baseline period. Comparing slopes between baseline and intervention periods across study arms, we estimated a strongly significant intervention effect of the PBI of 53%-points. The monthly changes during the observed intervention period showed relatively strong monthly improvements for both PBI and control facilities with an overall significant net effect attributable to the PBI of 6%-points. For graphical representation of estimated time trends, see Figure 3.

Hospitals: As with indicator 1, the baseline estimates for hospitals were much higher compared to health centers with an average of 38 (PBI) and 44 (control) women per month completing their fourth or higher ANC visit. At the interaction point, we observed a strongly significant effect of the PBI of 75%-points due to improvements across PBI facilities only. However, while this increased level in this service output indicator was sustained across PBI facilities during the following months, control facilities showed improved performance for this indicator from February 2015 (second payment cycle) onward, which resulted in a significant negative effect (decline in performance) of -87%-points due to the PBI. This negative trend continued for the remaining observation period. For graphical representation of estimated time trends, see Figure 4.

Figure 3. Time trends for health centers based on ITS for Indicator 2: Number of pregnant women completing four or more ANC visits during their pregnancy

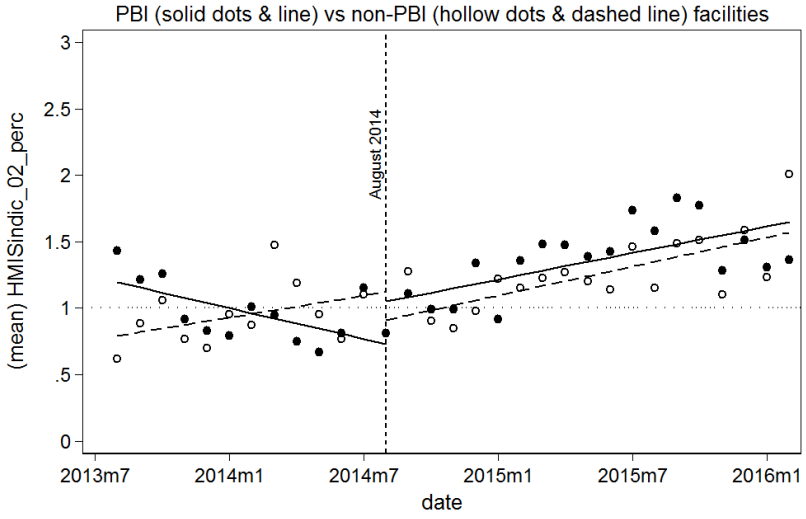
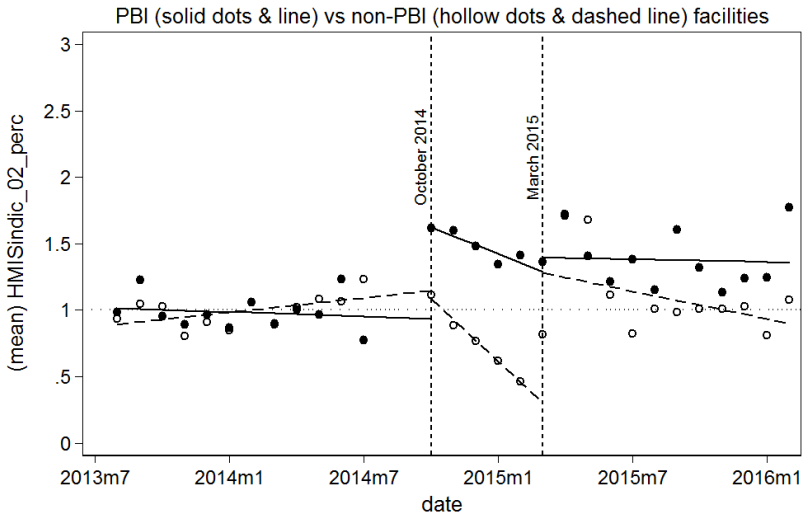


Figure 4. Time trends for hospitals based on ITS for Indicator 2: Number of pregnant women completing four or more ANC visits during their pregnancy



Interpretation: Overall, the PBI had positive effects on health centers' performance of indicator 2. For hospitals this effect was initially positive, but turned negative about six months into the intervention. One possible explanation for this finding might be due to errors in the reporting and recording of HMIS data for this indicator as data points from February 2015 onwards became more scattered. In case the data reflects real events, further investigation on contextual factors affecting performance of this indicator, especially among control hospitals, might allow a better understanding of the observed trends.

PBI effects on ANC service quality

While the PBI had an initial positive effect on number of women who received iron supplementation during their ANC visits across health centers, this effect became significantly negative later on. There was no significant PBI effect observed across hospitals for this indicator. For both health centers and hospitals there the PBI had significant positive effects on the number of women who received at least two doses of IPT during their pregnancy.

Indicator 13: Number of pregnant women receiving iron supplements during ANC visits

To best fit the trends in the data and to account for a national shortage in iron supplements, we assessed effects at February 2015 and August 2015 when there was an incline and decline in performances, respectively.

Health centers: PBI and control facilities showed significantly different baseline counts for the number of women receiving iron supplementation during their pregnancy (on average 34 vs. 25 women per month, p=0.017) due to a declining slope across PBI facilities during the baseline period. No change was observed at the beginning of the intervention (August 2014), but there was a large trend increase beginning in February 2015 (61%-points per month) and then a steep decline beginning in August 2015 (-109%-points per month), both statistically significant. For graphical representation of estimated time trends, see Figure 5.

Hospitals: PBI and control hospitals did not differ significantly in baseline counts (78 in PBI vs 93 in control, p=0.172) nor in effects over the intervention period. For graphical representation of estimated time trends, see Figure 6.

Figure 5. Time trends for health centers based on ITS for Indicator 3: Number of pregnant women receiving iron supplementation during ANC visits

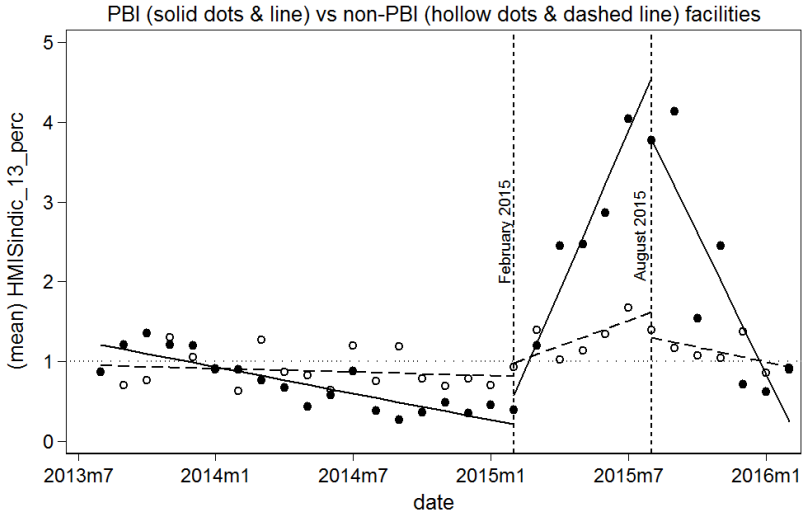
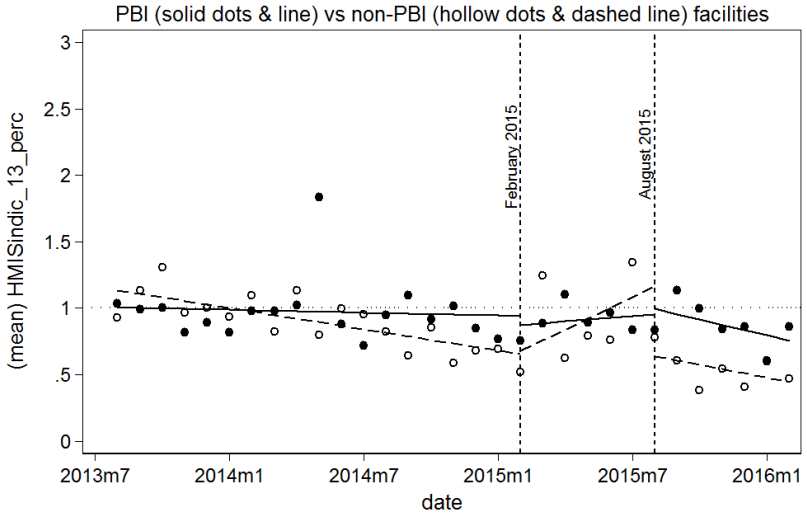


Figure 6. Time trends for hospitals based on ITS for Indicator 3: Number of pregnant women receiving iron supplementation during ANC visits



Interpretation: Overall, the PBI appeared to positively affect the performance of indicator 13 in health centers during the second pay period (February 2015 – August 2015) but the positive effect of the intervention could not overcome the effects of nationwide shortages. Still, based on iron-folate (FeFo) supplement stock-out frequencies measured across facilities, PBI facilities seemed less severely hit by this drug shortage (from 100% FeFo stock availability before to 53% during shortage) than control facilities (from 100% FeFo stock availability before to 18% during shortage). The overall effect of the PBI on FeFo stocks was that 35% fewer PBI facilities experienced a stockout of FeFo compared to control facilities (p=0.028). The overall effect of the PBI on women receiving iron supplementation was negated by plummeting rates through Feb 2016 when both PBI and control facilities were back to an average of 32 and 30 women per month, respectively.

Indicator 3: Number of pregnant women receiving at least two doses of IPT during their pregnancy

Health centers: PBI and control facilities showed significantly different counts for the number of women receiving at least two IPT doses during their pregnancy (on average 63 vs. 45 women per months, p= <0.001) due to opposite slope directions during the baseline period. Comparing slopes between baseline and intervention periods across study arms, we estimated a significant intervention effect of the PBI of 30%-points, given a change in the direction of slopes between baseline and intervention periods across PBI facilities. The monthly changes during the observed intervention period showed ongoing slight monthly improvements for both PBI and control facilities with an overall significant net effect attributable to the PBI of 2%-points. For graphical representation of estimated time trends, see Figure 7.

Hospitals: The baseline estimates of indicator 3 for hospitals were much higher compared to health centers with an average of 142 (PBI) and 174 (control) women per month having received at least two IPT doses. Differences between study groups at baseline were not statistically significant. At the interaction point, we observed a positive effect of the PBI of 20%-points that was statistically not significant. However, while monthly improvements for this service quality indicator continued throughout the observed intervention period, performance across control facilities declined, resulting in a strongly significant net effect attributable to the PBI of 5%-points. For graphical representation of estimated time trends, see Figure 8.

Figure 7. Time trends for health centers based on ITS for Indicator 3: Number of pregnant women receiving at least two doses of IPT during their pregnancy

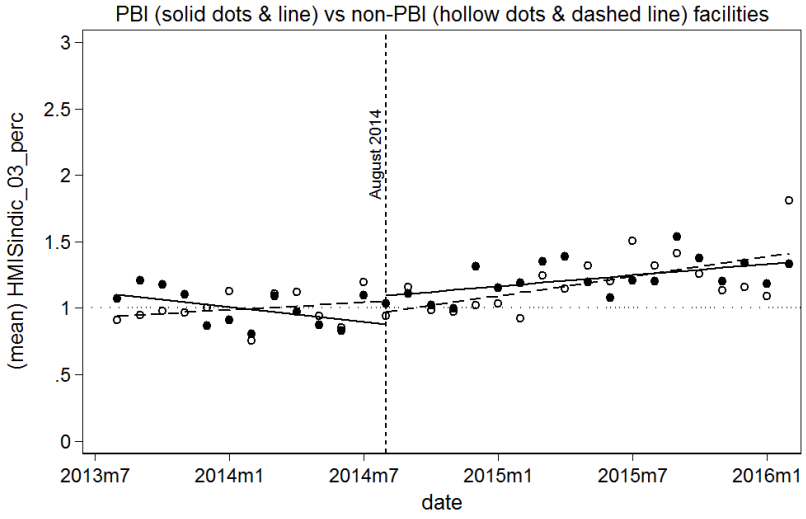
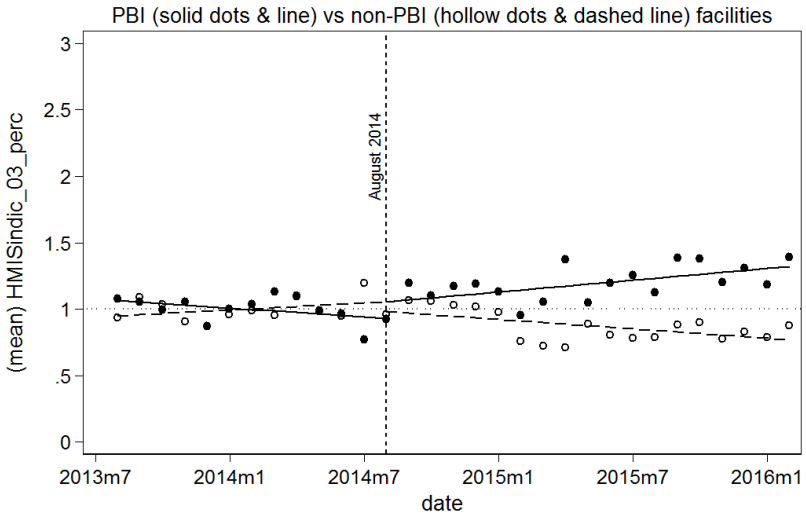


Figure 8. Time trends for hospitals based on ITS for Indicator 3: Number of pregnant women receiving at least two doses of IPT during their pregnancy



Interpretation: Overall, the PBI positively affected the performance of indicator 3. Immediate effects were statistically stronger among health centers compared to hospitals; however, hospitals had already achieved much higher counts for this indicator prior to the intervention. Consequently, the effect of the PBI on hospitals was statistically more pronounced in the monthly incremental changes during the intervention period. One aspect contributing to the greater effect of the PBI on IPT provision compared to iron supplementation (indicator 13) might have been that contrary to iron supplement stock, no stock-outs of Sulfadoxine-Pyrimethamine (SP) occurred during the same observation periods.

PBI effects on the provision of labor and delivery services

In respect to labor and delivery service performance, SSDI-PBI incentivized facilities only in terms of **labor and delivery service quality** by improving the *number of births attended by a SBA* (indicator 4 in Table 6). There was no statistically significant effect of the PBI on the number of delivery cases attended by SBAs at health centers. At hospitals, we found a significant negative effect of the PBI on this indicator.

Indicator 4: Number of births attended by a skilled birth attendant

Health centers: Although PBI and control facilities showed identical trends for this indicator during the baseline period, the average counts for the number of women attended by SBAs (on average 63 vs. 40 women per months, p= <0.001) differed significantly. We also observed a different development of post-intervention slopes from February 2015 (second payment cycle) onwards. Comparing slopes between baseline and early intervention periods (prior to February 2015) across study arms, we estimated a significant intervention effect of the PBI of 14%-points (this effect was only observable after splitting of the intervention period). The monthly changes during the initial intervention period showed an overall statistically significant negative intervention effect of -4%-points due to initial stronger improvements in the control arm. During the later intervention period after February 2015 the monthly trends showed a significant improvement of 5%-points. For graphical representation of estimated time trends, see Figure 9.

Hospitals: The baseline estimates of indicator 4 for hospitals were much higher compared to health centers with an average of 299 (PBI) and 353 (control) births per month attended by SBAs. Differences between study groups at baseline were not statistically significant. At the interaction point, we observed a negative effect of the PBI of -2%-points that was statistically not significant. There were no monthly performance changes detectable during the observed intervention period. For graphical representation of estimated time trends, see Figure 10.

Interpretation: Overall effects of the PBI on skilled birth attendance remained extremely limited and were strongest at the interaction point for health centers, while completely missing for hospitals. One possible reason for this might be that performance of this indicator was already at maximum capacity, meaning that almost all birth occurring at observed facilities were already attended by SBAs prior to the PBI intervention.

Figure 9. Time trends for health centers based on ITS for Indicator 4: Number of births attended by a skilled birth attendant

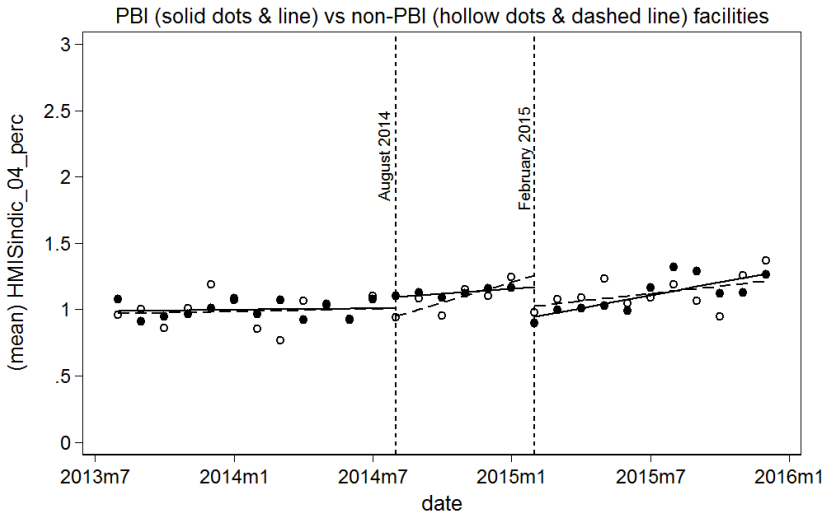
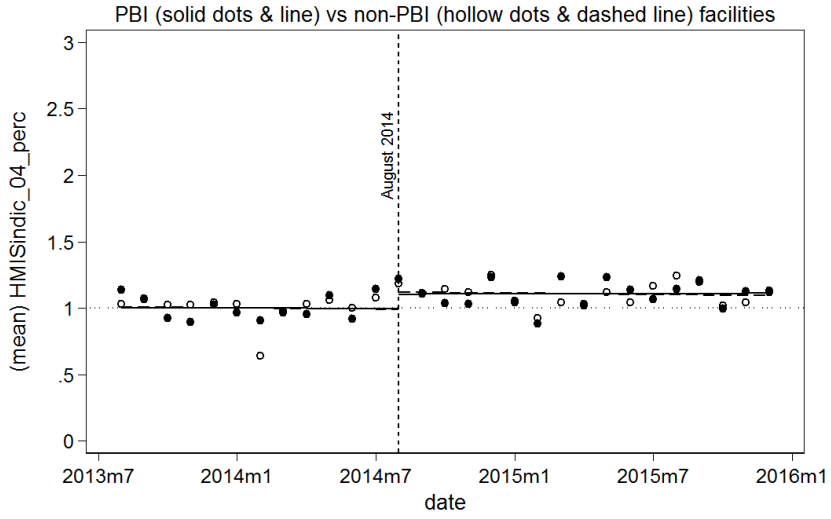


Figure 10. Time trends for hospitals based on ITS for Indicator 4: Number of births attended by a skilled birth attendant



Additional effect measures related to labor and delivery service provision: Since the intervention incentivized labor and delivery service quality only in respect to skilled birth attendance, which did not result in any effective changes, we also explored additional aspects of labor and delivery service quality based on service input measures. Although not directly incentivized by the PBI, we observed the availability of tracer items essential to routine birth attendance, such as partograph forms (essential to labor monitoring), delivery packs (essential to routine birth assistance), and oxytocin (essential to risk reduction of postpartum hemorrhage) in a simple before-and-after difference-in-differences analysis across all studied facilities.

At baseline, partographs were available at 82% of PBI facilities and 100% controls (p=0.013). At endline, all facilities in both study arms reported having blank partographs available, leaving no significant difference at endline (p=1.000). The initial difference was not large enough to result in more than a marginally significant difference-in-differences estimate for the overall PBI effect (p=0.083).

At baseline, delivery kits were available at 94% of PBI facilities and 100% controls (p=0.172). At endline, all facilities in both study arms reported having delivery kits available, leaving no significant difference at endline (p=1.000). The initial difference was not large enough to result in a significant difference-in-differences estimate for the overall PBI effect (p=0.344).
At baseline, oxytocin was available at all facilities in both study arms. At endline, only 88% of PBI facilities had oxytocin available, while 93% of control facilities reported sufficient stocks of this drug (P-0.555). This resulted in an overall negative effect of the PBI of –5% (0.669) that was statistically not significant.
Overall, SSDI-PBI did not appear to have a universally positive or negative spill-over effect on the examined service input items.

PBI effects on the provision of child health services

In respect to child health service performance, SSDI-PBI incentivized **child health service outputs** only, using the following performance indicators: a) the *number of one-year old children fully immunized* (indicator 5), and b) the *number of under-five-year-old children having received vitamin A supplementation* (indicator 8).

While there was no significant effect on the number of one-year old children fully immunized at health centers, we found significant negative effects of SSDI-PBI on the performance of hospitals in respect to this

indicator. We observed significant positive effects of SSDI-PBI on the number of under-five-year-old children having received vitamin A supplementation at both health centers and hospitals.

Indicator 5: Number of one-year-old children fully immunized

Health centers: PBI and control facilities showed significantly different counts for the number of 1-year-olds fully immunized (on average 84 vs. 54 children per months, p= <0.001) due to a positive slope development across control facilities during the baseline period. Comparing slopes between baseline and intervention periods across study arms, we estimated a non-significant intervention effect of the PBI of 28%-points. The monthly net changes during the observed intervention period resulted in a slight non-significant 2%-point increase due to the PBI. For graphical representation of estimated time trends, see Figure 11.

Hospitals: The baseline estimates for indicator 5 for hospitals were much higher compared to health centers with an average of 146 (PBI) and 211 (control) 1-year-olds per month fully immunized. Differences between study groups at baseline were statistically significant. At the interaction point, we observed a significant negative effect of the PBI of -30%-points. However, while monthly improvements for this service quality indicator remained minimal throughout the observed intervention period for both PBI and control facilities, the net effect attributable to the PBI was negative with -3%-points. For graphical representation of estimated time trends, see Figure 12.

Interpretation: Overall effects of the PBI on child immunization remained extremely limited and even lead to a decline in service outputs at hospitals. Based on the interaction effects of 28%-points at health centers and -30%-points at hospitals, a shift may have occurred wherein young children got vaccinated once PBI started (i.e. a shift from hospitals to health centers due to increased immunization outreach activities at heath center level). However, our data cannot further verify this assumption. Another explanation of the low effectiveness of this performance indicator might be that already at baseline maximum capacity for child immunization had been achieved.

Figure 11. Time trends for health centers based on ITS for Indicator 5: Number of 1-year-old children fully immunized

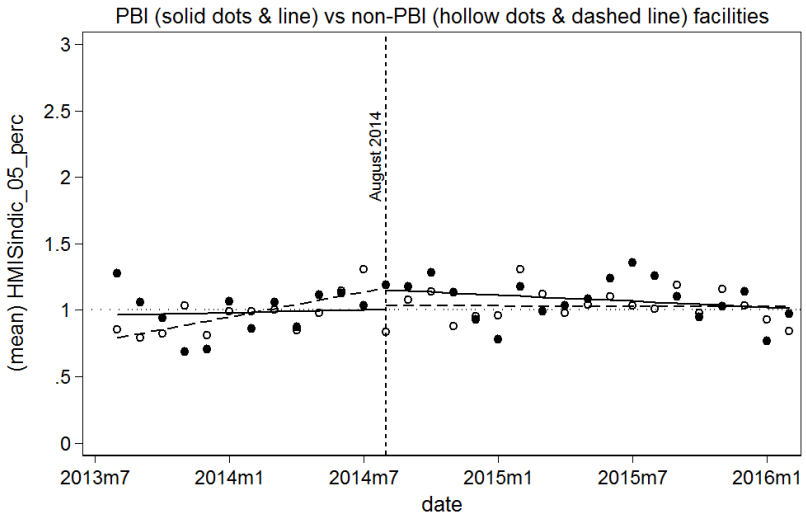
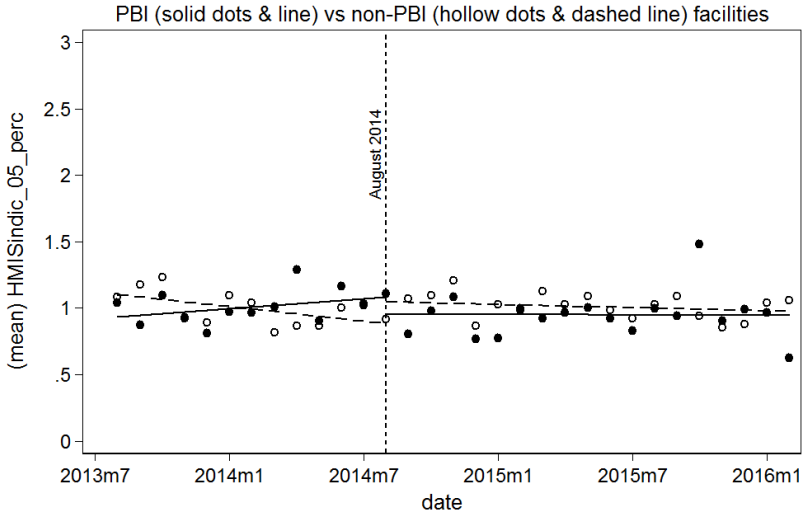


Figure 12. Time trends for hospitals based on ITS for Indicator 5: Number of 1-year-old children fully immunized



Additional effect measures related to the provision of child immunization services: Since the intervention incentivized only child immunization service output only in respect to full immunization of 1-year-olds, which did not result in any effective changes at health centers, we also explored additional aspects of immunization service outputs for single vaccinations, including BCG, pentavalent III, and polio III coverage rates (indicators 1, 2, and 3 in Table 7). Across health centers, we observed positive effects of the PBI on these coverage rates, with even a statistically significant improvement of 24%-points for pentavalent III coverage. Across hospitals, on the other hand, we observed negative effects for these indicators with even significant declines for BCG vaccine (-30%-points) and pentavalent III coverage (-40%-points). While SSDI-PBI induced impressive effects – positive or negative – at the interaction point, it hardly induced additional incremental changes for these indicators during the observed intervention period. This sub-analysis by vaccine type coverage demonstrates that the overall effect measured by indicator 5 in Table 6 has likely occurred across vaccination types and is unlikely due to any extreme trends within a single vaccine, such as a single-item stock-out.

Indicator 8: Number of under-five-year-old children fully immunized

Due to the nature of vitamin A supplementation outreach programs, which frequently and appropriately provide mass doses of Vitamin A once or twice a year, we analyzed the *number of five-year-old children supplemented with vitamin A* using 3-month interval sums instead of monthly sums. This approach prevented extreme variations among facilities given that vitamin A supplementation occurred every few months rather than on a regular basis.

Health centers: PBI and control facilities showed different counts for the number of under-5-year-olds supplemented with vitamin A (on average 75 vs. 226 children per month, p= 0.5). Given the 3-months estimation period, this difference was not statistically significant. Comparing slopes between baseline and intervention periods across study arms, we estimated a significant intervention effect of the PBI of 90%-points. The monthly net changes during the observed intervention period resulted in a non-significant 47%-point increase due to the PBI. For graphical representation of estimated time trends, see Figure 13.

Hospitals: The baseline estimates for indicator 8 for hospitals were much higher compared to health centers with an average of 140 (PBI) and 1835 (control) under-5-year-olds per month supplemented with vitamin A. Differences between study groups at baseline were statistically significant. At the interaction point, we observed only a small non-significant positive effect of the PBI of 3%-points. However, monthly improvements for this service output indicator were significantly large throughout the observed

intervention period for across PBI hospitals with a net effect attributable to the PBI of 53%-points. For graphical representation of estimated time trends, see Figure 14.

Interpretation: Overall effects of SSDI-PBI on vitamin A supplementation was extremely positive. Given the reduced number of data points (due to 3-months periods), these estimates only reflect the actual outreach activities during the Malawian winter months. Calculated across 12-months periods, effect sized would shrink significantly. Nevertheless, this PBI performance indicator improved these outreach activities significantly compared to similar activities at non-intervention sties.

Figure 13. Time trends for health centers based on ITS for Indicator 8: Number of under-5-year-old children supplemented with vitamin A

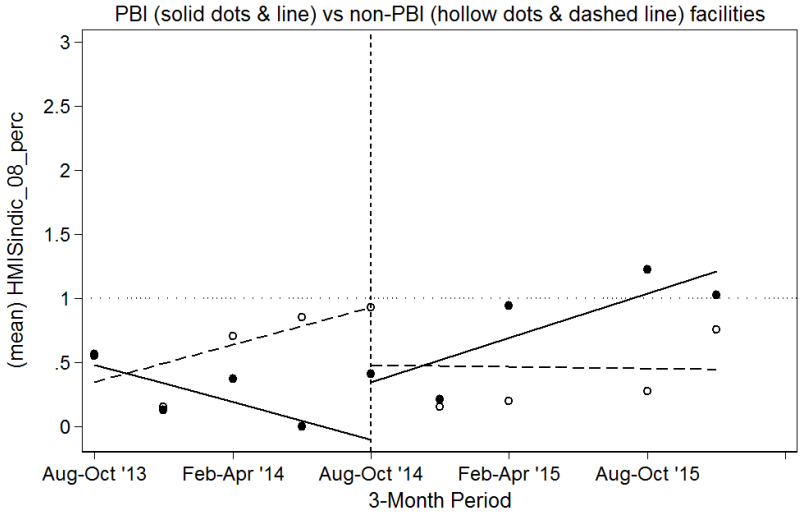
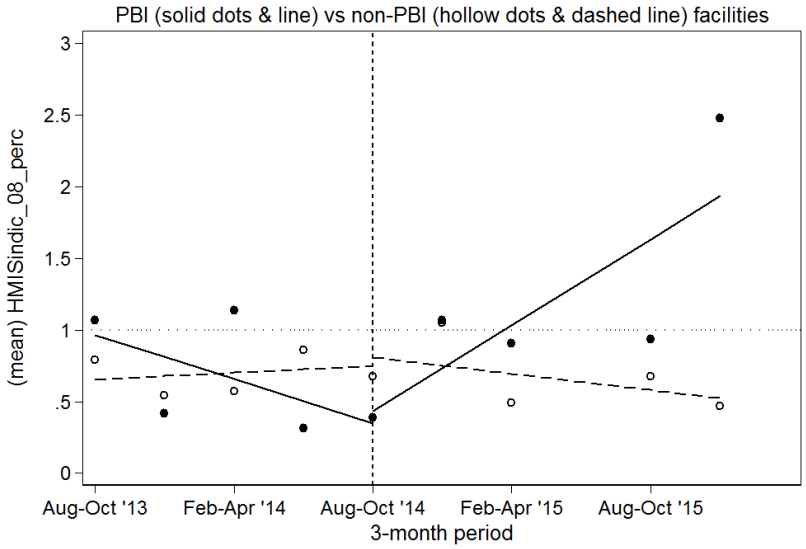


Figure 14. Time trends for hospitals based on ITS for Indicator 8: Number of under-5-year-old children supplemented with vitamin A



PBI effects on the provision of PMTCT services

In respect to service performance of prevention of mother-to-child-transmission of HIV (PMTCT), the PBI incentivized facilities only in terms of **PMTCT service outputs** by improving a) the *number of HIV-positive*

pregnant women initiated on anti-retroviral therapy (ART) (indicator 6 in Table 6) and b) the *number of infants born to HIV positive mothers tested for HIV* (indicator 11 in Table 6).
We observed a significant positive effect on ART coverage of pregnant HIV positive mothers across health centers, but a significant negative effect across hospitals during the late intervention period. We were unable to assess the PBI effect on HIV testing of infants born to HIV positive mothers due to low quality of available data on this indicator (many missing values in the database).

Indicator 6: Number of HIV-positive pregnant women initiated on ART

Health centers: PBI and control facilities showed identical counts for the number of HIV-positive pregnant women initiated on ART (on average 3 women per month, p= 0.6). Comparing slopes between baseline and intervention periods across study arms, we estimated a significant intervention effect of the PBI of 49%-points. The monthly net changes during the observed intervention period showed not further effects. For graphical representation of estimated time trends, see Figure 15.

Hospitals: The baseline estimates for indicator 8 for hospitals were similar with an average of 8 (PBI) and 9 (control) women per month. We split the analysis of the intervention period, as data trends differed from August 2015 (third payment cycle) onwards. At the interaction point, we observed only a non-significant positive effect of the PBI of 37%-points with another non-significant increase of 41% points in August 2015. While incremental improvements between August 2014 and August 2015 remained positive with 7%-points, this trend turned significantly negative afterwards with -13%-points. For graphical representation of estimated time trends, see Figure 16.

Interpretation: Overall effects of the PBI on the pregnant women's initiation on ART was positive during the interaction point for both health centers and hospitals. However, this effect was not long lasting. This flattening of the effect might be due to fact that the number of HIV-positive women enrolled in ART programs has reached its maximum. Given that data points were collected on monthly basis, but enrolment of HIV-positive pregnant women is a rather long-term (pregnancy-spanning) not monthly event and the prevalence of HIV-positive pregnant women might be relatively low even within the Malawi context, the ITS time intervals might not offer the most suitable analytical approach for this indicator.

Figure 15. Time trends for health centers based on ITS for Indicator 8: Number of HIV-positive pregnant women initiated on ART

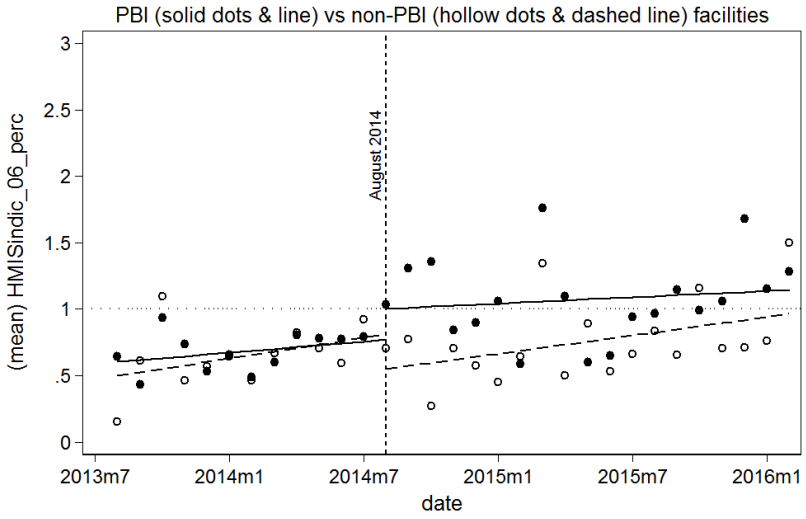
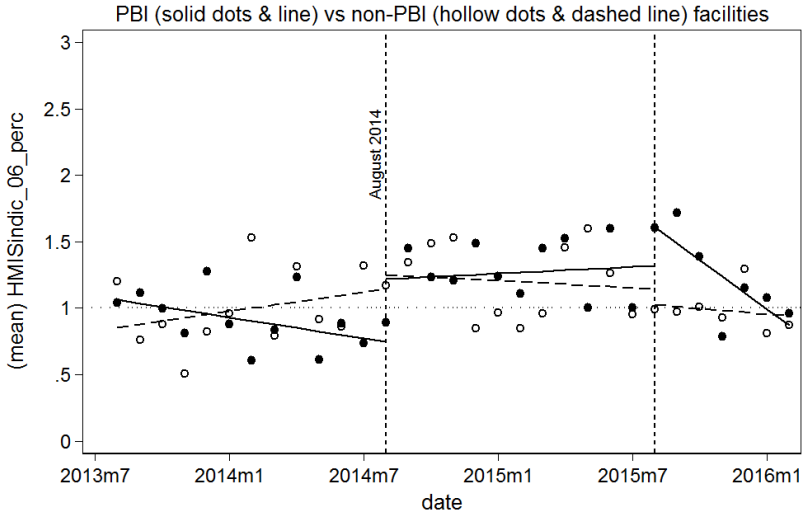


Figure 16. Time trends for health centers based on ITS for Indicator 8: Number of HIV-positive pregnant women initiated on ART



Additional effect measures related to the provision of PMTCT: The intervention incentivized both enrolment of HIV-positive women into ART programs and the testing of babies born to HIV-positive mothers. However, for the latter indicator we could not obtain sufficient data. Instead, we observed whether there were any spillover effects by the PBI on the number of HIV-tested pregnant women (indicator 4 in Table 7) as a PMTCT service output indicator related to indicator 6 in Table 6. Similar to the performance on ART enrolment of HIV-positive pregnant women, we found positive, but non-significant, PBI effects on the number of tested pregnant women across both health centers and hospitals. While this effect flattened as well across health centers, we observed significant incremental improvements of 13%-points during the observed intervention period for hospitals. Given the increase in testing of pregnant women and the decline in enrolment of HIV-positive women in ART programs might also hint at stock-outs of ARV medications, especially at hospitals. Further contextual analysis in respect to PMTCT service inputs may provide a better understanding of limiting factors of the PBI effect on these indicators.

PBI effects on the provision of HIV services

In respect to HIV service performance, SSDI-PBI incentivized facilities on improving **HIV service outputs** by improving a) the *number of couples tested for HIV during testing and counseling sessions (HTC)* (indicator 10 in Table 6) and b) the *number of HIV/AIDS cases screened for tuberculosis (TB)* (indicator 7 in Table 6). We observed significant positive effects on the number of couples tested during HTC. We were unable to assess the PBI effect on TB screening of patients with HIV/AIDS due to complete absence of data for this indicator in available databases.

Indicator 10: Number of couples tested for HIV during testing and counseling sessions (HTC)

Health centers: PBI and control facilities showed very similar trends and average counts (on average 30 vs. 28 couples per months, p= <0.8) for the number of couples counseled during the baseline period. Comparing slopes between baseline and intervention periods across study arms, we estimated a non-significant intervention effect of the PBI of 6%-points. The monthly changes during the intervention period showed an overall statistically significant positive intervention effect of 23%-points. For graphical representation of estimated time trends, see Figure 17.

Hospitals: The baseline estimates of indicator 10 for hospitals were much higher for control sites with an average of 57 (PBI) and 176 (control) tested couples per month. Differences between study groups at baseline were statistically significant. We split the analysis of the intervention period, as data trends differed from February 2015 (second payment cycle) onwards. At the interaction point, we observed a

non-significant negative effect of the PBI of -21%-points, followed by a significant increase of 193%-points in February 2015. While incremental improvements between August 2014 and February 2015 remained significantly negative with 18%-points, this trend turned significantly positive afterwards with 36%-points. For graphical representation of estimated time trends, see Figure 18.

Interpretation: Overall effects of the PBI on HIV testing and counseling of couples were positive and more pronounced across health centers. The relatively stronger improvements of control hospitals in the early intervention period remains unclear and might be due to reporting and recording errors of HMIS data, given that these data points resemble data errors that could result from simple data entry errors during a six-month period, especially since the linear trend prior to this episode seems to continue right after this episode, thus resembling the trends found across health centers.

Figure 17. Time trends for health centers based on ITS for Indicator 8: Number of couples tested for HIV during HTC

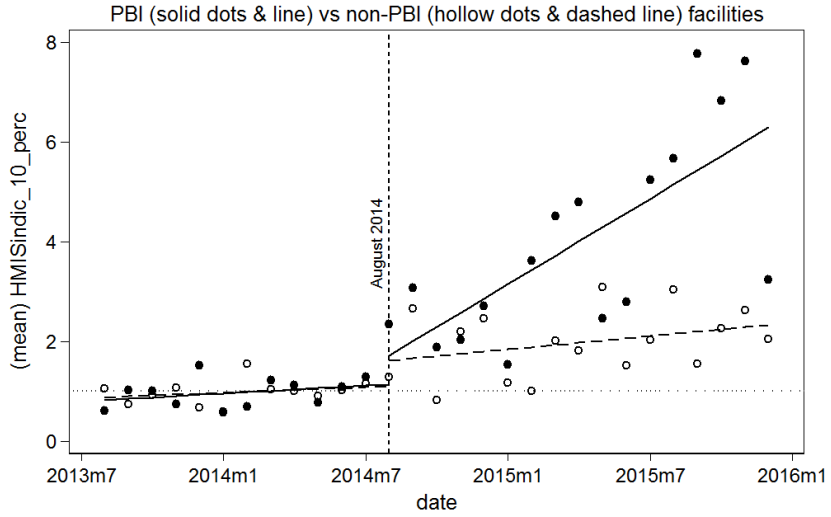
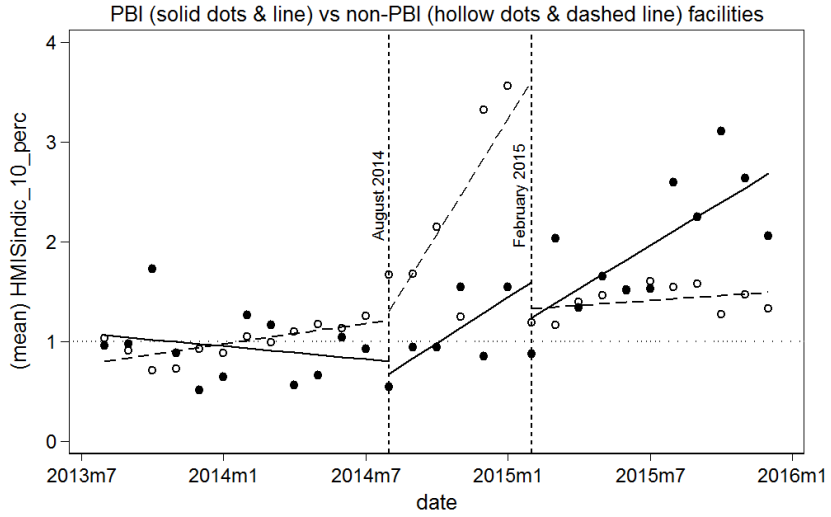


Figure 18. Time trends for hospitals based on ITS for Indicator 8: Number of couples tested for HIV during HTC



Additional effect measures related to the provision of HIV services: The intervention incentivized both couples testing and tuberculosis screening for patients with HIV/AIDS. However, for the latter indicator we could not obtain sufficient data. Instead, we assessed whether there were any spillover effects of SSDI-PBI on the number of HIV-tested males and HIV-tested non-pregnant females (indicators 5 and 6 in Table 7) as

HIV service output indicators related to indicator 10 in Table 6. Contrary to the performance on couples testing, we found negative, but non-significant, PBI effects on both of these indicators across both health centers and hospitals at interaction points. During the intervention phase, we observed statistically significant positive incremental improvements for both indicators across both facility types. This might indicate that initially HIV service performance was distorted towards incentivized service outputs (tested couples) with neglect of other client groups. During the course of the intervention, this distortive effect seemed to normalize.

We further assessed whether the availability of rapid HIV-test had an influence on the above service output performances. Before the intervention, 94% of PBI and 100% of control facilities reported stocks of rapid HIV-tests, while both types of facilities reported 100% stocks of these test at endline, not indicating any significant effect of the PBI on test availability (p=0.306). Unavailability of HIV test therefore is unlikely to explain any of the above service output findings.

PBI effects on the provision of family planning (FP) services

In respect to FP service performance, the PBI incentivized facilities on improving **FP service output** related to the *number of clients counseled for modern family planning methods* (indicator 9 in Table 6).
We observed mainly negative effects of the PBI on this indicator across both facility types.

Indicator 9: Number of clients counseled for modern family planning methods

Health centers: PBI and control facilities showed close to significance different counts for the number of clients counseled for modern FP methods (on average 386 vs. 146 clients per month, p=0.05). Comparing slopes between baseline and intervention periods across study arms, we estimated a negative non-significant intervention effect of the PBI of -40%-points. The monthly net changes during the observed intervention period resulted in a significant negative -1%-point decrease due to the PBI. Although incremental improvements were stronger among PBI faculties, this slight decline is due to a relative higher performance level at control facilities, which PBI facilities only surpassed in the later intervention period. For graphical representation of estimated time trends, see Figure 19.

Hospitals: The baseline estimates for indicator 12 for hospitals were much higher compared to health centers with an average of 560 (PBI) and 619 (control) clients counseled per month. This difference in average counts between study groups at baseline was not statistically significant. We split the analysis of the intervention period, as data trends differed between the periods August 2014 to February 2015 (first payment cycle), February 2015 to August 2015 (second payment cycles), and August 2015 onwards (third payment cycle). At the interaction point, we observed a non-significant negative effect of the PBI of -30%-points with another non-significant decline of -15% points in February 2015. In August 2015, we observed a positive effect of 102%-points. Incremental improvements were positive, but statistically non-significant, except for the period between February and August 2015, which was negative with -15%-points. For graphical representation of estimated time trends, see Figure 20.

Figure 19. Time trends for health centers based on ITS for Indicator 9: Number of clients counseled for modern family planning methods

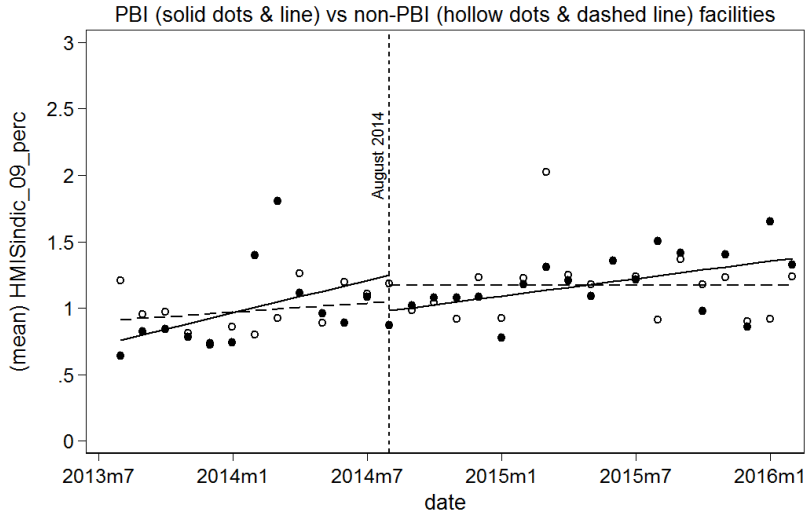
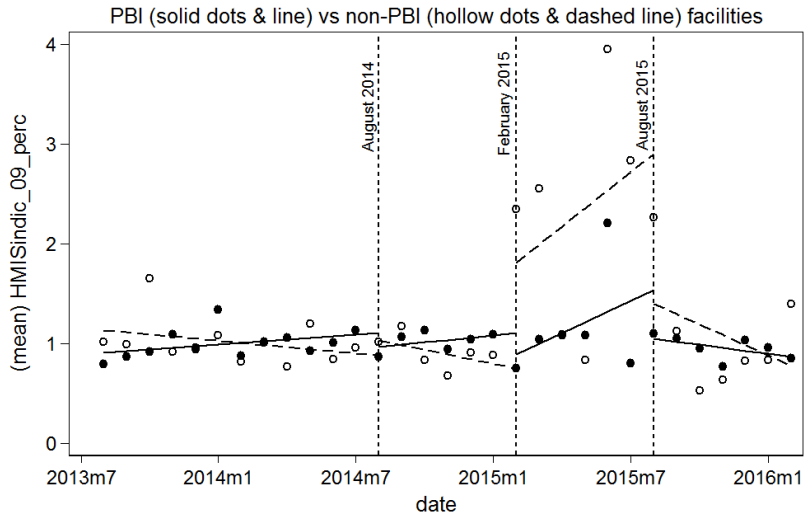


Figure 20. Time trends for hospitals based on ITS for Indicator 9: Number of clients counseled for modern family planning methods



Interpretation: Overall effects of the PBI on this indicator were rather negative, although not statistically significant. The trends observed between February and August 2015 (during the second payment cycle) might be again be due to data reporting and recording error on part of control hospitals. If this is, in fact, representing a real event, further exploration of contextual factors will be needed. We attempted this by assessing the effect of the PBI on service inputs for family planning, such as the availability of injectable FP methods – most commonly used method among Malawian women of reproductive age according to recent Malawi Demographic and Health Survey (DHS) assessment – using simple before-and-after comparison based on difference-in-differences analysis. While the proportions of PBI and control facilities with injectable FP methods available did not differ significantly at baseline (p=0.344) nor endline (p=0.302), the trend across control facilities (92% up to 100%) showed stronger, although not statistically significant, improvements over a relative decline in the proportion of PBI facilities in terms of injectable FP method availability (100% to 91%) (p=0.164). This finding might contribute, but not fully explain the patterns observed by the ITS for the performance of client counseling on modern FP methods.

PBI effects on the provision of postnatal care (PNC) services

In respect to PNC service performance, the PBI incentivized facilities on improving **PNC service outputs** by improving the *number of women who receive PNC by skilled health care workers within two weeks of delivery* (indicator 12 in Table 6)**.**

We observed a positive PBI effect across health centers and a negative effect across hospitals in terms of PNC.

Indicator 12: Number of women receiving postnatal care by a skilled health worker within two weeks of delivery

Health centers: PBI and control facilities showed significantly different counts for the number of women receiving postnatal care within two weeks of birth (on average 54 vs. 31 children per month, p=<0.001). Comparing slopes between baseline and intervention periods across study arms, we estimated a non-significant intervention effect of the PBI of 28%-points. The monthly net changes during the observed intervention period resulted in a significant negative -3%-point decrease due to the PBI. This slight decline is due to more extreme improvements across control facilities starting at a lower performance level, but surpassing PBI facility performance eventually. For graphical representation of estimated time trends, see Figure 21.

Hospitals: The baseline estimates for indicator 12 for hospitals were much higher compared to health centers with an average of 82 (PBI) and 150 (control) women per month. Although both PBI and control hospitals showed similar upward slopes during the baseline period, the differences in average counts between study groups at baseline were statistically significant. At the interaction point, we observed a non-significant negative effect of the PBI of -24%-points. However, monthly improvements for this service output indicator were significantly large throughout the observed intervention period across PBI hospitals with a net effect attributable to the PBI of 8%-points. For graphical representation of estimated time trends, see Figure 22.

Interpretation: Overall effects of the PBI differed between health centers and hospitals. Whereas initially the PBI affected health center performance for this indicator positively, the long-term effect of the PBI on hospital performance was much more pronounced. Again, this might indicate a shift in PNC service utilization by women towards health centers in the initial intervention phase that slightly reverted back to PNC service utilization at hospitals. However, more detailed information will be needed to explain this finings more definitively.

Figure 21. Time trends for health centers based on ITS for Indicator 12: Number of women receiving postnatal care by a skilled health worker within two weeks of birth

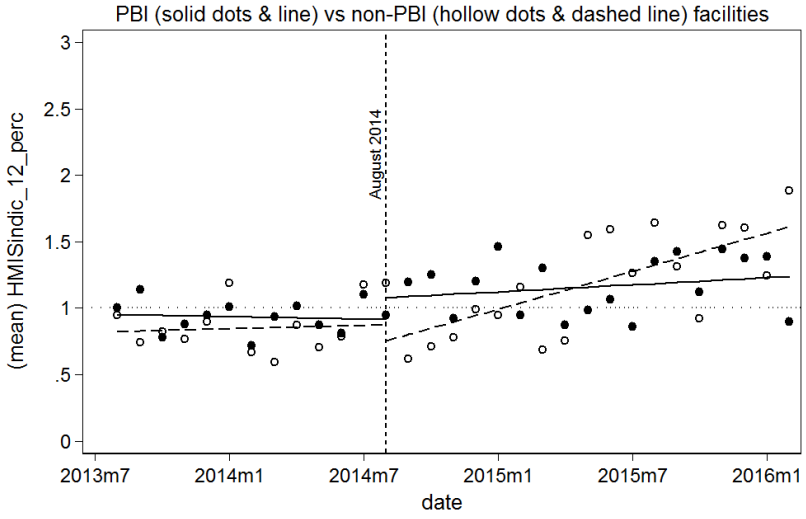
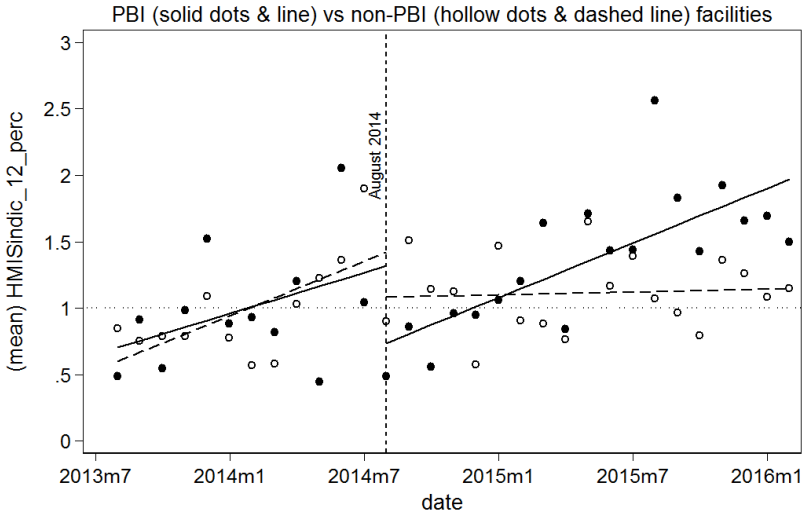


Figure 22. Time trends for hospitals based on ITS for Indicator 12: Number of women receiving postnatal care by a skilled health worker within two weeks of birth



PBI effects on the overall service volume of incentivized health services

To have a more general performance estimate, which allows us to assess the PBI effect on the provision of incentivized services, we created a simple composite score based on eight of the thirteen PBI performance indicators we considered most central to essential health care provision in Malawi. This composite score includes:

ANC services

- number of pregnant women with first ANC visit during first trimester
- number of pregnant women with four or more ANC visits during a pregnancy
- number of women receiving at least two dosed of IPT

Delivery services

- number of births attended by a skilled birth attendant

Child health services

- number of one-year old children fully immunized

PMTCT services

- number of HIV-positive pregnant women initiated on anti-retroviral therapy (ART)

HIV services

- number of couples tested for HIV during testing and counseling sessions (HTC)

PNC services

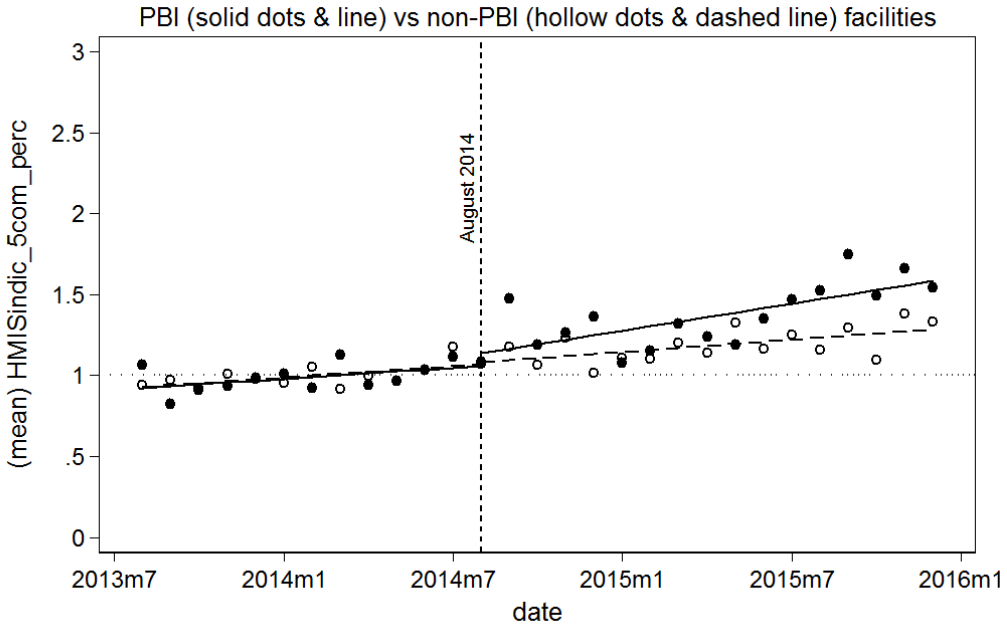
- number of women who receive PNC by skilled health care workers within two weeks of delivery

These eight indicators were summed to estimate overall service performance. For example, in a month when 10 women had their first ANC visit, 4 had their 4th or later ANC visit, 20 had received at least 2 doses of IPT, 3 births were attended by a skilled birth attendant, 20 children were fully immunized, 1 HIV-positive pregnant women was initiated on ART, 3 couples were tested for HIV during HTC, and 4 women received PNC by skilled health workers, then the service volume estimate for that month would be 10 + 4 + 20 + 3 + 20 + 1 + 3 + 4 = 65.

This analysis was performed for the entire sample, i.e. health centers and hospitals combined, in order to provide a general estimate on the overall PBI effect. Figure 23 presents the graphical representation of estimated time trends.

During the baseline period, control facilities had significantly higher average service performance at 559 vs. 361 items per month (p=0.008). We observed an immediate intervention effect of an 11%-point greater increase among PBI facilities compared to a -1%-point increase across controls; however, this immediate effect was not significant (p=0.282). Incremental changes during the intervention period were significantly greater among PBI facilities compared to control facilities (2%-points vs. 0.1 %-point per month, p=0.039). By the end of the observed intervention period, the control facilities had changed very little in terms of service performance, whereas PBI facilities demonstrated more pronounced improvements. Projected PBI facility performance was estimated to be around 612 items per month and control facility performance around 572 items per month. Overall, there was a positive effect of the PBI on service performance that became statistically significant during the later intervention period.

Figure 23. Time trends for entire facility sample based on ITS for composite performance Indicator



PBI effects on the general service management and quality assurance

Although not specifically incentivized, we assessed a set of indicators related to aspects of **service management and general quality assurance**, including: a) the *proportion of facilities that reported receiving external supervision within past 6 months***,** b) the *proportion of facilities that reported having management meetings at least every 6 months*, and c) the *proportion of facilities with client feedback system in place*.

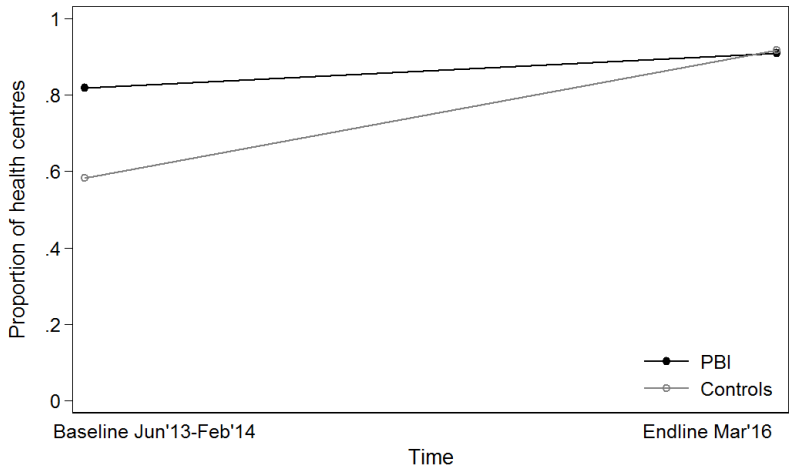
As performance-based programs carry the theoretical potential to improve organizational aspects of service delivery by aligning service management structures with performance outputs, we tried to assess whether and to what extent this was the case with this PBI. This analysis is based on simple before-and-after assessments using a difference-in-differences model across the entire facility sample (i.e. health centers and hospitals combined).

For both external supervision and client feedback we observed non-significant negative intervention effects. For periodic management meetings we did not observe any effects due to the PBI.

Indicator: Proportion of facilities that reported receiving external supervision within past six months

The proportion of control facilities (71% to 94%) improved more than the proportion of PBI facilities (82% to 88%) over the course of the intervention in respect to external supervisions of facility staff and of work processes. This change was not large enough to be significant when comparing before-and-after proportions. The differences at baseline (p=0.356), at endline (p=0.644), and in differences (p=0.328) were not significant. Figure 24 provides a graphical representation of estimated changes between baseline and endline.

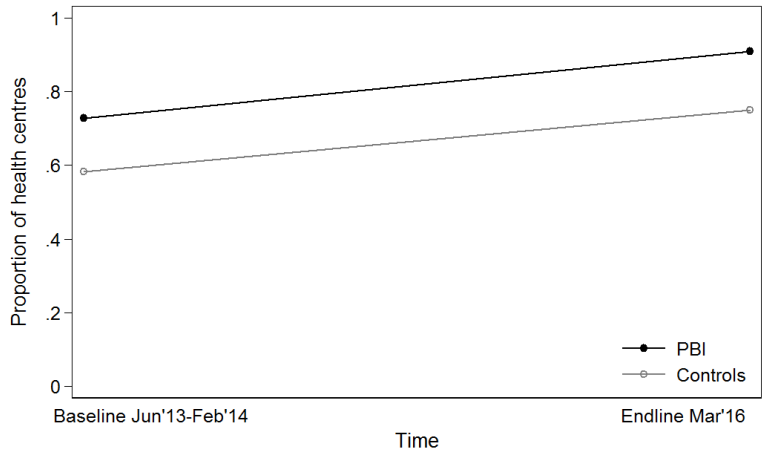
Figure 24. Changes between baseline and endline for entire facility sample based on DiD for proportion of facilities receiving external supervision within past 6 months



Indicator: Proportion of facilities that reported having management meetings at least every six months

Both the proportions of PBI (59% to 92%) and control facilities (47% to 80%) showed parallel upward trends in the proportion of facilities holding regular management meetings from baseline to endline. The differences at baseline (p=0.448), at endline (p=0.472), and in differences (p=0.981) were not statistically significant. Figure 25 provides a graphical representation of estimated changes between baseline and endline.

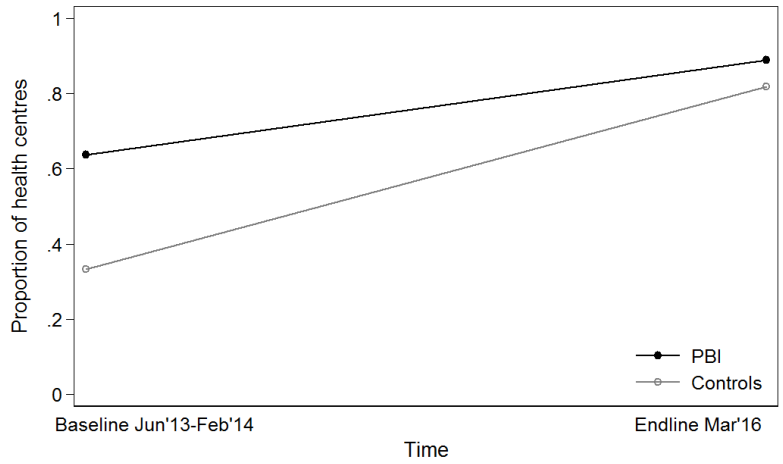
Figure 25. Changes between baseline and endline for entire facility sample based on DiD for proportion of facilities with reported management meetings at least every 6 months



Indicator: Proportion of facilities with client feedback system in place

The proportion of control facilities (71% to 82%) improved more than the proportion of PBI facilities (82% to 88%) over the course of the intervention in respect to the proportion of facilities with client feedback systems in place. This change was not large enough to be significant when comparing before-and-after proportions. The differences at baseline (p=0.398), at endline (p=0.715), and in differences (p=0.738) were not significant. Figure 26 provides a graphical representation of estimated changes between baseline and endline.

Figure 26. Changes between baseline and endline for entire facility sample based on DiD for proportion of facilities with client feedback systems in place



Discussion

Generally, service provision quality at health centers seemed to have improved more than at hospitals that partook in the SSDI-PBI program, which suggests that the intervention design may have targeted service performance issues that were more pertinent to first-level care. The SSDI-PBI approach of setting clear performance targets and the influx of additional financial income likely allowed health centers to better identify and prioritize those aspects of work performances and overall facility management that ensured higher service quality yields. It might therefore be useful to follow more differentiated understanding of service provision, possibly by creating different targets or indicators for hospitals versus health centers.

In terms of SSDI-PBI's effects by health service type, we noted overall positive effects on services related to maternal and newborn health (ANC, skilled birth attendance, PMTCT, and to some degree PNC). The PBI also positively affected service quality related to HIV counseling and testing, however had no effect on counseling services related to family planning, and affected child immunization services even negatively. Although the underlying causes for these observed patterns could not be identified by our data, findings underscore the point that performance incentives might not always result in desired outcomes. Our ability to assess quality of care indicators for which only binary (have/do not have) data were available and which were only collected at two time points was limited by little room for improvement from baseline and no access to information regarding what was happening between those two time points. Further understanding of how and why different health services respond differently to a PBI program might allow for some adjustments that ultimately improve quality more consistently across services and facility levels.

Depending on the design, PBI programs can restore or even reform certain functions essential to service organization, facility management, and quality assurance by aligning individuals' sense of responsiveness or accountability with general service efficiency. As PBI effects on these service elements are difficult to measure, especially using routine data, we nevertheless identified indicators on service and system organization (i.e. external supervision, facility management meetings, patient feedback procedures) as best possible proxies. Deeper understanding of the extent to which the current SSDI-PBI design contributes to an overall system-reform process within and beyond the enrolled facilities might provide useful information for future program expansion or scale-up.

Challenges or Modifications in Undertaking This Study Component

In order to assess changes in quality, the team expected to use two sources. The first source starts with Service Provision Assessment (SPA) data, which was accessible and could be complemented with primary data collection. The second source, Standards-Based Management and Recognition (SBM-R) data, proved unusable due to data deficiencies and could not be drawn upon for analysis purposes. The research team also expected to be able to draw upon a much wider array of indicators (given those that are listed in the DHIS2), however many indicators had data missing over large periods of time for many facilities or the data were of poor quality. Additionally there was no codebook available to interpret much DHIS2 data (when data had values that were not intuitive); the analysis was thus restricted to variables for which values were interpretable. Furthermore, obtaining data via the DHIS2 website was challenging due to website instability and a poor user interface design. Finally, in assessing service utilization, it was initially envisioned that the team would collect and analyze quantitative data and then use results from that analysis to inform the nature of qualitative data to be collected. Due to an inability to capture quantitative data in a timely fashion, it proved infeasible to delay primary, qualitative data collection until quantitative analysis could be completed.

Study Component 2b. Service Utilization – Qualitative Findings Related to Service Utilization

In order to gain a more nuanced understanding of service utilization, the study team also sought insights from community members who lived in the catchment area of facilities. This included women who had given birth within the preceding year and community leaders such as village health committee members, ward councilors, chief, village headmen, religious leaders and traditional healers. Due to an interest in capturing insights from those living far from facilities (a group that is often neglected in health utilization research), IDIs were later conducted with ten women living at least one hour from intervention facilities and ten women living at least one hour from control facilities.

Methods

Table 8. Service utilization qualitative sample characteristics

Respondent Type		Total
FGD		
	Community Leader	17
	Women	13
IDI		
	Women far from Intervention	10
	Women far from Control	10

Data collectors were trained for five days (including piloting), and all data collection was conducted in March 2016. In total, 13 FGDs were conducted with women and 17 FGDs were conducted with leaders residing in the catchment of intervention facilities. The FGD guide was semi-structured in format. Broadly speaking, FGDs with leaders covered the following topics: changes observed related to health facilities or facility-community relations in the preceding two years, knowledge and impressions of the SSDI-PBI program (if at all), and impressions of changes in the community that could be linked to the program (if at all). FGDs with women covered the following topics: impressions of care received at the facility, factors that compel or deter facility-based care, changes observed in the preceding two years related to the facility in terms of infrastructure or knowledge/attitudes/behavior of providers, recollections of the nature of services received across the maternal care continuum. All FGDs were conducted in Chichewa, simultaneously translated into English during transcription, and coded using NVivo[18]. Codes were deductive and based on discussion guides.

Recognizing that data collection primarily among those living near facilities was inadequate, a Malawian student who is pursuing her Masters in Public Health at Heidelberg University was employed to collect primary data via IDIs in May and June (this has served as the basis for her thesis). The student was mentored by the PI, co-PI and Scientific Coordinator. Her tool emphasized the experience of receiving care during the antenatal, intrapartum and post-partum periods, comparing insights from women living in the catchment areas of control and intervention facilities. She conducted all interviews herself, transcribed all data and is in the process of undertaking content analysis via manual coding (which may be complemented with coding using NVivo software). Codes for this data collection activity are deductive and based on interview guides.

While FGDs with community leaders proved exceptionally rich in detail and valuable in terms of providing insights related to SSDI-PBI, we caution that our report of findings from FGDs with women is restricted. During the course of data collection, FGDs with women were eventually cancelled due to a lack of knowledge among women related to the program and an inability for moderators to effectively convene an adequate participant group and/or to glean insightful information from women related to service utilization. In several FGDs, moderators appear to have grown frustrated during the course of the FGD when women had no insights to share. This apparent frustration led moderators to ask a series of leading questions about characteristics they expected the SSDI-PBI program to entail (for example, “What about outreach by providers in communities, did they do that?”). This scenario, wherein a moderator is leading participants toward a response, compromises the validity of the data thereby rendering it un-reportable. At other times, women would mention changes without being prompted but moderators did not probe on the temporal relationship, making it difficult to determine if changes could be attributed or somehow linked to the SSDI-PBI program. In two FGDs, there was an over-weighted inclusion of women who were new to the area and had no impression of delivery in the facility of interest.

Results

Insights from women related to service utilization

It appears that pregnant women and women of reproductive age were not explicitly sensitized regarding the SSDI-PBI program. To probe for women’s experiences of the program, FGD moderators asked for women’s impressions of changes within the past year, which elicited responses in roughly half of FGDs (no discernable breakdown by district). When asked if they could name any ongoing health programs in

facilities, women mentioned a Mother-to-Mother program, nutrition programs, and youth programs. In one FGD women described knowing of a *chitenje* (clothing wrapper) program (that is likely linked to SSDI-PBI), but when probed explicitly for the name or nature of the program women said "They explained but we have forgotten the name because we were excited about the *chitenje*."

In order to probe on the effects of SSDI-PBI, moderators asked women to describe differences they noticed related to the health facility or their own health care comparing the past year with previous years. In response to this question, women most often mentioned either no noticeable changes or they mentioned being encouraged by local leaders and providers to attend facilities earlier in pregnancy, and being encouraged by providers to bring their husbands to ANC visits. Women also described seeing certain objects purchased (most often, a scale or a blood pressure machine). In one focus group, women mentioned that the period for childhood immunizations had been extended from 10 months to 15 months. Two FGDs mentioned how rooms had been painted.

On the whole, FGDs with women suggest that they see improvements in facilities, but on the whole women continue to desire better care. Most FGDs discussed a lack of equipment or drugs, and a sense among women that providers are overworked, absent or uncaring. The experience of delivering alone or feeling neglected or mistreated was described in several FGDs, although in most discussions women said the level of respect accorded to women by providers has generally improved in recent years (it was not possible to relate the timing of this improvement in order to gauge if improvements were over the span of several years or more so in line with the advent of SSDI-PBI).

In IDIs with women living far from intervention facilities, respondents did not describe knowledge of a particular program but they did recognize that in recent months, facility-based providers had begun coming to their community to provide care. Women (in intervention facilities only; outreach was not described in control areas) described feeling enthusiastic regarding the advent of outreach clinics as it saved them the experience of losing a full day's work to attend care at a facility. Some women described confusion about the outreach clinics, saying that it would be helpful if clinic days could be announced in advance. A minority of women described wishing that the clinics did not take place in trucks, given that privacy can be compromised if a curious onlooker peers into the truck. Women in both control and intervention facilities expressed similar attitudes and understandings related to the need to start antenatal care (ANC) early, to engage partners in ANC visits and to be tested for HIV during ANC, but all respondents underscored that early initiation for those living in distant areas is exceptionally challenging. In relation to intrapartum care, women across intervention and controls highlighted that they struggle to reach facilities. A minority of women in both groups report disrespectful care, and across all respondents facilities are generally described as overcrowded and incapable of providing adequate privacy. While analysis is still underway for this data collection activity, it appears that women in intervention facilities seem slightly more likely to discuss facilities as being better equipped, but this finding will require more detailed analysis.

Insights from community leaders related to service utilization

With the exception of one FGD, community leaders spoke at length and with minimal prompting about their impressions of and experiences with SSDI-PBI. In general, community leaders describe SSDI-PBI as a program that focuses on maternal health, hygiene and cleanliness, and encourages providers and patients to "get along" then rewards these behaviors with equipment. By "getting along" community leaders said they have learned that they have to "send women to the facility" to get services and that providers "have to be nicer to us". In this respect, one leader used the phrase "kupha mbalame ziwiri ndi mwala umodzi" to describe SSDI-PBI; the phrase roughly translates as "killing two birds with one stone". See Table 9 for an overarching perspective on perceived benefits and persistent or emerging challenges related to SSDI-PBI from the perspective of community leaders.

Table 9. Community leaders' impressions of SSDI-PBI facilities

Positive Changes	Negative or No Change
Facilities look nicer (curtains, cleanliness) and have more infrastructure (toilets, labs, maternity wings)	Inadequate staffing, **overworked/ tired/ stressed providers have difficulty delivering quality care**
More community outreach/engagement by providers with community directly and with HACs (Health Advisory Committees)	Overcrowding in facilities
PBI facilities seem to have more drugs than non-PBI facilities	Chronic drug shortages
Impression of enhanced uptake of services (HIV, ANC, maternal health programs generally); Men being encouraged to accompany women for services	Patients are hungry
More materials (motorcycles, uniforms, computers, bed nets lawnmowers, BP machines)	Challenges of distance not overcome
More reliable operating hours	
Providers "forced" to be nicer to clients	

Among the positive aspects of service quality emphasized by community leaders, they most often described improvements in the appearance of facilities, a sense that the facility was trying to bring services closer to communities, that more equipment was available in facilities and that maternal health services seemed to become "more important" in recent months. In relatively equal measures, leaders described how their expectations were (or were not) met. In several cases, leaders were upset that equipment they expected to see delivered months earlier had still not arrived. Less often, leaders appeared to be under an impression that a portion of financial rewards received by facilities would go toward communities directly. The absence of rewards directly for communities was thus discouraging in those (few) instances. Frustrations or challenges with service delivery were discussed with much less depth, but the most notable negative change was that providers appear to be more overworked and tired.

In terms of the SSDI-PBI program specifically, leaders seem pleased with the intervention because they see tangible benefits and they feel sensitized by implementation staff. Nevertheless, community leaders are frustrated by procurement delays, would like to see the community sensitized even more, have concerns about the flow of funds (and question if delays in delivery are a by-product of underhanded financial management), and resent that many goods and labor are sourced from far away ("by and for the people of Lilongwe") rather than in or near the communities within the vicinity of SSDI-PBI facilities. As one leader said, "The money has come, but people from far away are benefiting from this program instead of us. The contractor has brought many people from Lilongwe to work in this project instead of employing us; they could only bring the supervisor. The laborers would be from here. Actually, I am not happy with this."

Discussion

In general, female clients were not conversant on the SSDI-PBI program and did not note any particularly pronounced changes in the year since the program began. Community leaders appear to be well-sensitized on and engaged in the SSDI-PBI program. Leaders note many positive aspects of services offered in facilities amid SSDI-PBI; they also have several suggestions regarding how the program could be improved. Suggested improvements largely center on providing more opportunities to the community (in terms of employing community members to undertake labor), streamlining procurement and keeping the community informed regarding the flow of funds.

Challenges or Modifications in Undertaking This Study Component

In general, this study component encountered relatively minor challenges that are routine in qualitative research (such as a broken tape recorder, a virus in the software used to upload recordings and, in case, language challenges).

Study Component 3. Provider Motivation

It is commonly assumed that one of the key mechanisms through which PBF brings about change is by enhancing health workers' motivation. However, this assumption is rather poorly researched to date. Little is known about whether motivation is affected by PBF, the mechanisms through which PBF achieves effects on motivation, and how a specific intervention design or program implementation shapes PBF. These questions remain pertinent to stakeholders as they attempt to design interventions that are effective, efficient, and side-effect free. This study component aims to generate both specific information that will help improve the SSDI-PBI intervention as well as draw more general conclusions to feed into this broader gap in knowledge.

Methods

Data collectors were trained for five days (including piloting), and all data collection occurred in March 2016. The study component relied on in-depth interviews with health workers in intervention facilities (n=29), as well as a structured survey administered among health workers in intervention and control facilities (n=76). All providers who were available on the day of the visit were interviewed. Respondents for both quantitative and qualitative activities included medical assistants, nurse/midwives, health surveillance assistants (HSAs), one clinical officer and one lab technician. Interviews were performed in English by trained interviewers who used a semi-structured interview guide that covered the following topics: perceptions of the intervention and its components; changes at the workplace as a result of the intervention; and changes in motivation and other intrapersonal aspects as a result of the intervention. Interviews were recorded and transcribed verbatim. The quantitative component entailed a series of questions asking health workers in both intervention and control facilities retrospectively about whether they have or have not witnessed changes in their own effort at work as well as in certain aspects of their working environment in the preceding year.

In terms of analysis, qualitatively we used a hybrid of deductive and inductive coding to analyze IDIs, proceeding through the material and applying a series of codes defined *a priori* on the basis of themes in the interview guide as well as allowing for additional codes to emerge. Qualitative data analysis was supported by NVivo. Quantitatively, we used chi-squared difference tests, assessing whether participants in the intervention group differed significantly from participants in the control group in regard to their perception of changes in their work effort and working conditions.

Results

How do providers perceive the intervention?

One year into the intervention, **health workers expressed very positive overall perceptions of SSDI-PBI**. Originally sceptical and somewhat wary of yet another intervention with labor intensive demands, health workers described how they became convinced of the value of the intervention for themselves, their facilities, and their patients over time. At the time of our study, virtually all respondents expressed high levels of endorsement of the intervention goals. Critical voices were rare and limited to specific, non-fundamental aspects of the implementation. Most respondents expressed feelings of ownership of the intervention, although many suggested minor changes to

Desire for Continuation

"I would love if this PBI did not come to an end because it has brought tremendous results. […] We do not feel like it has been imposed on us. It's like it has put us back on track. We are benefitting a lot." - Senior HSA

the intervention design to enhance ownership further **(see quote box "Desire for Continuation")**. Almost all expressed a wish for a continuation of the project.

Health workers also endorsed most of the **indicators** and generally perceived them as very useful to focus their efforts on critical aspects of care. They expressed appreciation for having **targets** to work towards and guide their effort, but many stated that certain targets are unreachable due to contextual realities and health systems constraints (e.g. HR shortages, drug stockouts). For instance, several respondents complained that the target on women positively tested for HIV was set too high in light of declining incidence rates in their catchment areas. In regards to family planning, some health workers expressed preference for an indicator on counselling activities rather than the provision of family planning methods as such, stating that this would be better aligned with the cultural situation in their catchment areas. Alternatively, the targets attached to the family planning indicator would have to be better tailored to the specific situation in the area of operation. In terms of health systems constraints, health workers complained that reaching certain targets is unrealistic in light of persistent shortages in human resources and drugs, which are both beyond health workers' range of influence.

> *"The indicators are very helpful. Previously we were working without targets but with PBI, there are targets set for us so we always try much as possible to reach our target and even to surpass our targets, so it's really helpful."* (Nurse)

There was initial confusion and dismay among health workers about the **peer-to-peer** process. Many did not find it fair and sensible to be evaluated by people they perceive to be in competition with, and described how this situation has led to instances of purposely lowered performance scores.

> *"They mismarked us because they were revenging on what the previous group had done to them. In the end, instead of scoring good grades, they failed you deliberately."* (Senior HSA)

A year into the intervention, however, most perceived the system as constructive, fair, and helpful in preparation for the external verification.

> *"I feel that it is a very good start. […] Peer to peer helps us to know our strengths and weaknesses so that when the external assessors come, we show them where we have improved."* (Nurse)

The **business plan** model was generally appreciated as a means to obtain resources to improve the working environment. A number of health workers further acknowledge that the business plan drafting process has taught them to think about and manage their work and workplace in a more strategic and foresightful way.

> *"Our facility has benefitted from the business plan in terms of enabling us to procure some essential items for the facility."* (Medical assistant)

> *"The business plan has actually helped me grow in terms of work. I was just sitting down, saying 'This day is gone, what does tomorrow bring?' Now we are able to think 'I think our facility needs to have that and that'."* (Nurse)

However, health workers were rather unsatisfied and frustrated with several aspects of the process of **procurement** and delivery of items in the business plan. First, they perceived the procurement and accounting process as rather opaque. They described how the lack of insight into item prices and remaining account balances made the business plan drafting difficult and arbitrary, and at least initially resulted in the draft of a wish list rather than a strategic procurement plan.

> *"When we're planning, we allocate a certain amount of money to every item, but […] we are not given any receipt on any item procured [...] which would enable us to follow how much money is remaining with us."* (Medical assistant)

Many health workers were further dissatisfied with the quality of items purchased for them by SSDI, particularly in relation to price. There seemed to be some confusion around where items are purchased (US vs locally). Several respondents stated that similar items of better quality could have been procured at lower cost in-country.

> *"We want better transparency and accountability when items are being procured and how money is being used. For example, we received a second hand generator and we sent it back. There is also a delay in delivery which needs to be changed."* (Nurse)

Health workers also complained about substantial delays in the procurement and delivery of items. They perceived this as unfair particularly in light of the fact that they could have improved their performance in the intervention faster – and thus earned more reward budget in the next cycle – had it not been for these delays. Several explicitly spoke about feeling penalized for SSDI's underperformance.

> *"We were penalized on the resources that we do not have. Of course we have requested them on our business plans, but the PBI central office has not yet procured the material for us."* (Nurse)

Many respondents thus pleaded for a more direct involvement of health facilities in the purchasing process.

> *"We participate in the development of the business plan and the specification of what we want to be procured. So in the procurement process itself, they should also involve people more."* (Nurse)

Only 6 of our qualitative interview respondents had participated in any of the **reward announcement meetings**. Those respondents who had attended, however, found it inspiring and motivating, and recommended their continuation.

> *"We see what other facilities have earned. This helps you to work hard because you are motivated that if this facility has done it, why can't we do it?"* (Nurse)

Respondents addressed two challenges in addition to those mentioned above. First, several described how the availability of additional **financial resources**, while beneficial for the facility, acted **as a new source of conflict with the community**, specifically the Health Advisory Committees. According to health workers, community members, possibly due to lack of understanding of the project in the early implementation stages, seemed to perceive ownership of PBI resources, and accuse health workers of embezzlement of 'their' funds. This situation seemed to have improved over time, however.

> *"The relationship with the community has improved, but with some negatives because if staff members procured an item, the community would assume that the facility is using the community's resources."* (Medical assistant)

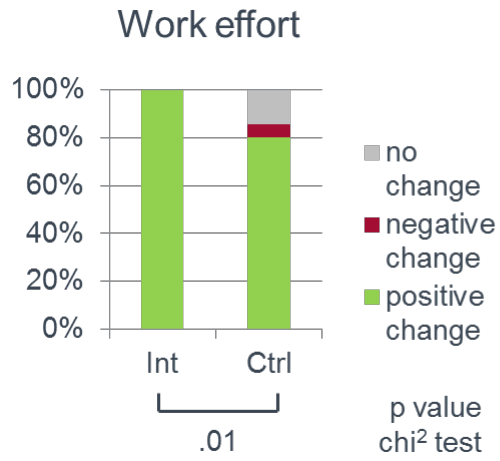
Several respondents also voiced **concerns regarding the sustainability of improvements** in case the program does not continue. While they acknowledged that certain changes in health worker and community mindsets might last beyond the duration of the PBI project, they clearly stated that certain main areas of change, particularly those of regular meetings and enhanced outreach activities, depend so strongly on PBI funds that they will likely collapse with the end of the project. Few health workers even fear a 'below baseline' situation should PBI end.

"Once PBI is gone, there would be total chaos. […] An example: Whenever we are having a staff meeting, everybody is present. All 40 members of staff are there because they know refreshments are there. When refreshments are not there, only 10 are present. Last December, we had a staff meeting with no refreshments. This February, people were nowhere to be seen. So I wish PBI never left so that this togetherness still remains." (Medical assistant)

How has the SSDI-PBI intervention changed motivation of health care providers, and through which mechanisms? Are changes in motivation reflected in changed behavior at work?

In-depth interviews with health workers revealed that motivation to work hard has increased as a result of PBI for almost all respondents. In the health worker survey, all respondents from intervention facilities stated that the effort they put into their work has increased in the year prior to the survey, while respondents from control facilities did so to a significantly lesser degree (see Figure 27).

Figure 27. Work effort according to health worker survey



Health workers explained that the intervention achieved this improvement in work motivation through the following mechanisms:

PBI served as a wake-up call to health workers

Health workers described how the introduction of PBI had acted as a wake-up call to them in that it opened their eyes to shortcomings in their performance by introducing indicators and allowing them to evaluate their current performance against these indicators. Most health workers perceived this wake-up call as a motivation to do better in the future.

"The [indicators] are helpful because they gave us a wake-up call. We know now that we should do more." (Nurse)

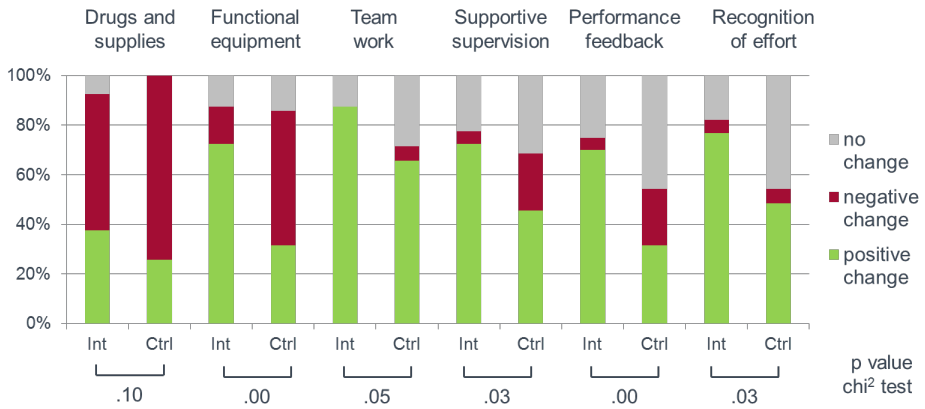
PBI motivated health workers by introducing goals to work towards

Although treatment standards and protocols had existed prior to the intervention, PBI provided health workers with a relatively concise list of priority indicators and specific targets to work toward. Health workers described how their desire to achieve the targets set for them, stemming both from a desire to earn the rewards and an intrinsic enjoyment of the challenge, motivated them to work hard.

"We feel motivated. We were working in an inappropriate manner. Now, we strive to attain the set targets and work on the areas where we were not performing well in order to increase the quality of care we provide." (Nurse)

Health workers are motivated by the perceived substantial improvements in their working environment due to PBI, which facilitates easier and more effective work

Figure 28. Provider perspectives on changes in work environment



Virtually all health workers in intervention facilities described how PBI has changed their working environment for the better in one way or another. Results from the health worker surveys corroborate these findings (see Figure 28). Health workers described how these improvements eased their daily work and made them feel more effective and successful in their work. This not only motivated them to work hard towards the PBI indicators so as to gain even more resources for further workplace improvement in the future. Health workers also indicated a positive effect on their overall work motivation beyond only the PBI-incentivized tasks.

> *"If someone is working in a conducive environment, he or she is willing to give the best he or she can. With this [PBI], I am really motivated."* (Medical assistant)

Specifically, health workers explained that the intervention through the reward budgets and business plans allowed them to increase the availability of medical equipment and certain supplies (e.g. IP material, test kits) which they were previously in dire need of.

> *"There are a lot of changes in terms of equipment. Now I am enabled to work efficiently and effectively, knowing that I will examine my patients properly and treat them accordingly."* (Medical assistant)

Health workers further described that they receive more and better supportive supervision and helpful feedback on their performance than they did prior to the intervention, both from SSDI and through enhanced engagement of the DHMTs. This improved supervisory situation is generally appreciated and perceived as helpful by health workers.

> *"Way back the supervision was not being done according to protocols. They could come supervise and leave without giving any feedback. But now they are using consolidated check list, and after supervising us we sit down and they give us feedback and see where we have challenges and discuss the way forward."* (Medical assistant)

PBI has also improved the quality of teamwork in health workers' opinions, as well as greatly impacted on health facilities' collaboration with and involvement of the communities. Health workers perceived this improved situation as both helpful and gratifying.

> *"Through this project, there has been such a tremendous relationship with the community. […] For instance, one part that would bring down our scores is [lack of] electricity. So we had a meeting and I was the one who introduced this topic [to the community]. […] The community took up the whole role and divided the chores amongst them. It was agreed that each household was to contribute K200. […] I am very proud of my community. Now the community is in the process of building up an incinerator."*

(Medical assistant)

"Staff members are now working as a team compared to before the project was introduced. […] Like the distribution of jobs, in case of shortage of staff, […] other team members are willing to help." (Medical assistant)

We quantitatively assessed a few further dimensions of change assumed to result from PBI from the health workers' perspective, but did not find a statistically significant intervention effects. Specifically, we were unable to show an effect on health workers' **perceived accountability for their performance**. However, 73% in the intervention group said that they feel more accountable than before PBI, but so did 66% in the control group. Similarly, the majority of respondents in the intervention and – to an only slightly lesser degree – in the control group stated that they perceived more **autonomy to decide how to go about their work**, more **influence on decision-making** at facility level, and a better **idea of what is expected** of them at work.

However, health workers also made clear that their working conditions are still far from ideal. They explained that they in particular continue to struggle with shortages in drugs, as drugs need to be provided through official government channels and cannot be purchased through SSDI business plans.

"Sometimes we run short of drugs because of the system that the drugs should come straight from headquarters. We have stayed without LA [antimalarial] and ferrous sulphate for about a month or two." (Nurse)

This not only continued to contribute to challenges in patient care, but for certain health workers also dampened enthusiasm about the intervention and limited the motivational effect the intervention could have had.

"On the one hand, I have been motivated while on the other hand, I have not been motivated. Because we received equipment for the maternity ward which was motivating for me. There was however a shortage of drugs which did not motivate me as much." (Nurse)

They indicated that the motivating potential of the intervention in regards to improvements in their working environment was further limited by the above-discussed delays in provision and the perceived substandard quality of equipment and material by SSDI.

"When we are rewarded we feel more motivated, but when we don't get the resources in time, maybe we were promised to get something and then we don't get it, it is like the people are more demotivated than they were before we started the PBI project." (Nurse)

PBI motivated health workers by fostering a sense of initiative and proactivity towards shaping one's work environment

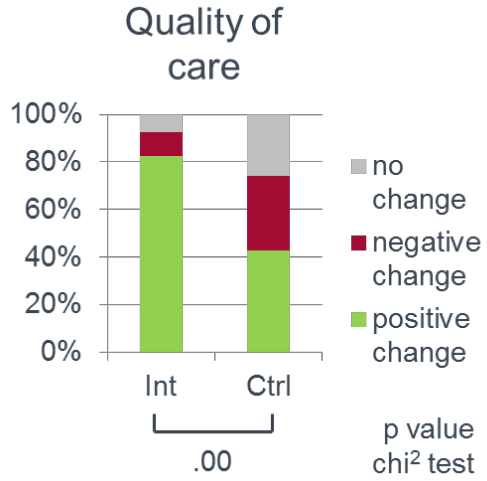
Prior to PBI, health facilities fully depended on the DHMTs for resources, and often experienced stockouts of materials and drugs as a result of suboptimal communication and logistic channels. As discussed above, health workers appreciated that PBI allows them to bypass this official supply system by allowing them to procure certain resources directly and according to their specific need. However, several health workers also acknowledged that PBI made them realize their role in the malfunctioning of the official system. Specifically, they explained how experiences in the context of PBI showed them the potential of taking initiative and leadership in improving their work situation and overcoming feelings of being stuck and helpless in the system. Health workers gave examples of this newfound sense of proactivity in relation to involving the community in problem solving as well as in regards to improved communication with the DHMTs.

"Previously, if the health centers had problems, they would wait for the district hospital to do something. [...] But now, it's like people are taking a leading role like where they have a problem, they are able to motivate the community or are able to organize themselves and get things done that previously were left for management to do." (Nurse)

Health workers are motivated by feeling able to provide better services to more patients

Health workers described how the project has enabled them to provide better care as a result of improved working conditions as discussed above. This is also reflected in answers to the health worker survey question as to whether quality of care has changed for the better in the last year (Figure 29). They further explained that the intervention enabled them to provide care to more patients then previously by allowing them to conduct outreach clinics and incentivize utilization of health care services through handouts/patient incentives.

Figure 29. Provider perspectives on changes in quality of care



"This project has helped us to give quality care, and because of this people are more motivated to come to this facility." (Nurse)

Many health workers expressed that being able to provide better care and reach patients they were previously unable to reach is highly motivating as it allows providers to act on the altruistic motives that originally drew them to their profession and to enact the principles and standards they committed to when joining the profession.

"[PBI] has helped us be more motivated and more hard working not only because of the support with equipment, it has also helped us on how we care for our patients, on how to reach people who we were unable to reach." (Nurse)

Although health workers perceived these improvements in patient care as overall positive and felt proud of them, several also acknowledged that they entailed a substantial increase in workload. This is also reflected in the quantitative data, where 85% of intervention respondents (but also 77% of control respondents) stated that their workload had increased in the previous year. In the in-depth interviews, several health workers expressed difficulty in handling this increased workload in the absence of increases in staffing levels, and described how fatigue and feelings of being overwhelmed dampened the effect of PBI on their work motivation

"With the coming of PBI, the workload is just high. To provide quality services we need a lot of human resources." (Nurse)

"It affects my motivation. We are always tired." (Nurse)

Discussion

Perceptions of the intervention

Our findings show that the intervention was well understood and generally appreciated by health workers. Health workers liked having indicators and targets, even though some of the targets should be revised in their opinion. Following initial dismay, health workers also came to appreciate the peer-to-peer process as constructive feedback and preparation for the external verification. Health workers were also generally appreciative of the business plans as a means to obtain resources for the facility, especially in light of the particularly severe resource shortage in the Malawian health sector during the study period. Only few of our interview partners had attended a reward announcement meeting, but those who had found it helpful and inspiring.

However, our study also revealed implementation challenges particularly in regards to the procurement process that significantly dampened health workers' enthusiasm about the intervention and should urgently be addressed. Health workers were frustrated by a lack of transparency in the entire procurement process, about expensive but low-quality material bought, and about substantial delays in the delivery of items. For the future, they wished for more transparency in the sense of a better overview of their accounts and account balances as well as over supplier options, item prices, etc. Many would prefer to be actively involved in the purchasing process. Lastly, health workers pleaded for timely delivery of items and particularly cash for outreach and meetings, as problems related to the latter had substantially impeded progress in the first year of the intervention.

PBI influence on motivation

The intervention positively impacted health workers' motivation to make an effort and perform well, both in relation to the PBI indicators and beyond. PBI seems to have done so via a combination of several mechanisms: First, the intervention acted as a wake-up call to health workers, reminding them of their professional duties and commitments. Second, it motivated health workers to improve their performance by giving them tangible goals to work towards, introducing a welcome element of challenge and competition to a previously 'business as usual' environment. Third, health workers perceived substantial improvements in various dimensions of their working environment as a result of PBI, including equipment and supplies, supportive supervision and performance feedback, team work and community collaboration and involvement. Virtually all respondents found these improvements highly motivating and enabling. Forth, a number of respondents explained that PBI had motivated them by instilling a sense of proactivity and initiative in them. The intervention achieved this by making them realize that unlike previous beliefs and practices, they are in fact able to positively shape their working conditions. Fifth, health workers described that the intervention had enabled them to provide better care to more patients, motivating them by allowing them to act on altruistic motives which for many were central in their decision to join health care.

Despite all positive change, certain aspects hindered the intervention from reaching its full motivational potential. As described above, health workers were dissatisfied with the SSDI procurement process, which dampened their enthusiasm. Although the intervention allowed health facilities and health workers to address a number of challenges of the broader health system in a decentralized manner, others were beyond their control. Specifically, health workers explained that persistent shortages in drugs and staff hindered them in progressing with the performance improvement process. Many perceived this as frustrating and demotivating in light of the above-described "spirit of change" in the health workforce and a certain social and financial pressure to perform better introduced by the intervention.

In future continuation and potential scale-up of the intervention, our findings suggest that the project will gain from revising the procurement process, and from adapting targets even better to health facilities' specific realities as well as broader contextual constraints. In regards to the latter, it might make sense to

discuss certain policy changes regarding drug supply chains and human resource practices if the intervention is to be scaled-up. Given the importance of the indicators and targets as a wake-up call and as a 'fun challenge' to work towards for the overall motivational effect of SSDI-PBI, we further recommend that the implementation team pay specific attention to finding ways to sustaining these effects and prevent the health workforce from going back to 'business as usual' over time.

Challenges or Modifications in Undertaking This Study Component

The evaluation had planned to collect more extensive quantitative data from providers in both intervention and control facilities related to motivation. As the study team learned more about the implementation of the SSDI-PBI program, we decided to forgo this approach (due to the absence of baseline data as well as the small sample size); while quantitative data was collected in both intervention and control facilities the application of a more detailed motivation scale was not used. Instead, qualitative data (in intervention facilities only) was used to more deeply explore motivation in the context of SSDI-PBI.

Study Component 4. Economic Evaluation

Our economic evaluation of the SSDI-PBI intervention situates itself within a context characterized by the almost complete absence of comparable evaluations addressing the relationship between costs and benefits of Performance Based Financing. This is largely due to the complexity of such interventions, which operate at multiple levels and affect a multitude of health service outcomes. This complexity poses a challenge both in tracing economic costs at all relevant levels and in identifying meaningful outcome measures for an economic evaluation.

Methods

Study Design

The common approach for measuring "value for money" entails the estimation of an incremental cost effectiveness ratio (ICER) that relates the incremental costs of an intervention with its added benefits. By comparing an ICER with a reference value, it can be assessed if the change in benefits is worthwhile given the change in costs associated with the intervention. However, in our study we had to refrain from calculating an ICER of SSDI-PBI given that it was not possible to establish a direct link between the services targeted by the intervention and an actual benefit measured as improved health - due to a context-relevant lack of epidemiological and clinical data. The appraisal and interpretation of an ICER is only possible when the value can be expressed in terms of a health gain (life-years saved or Disability Adjusted Life Years) - not in terms of increases in service provision, which was the only meaningful measure of change in the SSDI-PBI intervention.

In the light of the abovementioned limitations, we conducted a cost-consequence analysis. A cost-consequence analysis tracks the total costs of implementing an intervention as well as the benefits produced by this same intervention. Costs and benefits are presented separately, allowing decision makers to judge whether the benefits accrued are sufficient to justify the costs incurred to generate them.

Cost analysis

Our analysis aimed to estimate the **economic costs** of all activities related to the implementation of the SSDI-PBI intervention. Economic cost analysis considers the value of all resources used in providing a service, whether or not the resources were purchased directly by the program. Since economic resources are limited, any decision to undertake a healthcare program will divert resources from an alternative use. This means that the benefits to be derived from the expenditure in question should be compared to the benefits that would have been obtained if the money had been used elsewhere. This is the concept of opportunity cost. In essence what we have to measure when we are dealing with the costs of a healthcare program is the opportunity cost of using a resource in the program instead of in its best alternative. Costing healthcare programs/services requires identifying all cost-generating components and attributing a monetary value to them.

A common approach to calculating program costs is the resource-cost method. This approach includes itemizing the resources necessary to provide services, and calculating or estimating the costs of each resource. Researchers have advocated this method as a means to develop cost estimates that reflect the value of all resources required for delivering a program.

In our study, the costs associated with SSDI-PBI included the following cost categories:

1) Personnel costs
2) Design costs (design of program, training, initial dissemination)
3) Implementation costs (including verification and counter-verification);
4) The cost of incentives

Our initial aim was to estimate the cost of single activities across different phases (e.g. training, supervision, verification) entailed by the SSDI-PBI intervention. However, due to lack of adequate data, we could not compute the full cost of the single activities. This was mostly linked to the fact that, due to the structure of the data provided by the implementing partners, we could not allocate personnel costs across activities. Regarding the cost of incentives, we include them in our analysis (measured in relation to what was "earned" by the single facilities and not what was effectively "spent") to measure the economic value of the additional effort needed to produce an increase in quantity and quality of service provision. This decision was based on the assumption that the incentives represent a good proxy of the value for which healthcare providers are willing and able to increase their level of effort and produce increases in service provision.

The perspective of the analysis affects the costs considered. In our study we adopted a health system perspective, which reflects a combination of unduplicated costs to the government, funding agency and the service provider. Given the short time and the resources available and given the specific interest of the funding agency, we did not consider the adoption of a societal perspective, accounting also for costs incurred or avoided at the societal level.

Benefits (consequences)

Benefits were computed looking at the increase in service provision accrued during the abovementioned implementation period on 11 out of 13 SSDI-PBI indicators for which we could access sufficient data (see Table 10).

Table 10. Effects detected across incentivized indicators

Incentivized Indicator	Estimated total effect 17 months post-intervention (percentage points)	Immediate effect attributable to the intervention (percentage points)	Change in long-term monthly trend attributable to the intervention (percentage points)
1. Number of pregnant women starting antenatal care during the 1st trimester	**208**	**20**	**11 *** †††**
2. Number of women completing the 4 ANC visits	**151**	**55 ****	**6 ****
3. Number of pregnant women receiving at least 2 doses of IPT	**81**	**23 ***	**3 ****
4. Number of births attended by skilled birth attendants	**1**	**1**	**0**
5. Number of 1 year old children who are fully immunized	**23**	**8**	**1**

Incentivized Indicator	Estimated total effect 17 months post-intervention (percentage points)	Immediate effect attributable to the intervention (percentage points)	Change in long-term monthly trend attributable to the intervention (percentage points)
6. Number of HIV-positive pregnant women who were initiated on ART	**85**	**43****	**2**
7. Number of HIV/AIDS cases screened for TB			
8. Number of children receiving Vitamin A supplementation (reported using non-zero average of **3-month periods** and slopes over **3-month periods** rather than monthly periods)	**269**	**59**	**52 ** ††**
9. Number of clients counseled for FP	**-29**	**-9**	**-1**
10. Number of couples tested for HIV during HTC services	**374**	**-19**	**23** ††**
11. Number of infants born by HIV positive mothers tested for HIV			
12. Number of women who receive PNC by skilled HCWs within 2 weeks	**32**	**11**	**1**
13. Number of pregnant women attending ANC receiving iron supplementation	**218**	**-14**	**14* †**

Stars mark significant changes in intervention group compared to control:
* p < 0.05 ** p < 0.01 *** p < 0.001
Daggers mark significant difference in slopes after intervention
† p < 0.05 †† p < 0.01 ††† p < 0.001

Due to a lack of adequate data, our estimate of the benefit is purposely not adjusted for quality of care considerations. In addition, please note that due to the need to align cost and benefit data, the analysis of the consequences included in the economic evaluation differs from the analysis of the consequences displayed above in relation to changes in service provision. The analysis of the benefits included in the economic evaluation is limited to the period Aug 2014 to Dec 2015.

Data Collection

Our aim was to estimate the full cost of SSDI-PBI through the mixture of data from different sources, combining secondary with primary data collection. We collected the data from the financial statements of two implementing agencies, Jhpeigo and Abt. They provided Excel spreadsheets including their financial costs associated with the SSDI-PBI intervention. The information contained in these files had a variable level of detail. Abt data allowed for the analysis of two different phases (design and implementation) while Jhpeigo data were considerably more disaggregate across the time (on monthly basis). Based on these sources, we were able to estimate full costs of design and implementation phases but not of specific activities (e.g. training, supervision, verification, etc.). More specifically, neither of the two implementing agencies provided the breakdown of personnel cost across different activities, but only an aggregate measure associated with the two phases.

In order to capture the full costs of all personnel involved in the SSDI-PBI intervention, we estimated personnel costs incurred by the other two implementing partners, MoH and USAID, through activity costing. We used project reports and documents to identify all activities to which MoH and USAID staff had participated. The documents allowed us to draft a detailed list of the single people involved in the project as well as details on the time committed by each person to the intervention through participation in

different activities (e.g. design workshop, supervision, etc.). To estimate the total value of the time committed by MoH and USAID personnel, we used the human capital approach (according to which value of time is measured through the earnings of an individual) and for each individual, we multiplied the time committed by their earning. We collected information on earnings through direct contact with the USAID and MoH human resource offices.

As mentioned above, we used the value of the incentives to estimate the value of the additional effort needed by providers to produce increases in health service provision. Information on the value paid out as incentives was provided directly by Jhpiego for the 16 months included in the evaluation.

Data Analysis

Our evaluation distinguishes costs between the SSDI-PBI design (Sept 2012 to July 2014) and the implementation phase (Aug 2014 to Dec 2015). All costs were computed in US dollars adjusted for inflation from the year in which the costs were incurred to the year 2015. The average exchange rate for the period 2014-2015 was used to convert values that were incurred in Malawian Kwacha (MKW) into US Dollar (USD).

Results

Counting the period Sept 2012 to Dec 2015, the economic value of the SSDI-PBI intervention amounts to USD 3,402,187, with about one third (USD 1,161,332) being absorbed by the design phase.

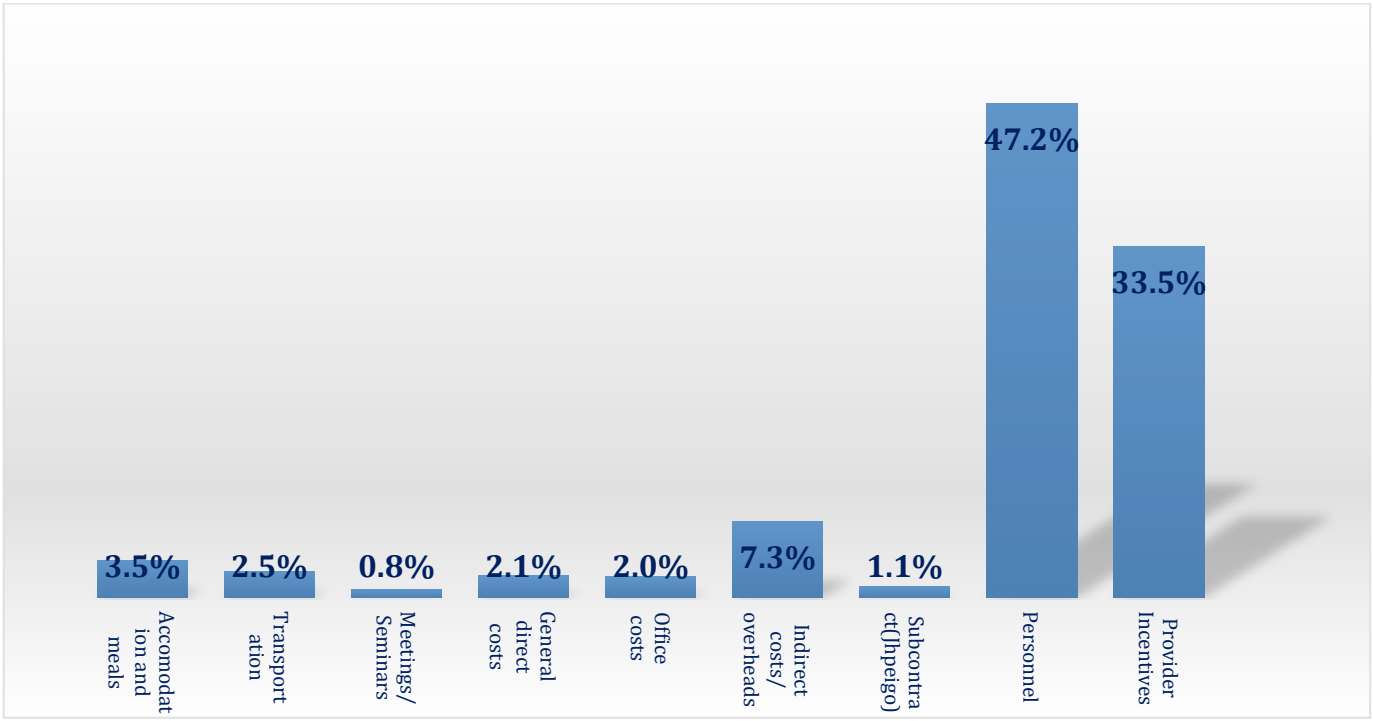
Table 11 shows the total economic value across cost categories by phase. With an estimated value of 1,140,436, the incentives represent about one third of the total value of the intervention and about half the value of the implementation costs. With a value of USD 1,605,178, personnel represent the single most relevant cost of the SSDI-PBI intervention. It is worth noting that the economic value of personnel was substantially higher during the design (USD 934,045) than during the implementation phase monitored in our analysis (USD 671,133). The high personnel costs during the design phase are driven by a high time commitment on the part of USAID staff, who were advising on the development of the intervention.

Table 11. Total costs of SSDI-PBI by cost category and phase

	PBI Design	PBI Implementation	Total
Accommodation and meals	$ 42,528.79	$ 77,739.73	$ 120,268.52
Transportation	$ 47,398.95	$ 38,968.51	$ 86,367.46
Meetings/Seminars	$ 19,312.18	$ 7,764.80	$ 27,076.98
General direct costs	$ 17,058.67	$ 54,815.53	$ 71,874.20
Office costs	$ 24,226.36	$ 43,449.68	$ 67,676.04
Indirect costs/overheads	$ 76,762.26	$ 170,148.07	$ 246,910.33
Subcontract (Jhpeigo)		$ 36,398.74	$ 36,398.74
Personnel costs			
Abt	$ 238,937.08	$ 65,281.68	$ 304,218.76
Jhpeigo	$ 374,063.51	$ 527,716.56	$ 901,780.07
Ministry of Health	$ 4,882.97	$ 8,085.77	$ 12,968.74
USAID	$ 316,160.94	$ 70,049.24	$ 386,210.18
Total personnel			***$ 1,605,177.75***
Incentives	0	$ 1,140,436.89	$ 1,140,436.89
Grand Total	**$ 1,161,331.71**	**$ 2,240,855.22**	**$ 3,402,186.93**

Figure 30 shows the distribution of cost across costs categories to show the most relevant costs incurred by personnel (47.2%) and cost of incentives (33.5%). All the remaining costs accounted for a bit less than 20% of total costs.

Figure 30. Distribution of costs across categories



By focusing only to PBI design phase, we estimated cost in different categories which are presented in Table 12. The total amounted to USD 227,287, excluding personnel costs of USD 934,045 (see above). Almost 50% was due to indirect and general costs (overheads) while the rest incurred to cover variable costs of the meetings (accommodation, meals, transportations etc).

Table 12. PBI design costs (excluding personnel)

Cost category	Costs	
Accomodation and meals		42,528.79
Transportation		47,398.95
Meetings/Seminars		19,312.18
General direct costs		17,058.67
Office costs		24,226.36
Indirect costs/overheads		76,762.26
TOTAL	**$**	**227,287.21**

Table 13 shows the estimated costs for implementation of SSDI-PBI, excluding the personnel cost (of USD 671,133). Also in this phase, an important portion of the total costs was due to overheads and general costs of the implementing agencies (approximately 50%).

Table 13. PBI implementation costs by category (excluding personnel)

Cost category	Costs
Meetings/Seminars	7,764.80
Subcontractor	36,398.74
Transportation	38,968.51
Office costs	43,449.68
General direct costs	54,815.53
Accomodation and meals	77,739.73

Indirect costs/overheads	170,148.07
TOTAL	**$ 429,285.07**

Estimating the full costs of the personnel was a major challenge. Since our aim was to capture full costs of all personnel involved in the program, we didn't base our estimates on the financial costs of personnel reported by the two implementing agency only. We made an effort to include the value of human resources employed at donor agency (USAID) as well as MoH who participated in numerous meeting and initiatives associated with the design and implementation of SSDI-PBI.
Table 14 shows the distribution of total personnel costs across the two phases and by actor. As it could be expected given the different roles of the two implementing agencies, personnel costs at Abt level was relatively higher in the design phase versus implementation, while Jhpiego personnel was more involved in the implementation phase. USAID personnel were proportionally more dedicated in the design phase while the MoH resources were almost doubled in the implementation phase in comparison to the design. The total cost of personnel by Jhpiego accounted for approximately 56% of total personnel costs, while MoH personnel resulted in less than 10%. It is important, however, to underline that the main driver of the low value of MoH personnel are the unit costs of time assigned to this category which are determined by their local salaries.

Table 14. Personnel costs by agency and phase

Actor	Design	Implementation	Total
Abt	238,937.08	65,281.68	304,218.76
Jhpeigo	374,063.51	527,716.56	901,780.07
Ministry of Health	4,882.97	8,085.77	12,968.74
USAID	316,160.94	70,049.24	386,210.18
Total personnel costs	**$ 934,044.50**	**$ 671,133.25**	**$ 1,605,177.75**

Finally, we estimated the costs of provider incentives to be included in the full costs of the program. The underlying assumption, as mentioned above, is that the value of "earned" or budgeted incentives corresponds to the value of additional services provided under the SSDI-PBI regime. Data didn't allow us to analyze in more detail these "provider costs". For example, we were unable to estimate the costs of setting up and managing the performance data collection at provide level to assess the incremental costs of the new system.

Table 15 shows the economic value of the incentives accrued by each facility over time (we have purposely removed information that would allow the identification of the facilities). Two elements clearly emerge. First, the value of the incentives varies substantially by facility, showing different responses to the intervention. Second, for most facilities, the value of the incentives increased substantially between the first and the second payment cycles and increased by an additional 15% in the third payment cycle. This finding is coherent with prior studies on PBI and indicates that health providers' capacity to respond to the targets set by a PBI intervention increases over time. This can be attributed to two factors: an increased understanding of the intervention and requirements entailed over time, allowing providers to work on improving their performance in relation to the PBI indicators; and the improved working environment resulting from utilization of PBI rewards which physically enabled facilities to do so. The facilities that persistently display a low incentive value across all payment cycles probably suffer from systemic challenges, impeding them to respond to the SSDI-PBI intervention. It is advisable to identify them nominally and address their weaknesses with specific additional interventions.

Table 15. Economic costs of provider incentives

	September 2014-March 2015	March 2015-September 2015	September 2015-December 2015
Facility 1	56,567.56	53,525.80	41,036.45
Facility 2	8,308.74	8,172.44	6,265.54
Facility 3	21,194.33	19,849.11	15,217.65
Facility 4	50,740.55	97,312.90	74,606.56
Facility 5	13,083.01	25,917.19	19,869.84
Facility 6	12,745.56	16,617.68	12,740.22
Facility 7	9,278.63	20,181.60	15,472.56
Facility 8	11,699,34	22,855.65	17,522.67
Facility 9	17,946.33	38,085.23	29,198.67
Facility 10	13,026.00	12,807.22	9,818.87
Facility 11	17,111.82	33,969.86	26,043.56
Facility 12	4,214.08	5,233.98	4,012.72
Facility 13	3,597.18	3,928.86	3,012.13
Facility 14	16,119.50	31,941.99	24,488.86
Facility 15	5,937.97	7,286.76	5,586.52
Facility 16	4,334.89	10,455.83	8,016.14
Facility 17	31,806.04	68,872.24	52,802.05
	$ 297,711.53	**$ 477,014.36**	**$ 365,711.01**

Discussion

The inability (related to the structure of the study) to derive an ICER combined with lack of comparable studies makes it difficult for us to appraise the relationship between costs and benefits observed in our study. We leave it to the concerned stakeholders to appraise whether the amount of resources consumed by the intervention is justifiable in the light of the health service benefits produced. This assessment can be informed by knowledge on the resources consumed by other interventions with similar objectives and scope. In guiding the appraisal of the findings, we would like to draw attention to the fact that half of all resources were absorbed by personnel costs and that one third of total costs were incurred during the

design phase. This suggests that the implementation of such a novel intervention requires a very high initial level of effort.

Given lack of relevant data, our evaluation could not estimate the relative cost of activities within the SSDI-PBI intervention. This represents a major limitation, considering that activity costing would have been beneficial to inform decisions on potential further scale-up by indicating the relative cost of the different activities. In particular, it would have been desirable to discern the cost of the verification activities to fully appraise the value of peer-verification as an innovative verification strategy.

Interestingly, the incentives themselves absorbed one third of the overall costs of the intervention. The fact that the amount paid out as incentives increased over time is an indication of how health facilities' responsiveness to the intervention increased over time. The fact that the amount paid out as incentives differed substantially across facilities suggests that different facilities experienced both different starting points and a different capacity to respond to the PBI intervention over time.

It is essential to note that our estimation of the costs and benefits assumes the existence of SSDI as an underlying health system intervention. This is to say that we estimated the additional costs and the additional benefits accrued by the SSDI-PBI intervention as compared to SSDI alone. Should SSDI-PBI be compared to status quo, i.e. to facilities not included in the SSDI intervention, the estimation of costs and benefits may appear radically different. Specifically, we postulate that the mere existence of SSDI as an underlying intervention substantially lowered the costs of implementing PBI compared to the level of resource consumption that would have been needed to implement a comparable intervention in standard Malawian health facilities.

Likewise, in spite of our efforts to carefully trace the full economic cost of the intervention, it is possible that our results represent an under-estimation of the full economic cost of the SSDI-PBI. Specifically, we postulate that this is a possibility to be considered due to two elements. First, we do not know and could not assess if and to what extent the overall SSDI intervention cross-subsidized the SSDI-PBI, with time from SSDI staff being devoted to supporting SSDI-PBI, without this time being formally traced as SSDI-PBI resource consumption (the assumption being that limited task-shifting is common rarely traced in an organization). Second, we could estimate and value the time committed by MoH staff only for the core list of activities provided to us by the implementation team. We cannot exclude that MoH staff committed some time to the intervention beyond this formal list of activities.

Challenges or Modifications in Undertaking This Study Component

- The economic evaluation team would have liked to compute the cost of the single activities (e.g. training, supervision, verification) entailed by the SSDI-PBI intervention. Due to a lack of adequate data, this was not possible.
- The SSDI-PBI intervention relies on multiple indicators, all measured in relation to increases in service use and all directed at different target populations. As such, each indicator was meant to produce a health gain through a different causal path. The multiplicity of targets and their nature made it impossible for the evaluation to produce an incremental cost effectiveness ratio (ICER) since modeling the link between the observed increase in service use (over a wide range of indicators) and an ultimate health gain was beyond the scope of our study. Likewise, given that the indicators targeted different services directed at different target populations, we could not condense them into a meaningful single outcome measure simply reflecting increases in "general" service use.
- Due to time and resource constraints, our estimation of the costs relied primarily on secondary data, mostly the financial statements produced by the implementing agencies. This approach allowed us to accurately trace all actual costs incurred by the implementing agencies. However,

given the way data was stored by the implementing agencies, this approach did not allow us to carry out detailed activity costing, primarily because we could not trace personnel costs (the major cost category) across the single activities.

- The need to rely on the financial data provided by the implementing agencies delayed our work substantially. Unlike what was originally expected, the implementing agencies experienced difficulties and challenges in sharing their financial data. Permission to share financial data was sought at multiple levels, each time introducing further delays. In the future, it would be advisable for the funding agency to clear issues pertaining to the sharing of financial data before such an evaluation is commissioned. This would allow better prospective planning.
- To estimate costs not traced by the implementing agencies, such as USAID and MoH costs, we relied on primary data collection, estimating both resource use and unit costs. At times, the estimation of resource use for personnel proved challenging, given that some people did not respond to the request to provide information on their time commitment to the intervention and their resource use had to be approximated using estimates provided by their colleagues.

Overarching Evaluation Modifications and Adjustments

During the course of the study, the evaluation team made changes to the initially planned methodological approach, which are highlighted at the conclusion of each study component. The following is a list of overarching changes:

- While not explicitly stated in the original proposal, the study team envisioned having the same observation period across study components. Ultimately, the components have slightly different observation periods.
    - For the economic evaluation component, data is secondary and draws from information provided by implementers to the team and complemented with routine health system data. The data provided by implementers is focused on cost and covers the periods from September 2012 to Dec 2015 (the last period for which data was available). The data on health service utilization was provided by the Ministry of Health (via its DHIS2) and covers the periods from August 2013 to December 2015 (12 months of baseline data; 17 months of intervention period data).
    - For the service utilization component, quantitative data draws largely from secondary sources and analysis was conducted for the periods from August 2013 to February 2016 (12 months of baseline data; 19 months of intervention period data). For primary data collection in the service utilization component, data was collected in March and April of 2016.
    - For the fidelity of implementation component and motivation components, data relies on primary data collected in March and April 2016.
- The team initially envisioned conducting a series of webinars to share preliminary findings across various stakeholders. This proved unnecessary as presentations of preliminary findings were ultimately conducted in person in May 2016, while members of the evaluation team were in-country to present findings from another study.

Significance of the Findings

An in-depth assessment with relatively small sample, absent more ideal baseline data and after only one year of program implementation is challenging. Many of the findings, particularly in IDIs with providers and relevant stakeholders within the MoH and implementing partners give an impression that the intervention is still 'settling', that facilities are only starting to fully understand both the concept and operational aspects of PBI. A more rigorous impact evaluation would entail a more tailored baseline, a longer period of time for a program to mature (possibly with a midline study) and, finally, an endline evaluation. With these caveats, we nevertheless view the findings from this study as meaningful in terms of providing timely,

relevant advice in the event of program scale-up or expansion in Malawi or a similar setting. We also sought to draw on novel data collection methods in order to harness wide swaths of existing, secondary data.

In terms of barriers to undertaking the study and efforts to mitigate such barriers, please review the sections “Challenges or Modifications in Undertaking This Study Component” within each study component and the “Overarching Evaluation Modifications and Adjustments “ section above. Almost all challenges are linked to difficulties inherent to a study that relies heavily on secondary data (delayed or denied delivery of data, or inconsistencies in the data once received). The team made several efforts to address this situation in a timely, yet scientifically sound manner. While the analysis of secondary data portends tremendous benefits (especially in terms of capturing more data at a lower cost), in countries with less-than-ideal data repositories -- challenges persist.

Concluding Comments and Recommendations

The SSDI-PBI program has garnered tremendous buy-in across several stakeholders and, even relatively early in the program implementation, promising improvements in service utilization are evident. In order to bolster the program, some adjustments are worth considering:

- Revise the procurement process: create transparency in relation to what exactly is procured and how; involve facilities more in the actual procurement process; create transparency over accounts; procure high quality material; improve the timeliness of supply of items and cash
- Revise targets to better reflect facilities‘ realities; consider tailoring the program based on facility level given that hospitals and health centers have varying needs
- Consider providing start-up support to bring facilities up to an adequate level in terms of infrastructure and equipment, to assure sufficient human resources to cover increases in workload, and to assure steady drug supply pre-roll out of the intervention
- Better sensitize clients – particularly women – on the program and what they should expect from it
- Better sensitize both communities and providers regarding the flow of funds and goods, and the purpose of business plans (particularly at program outset)
- In the event of case of scale up, consider planning an economic evaluation from the onset of an intervention, to ensure the most accurate data collection. Also consider implementing activity costing to allow future costing studies to track the specific value of each activity.

Acknowledgements

This research project is made possible through Translating Research into Action, TRAction, which is funded by United States Agency for International Development (USAID) under cooperative agreement number No. GHS-A-00-09-00015-00. The project team includes the Heidelberg University as prime recipient and the University of Malawi at the College of Medicine in Malawi as a sub-recipient.