



Effect Heterogeneity in Responding to Performance-Based Incentives: A Quasi-Experimental Comparison of Impacts on Health Service Indicators Between Hospitals and Health Centers in Malawi

Stephan Brenner , Rachel P. Chase , Shannon A. McMahon , Julia Lohmann , Christopher J. Makwero , Adamson S. Muula & Manuela De Allegri

To cite this article: Stephan Brenner , Rachel P. Chase , Shannon A. McMahon , Julia Lohmann , Christopher J. Makwero , Adamson S. Muula & Manuela De Allegri (2020) Effect Heterogeneity in Responding to Performance-Based Incentives: A Quasi-Experimental Comparison of Impacts on Health Service Indicators Between Hospitals and Health Centers in Malawi, Health Systems & Reform, 6:1, e1745580, DOI: [10.1080/23288604.2020.1745580](https://doi.org/10.1080/23288604.2020.1745580)

To link to this article: <https://doi.org/10.1080/23288604.2020.1745580>



© 2020 The Author(s). Published with license by Taylor & Francis Group, LLC.



Published online: 10 Jun 2020.



Submit your article to this journal [↗](#)



View related articles [↗](#)



View Crossmark data [↗](#)

Effect Heterogeneity in Responding to Performance-Based Incentives: A Quasi-Experimental Comparison of Impacts on Health Service Indicators Between Hospitals and Health Centers in Malawi

Stephan Brenner^{a*}, Rachel P. Chase^{a*}, Shannon A. McMahon^a, Julia Lohmann^a, Christopher J. Makwero^b, Adamson S. Muula^b, and Manuela De Allegri^a

^aHeidelberg Institute of Global Health, University Hospital and Medical Faculty, Heidelberg University, Heidelberg, Germany; ^bSchool of Public Health and Family Medicine, College of Medicine, University of Malawi, Blantyre, Malawi

ABSTRACT

Heterogeneity of effects produced by performance-based incentives (PBIs) at different levels of care provision is not well understood. This study analyzes effect heterogeneities between different facility types resulting from a PBI program in Malawi. Identical PBIs were applied to both district hospitals and health centers to improve the performance of essential health services provision. We conducted two complementary quasi-experiments comparing all 17 interventions with 17 matched independent control facilities (each 12 health centers, five hospitals). A pre- and post-test design with difference-in-differences analysis was used to estimate effects on 14 binary quality indicators; interrupted time series analysis of monthly routine data was used to estimate effects on 11 continuous quantity indicators. Effects were estimated separately for health centers and hospitals. Most quality indicators performed high at baseline, producing ceiling effects on further measurable improvements. Significant positive effects were observed for stocks of iron supplements (hospitals) and partographs (health centers). Four quantity indicators showed similar positive trend improvements across facility types (first-trimester antenatal visits, voluntary HIV-testing of couples, iron supplementation in pregnancy, vitamin A supplementation of children); two showed no change for either type of facility (skilled birth attendance, fully immunized one-year-olds); five indicators revealed different effect patterns for health centers and hospitals. In both health centers and hospitals, the largely positive PBI effects on antenatal care included resilience against interrupted supply chains and improvements in attendance rates. Observed heterogeneity might have been influenced by the availability of specific resources or the redistribution of service use.

ARTICLE HISTORY

Received 17 January 2020
Revised 18 March 2020
Accepted 18 March 2020

KEYWORDS

Performance-based incentives; impact evaluation; effect heterogeneity; quality of care; health facility

Introduction

Performance-based incentives (PBIs) entail the payment of financial and in-kind rewards to health system actors upon the achievement of pre-defined quantity and/or quality performance outputs.¹ In Sub-Saharan Africa (SSA), PBIs have been increasingly implemented in the form of Performance-Based Financing (PBF) (i.e. public sector contracting of health service providers).² The aim of PBF programs is to bolster autonomy and control among health providers in terms of how funds are earned and allocated, and to increase self-governance among health care providers by reducing their dependence on centrally allocated operational budgets.³

Recent reviews on PBF implementation in SSA have shown mixed effects in terms of expected changes in both the quantity and quality of service delivery.^{4,5} This evidence suggests that PBF success

depends on a number of factors related to both the health system macro-context (e.g. status quo of key health indicators, political stewardship, governance, decentralization, financial flows, existing purchasing structures) and the service provision micro-context (e.g. choice of performance indicators, type of service providers, degree of provider's financial autonomy, verification mechanisms, implementation capacities).⁶

As PBF programs are usually designed to respond to a given macro- and micro-context, comparison of PBF effects across different settings remains challenging.⁷ While evidence generalization is often limited given the uniqueness of each PBF program within its specific context, individual program evaluations still offer opportunities to generate specific knowledge on the effect of individual PBF designs within their given macro- or micro-contexts.

CONTACT Stephan Brenner  stephan.brenner@uni-heidelberg.de  Heidelberg Institute of Global Health, University Hospital and Medical Faculty, Heidelberg University, Im Neuenheimer Feld 130.3, Heidelberg 69120, Germany

*Joint First Authorship.

© 2020 The Author(s). Published with license by Taylor & Francis Group, LLC.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In this study, we focus on aspects related to the micro-context — namely the extent to which PBF effects differ between sub-district level health centers and district-level hospitals—in the case of the *Support for Service Delivery Integration Performance-Based Incentives* (SSDI-PBI) program in Malawi. This program primarily intended to improve the provision of primary health services defined by the country's Essential Health Package (EHP); hence, both health centers and hospitals were subject to identical performance indicators, verification and reimbursement mechanisms, as well as procurement regulations. Our study objective was not only to estimate the impact of SSDI-PBI on a number of quality and quantity service indicators but to more specifically assess similarities and differences in observed PBI effects between hospitals and health centers, thereby further contributing to understanding on the micro-context.

Methods

Macro Context

Malawi is a low-income country in SSA, which suffers from a heavy burden of HIV and communicable diseases (especially tuberculosis and malaria) and, more recently, increases in non-communicable diseases (i.e. hypertension, diabetes, cancer).⁸ Over the past decade, there has been a downward trend in maternal, newborn, and child mortality,^{9,10} with a recent maternal mortality ratio of 439 deaths per 100,000 live births, neonatal mortality rate of 27 deaths per 1,000 live births, and under-five mortality rate of 63 deaths per 1,000 live births.¹¹ Care during pregnancy and birth has improved in the past decade,¹⁰ with about 51% of pregnant women attending at least four antenatal care (ANC) visits, 24% having a first ANC visit during their first pregnancy trimester, 33% taking iron tablets for at least 90 days during pregnancy, 63% taking at least two doses of preventive malaria treatment during pregnancy, and 90% of births being attended by a skilled birth attendant (SBA).¹¹ Similar upward trends have been achieved for HIV care and family planning,¹⁰ with about 44% of women and 42% of men having been tested for HIV during the past 12 months, 58% of married women of reproductive age using modern contraception, and 49% of currently married women and 43% of currently married men aged 15–49 wanting to limit childbearing.¹¹ In terms of child nutrition and immunization, 64% of under-five children have received vitamin A supplements in the past 6 months, and 70% of under-one children are fully immunized.¹¹ 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, and support of service delivery.^{12–14}

In 2014, Malawi counted 1060 formal-sector health facilities (including 509 government-owned facilities) offering primary health services, of which 489 were health centers and 119 hospitals.¹² Primary health care services included in the EHP cover reproductive and child health, and the prevention, detection, and management of infectious and non-communicable diseases.¹⁵ EHP services are intended to be provided free of charge at the point of use in both public and private not-for-profit facilities contracted by the Ministry of Health (MoH).¹⁵ Evidence indicates, however, that EHP services are not as effectively available as they should be, thereby subjecting clients to substantial out-of-pocket expenditures.¹⁶

Malawi's centralized tax-based health system receives most of its funding from external donors.¹⁵ In 2013, Malawi's total health expenditure was 39 USD per capita or nearly 11% of GDP, with about 18% from public, 14% from private, and 68% from donor sources.¹⁷ Annual budgetary support from both central-level ministries and local governments is allocated to each district where funds are managed and allocated to health facilities by District Health Management Teams (DHMTs).¹⁸ Health workers employed by government health facilities receive salaries paid directly by the central MoH.

Micro Context

For both health centers and district hospitals, the DHMTs serve as a management hub with respect to service delivery, including the coordination of drug budgets, provision of equipment, assignment of clinical staff. However, compared to district hospitals, health centers have a different size and composition of clinical staff (doctors/clinical officers: median 2 versus 7 in hospitals; nurses: median 2 versus 18; health surveillance assistants: median 13 versus 19).¹² Given their generally more remote location compared to hospitals, more health centers have only unreliable access to electricity and running water, as well as less reliable supply and procurement chains in respect to essential drug and supply stocks. Compared to district hospitals, health centers generally cover a more remote catchment area and have to rely more on outreach activities.

While in theory, all facilities develop their own business strategies toward improving service coverage for each quantity indicator and improvement along with quality scores in collaboration with their respective

DHMT, implementation research demonstrated that the highly regulated procurement processes failed to support facilities in guaranteeing complete drug stocks and other timely investments.¹⁹

PBF Implementation

In 2011, the *Support for Service Delivery Integration* (SSDI) project, a United States-sponsored bilateral health sector program to strengthen EHP service provisions, started at 301 health facilities in 15 of Malawi's 28 districts.²⁰ In August 2014, SSDI together with the MoH launched a PBF program (referred to as SSDI-PBI) in only three of these 15 districts (Chitipa, Mangochi, Nkhoswe) introducing PBIs to only 17 government-owned facilities (12 health centers, five hospitals) out of a total of 53 SSDI facilities in these three districts.²¹ Selection of these 17 facilities was nonrandom and based on the following criteria: minimum of four qualified health staff per facility, provision of all primary level EHP services, availability of essential equipment and infrastructure, and facilities' prior inclusion in the SSDI-sponsored *Performance Quality Improvement* (PQI) program (a standards-based quality improvement approach that emphasizes root causes and provider-led solutions to address poor performance).²²

Facility performance was defined by 13 quantity indicators (see [Table 1](#)) covering six health care domains: antenatal, obstetric, postnatal, under-five child, family planning, as well as HIV and AIDS care. A checklist consisting of 107 quality indicators assessed performance related to facility organization and patient satisfaction with respect to EHP service provision (e.g. service management, routine performance reporting, hygiene and sanitation, laboratory maintenance, drug, and commodity management). An overall PBI envelope was determined annually for each facility. Within this envelope, financial incentives were paid in form of performance bonuses reflecting a combination of fee for service payments (i.e. reimbursement of the number of services provided for each quantity indicator by an annually set service unit fee) and target-based payments (i.e. proportional reimbursement based on the extent facilities reached preset targets for quantity and quality performance). The performance was verified by peer review and counter verified by an independent private firm with payments scheduled on a 6-monthly basis.²¹

For each 6-monthly payment cycle, facilities proposed investment activities outlined in their annual business plans to be procured and/or implemented by SSDI after approval and in alignment with the regulatory procedures stipulated by donor and implementer

(e.g. tender procedures for open market purchases).²¹ Unique to this PBI program, no portion of the PBI rewards could be used as individual bonus payments for facility staff.

Study Objectives and Design

Our study objectives align with the theory of change underlying SSDI-PBI, which expected additional rewards earned by facilities based on their performance to enable health centers and district hospitals in identical ways to more strategically invest into inputs and processes related to quality EPH service delivery and to guarantee successful achievement of service outputs over time (see [Figure 1](#)). Based on these expected changes, we aimed to explore whether patterns of change reflected by SSDI-PBI effects on selected service indicators differed between health centers and hospitals.

To make the best use of the data available at the start of this study, our evaluation was built on a quasi-experimental design using a multi-methods approach,^{23,24} which relied on two parallel analytical strategies: (a) a controlled pre- and post-test analysis of 14 binary quality indicators available through primary and secondary data and (b) a controlled interrupted time series (ITS) analysis of 11 quantity indicators available through secondary routine health facility data.

Pre- and Post-Test Study Component

Outcome Variables

For baseline quality indicators, we relied on dichotomous data from the Service Provision Assessment (SPA) survey conducted in Malawi between 2013/2014.¹² Reviewing the SPA database, we selected 14 quality indicators directly or indirectly relating to the 13 quantity indicators (shown in [Table 1](#)) we expected to improve given the SSDI-PBI theory of change. These quality indicators represented measures of service and staff organization (i.e. supervision, management meetings, patient feedback) or of the availability of supplies and medicines essential to EHP provision.

Sample

The sample included all 17 SSDI-PBI and 17 control facilities (one matched control per each SSDI-PBI facility). The following criteria were used to match facilities at baseline: facility type, general service support through SSDI, government ownership, PQI enrollment, as well as similarities in geographic characteristics, catchment area size, and physical accessibility. Within-district



Table 1. List of quantity indicators incentivized by the SSDI-PBI together with related quantity and quality outcome variables included in our study.

SSDI-PBI quantity performance indicators	HMS quantity indicators used as outcome variables in ITS analysis	SPA quality indicators used as outcome variables in DID analysis
Antenatal Care (ANC):		
Number of pregnant women starting antenatal care during first trimester.	Number of pregnant women who started ANC visits in pregnancy weeks 0–12.	-/-
Number of pregnant women completing four ANC visits.	Number of pregnant women with a total of four ANC visits.	-/-
Number of pregnant women attending ANC receiving iron supplementation.	Number of pregnant women who received 120+tablets of ferrous-folate during ANC visit.	Proportion of facilities with at least one supplement of iron available
Number of pregnant women receiving at least two doses of IPTp.	Number of pregnant women who received two times the tablets of sulphadoxine-pyrimethamine.	Proportion of facilities with at least one treatment of sulphadoxine-pyrimethamine available
Birth and Postnatal Care (PNC):		
Number of births attended by a skilled birth attendant.	Number of births with skilled birth attendance.	Proportion of facilities with at least one uterotonic treatment available Proportion of facilities with at least one delivery pack available at maternity unit
Number of women who received postnatal care within seven days after delivery by a skilled health worker.	Number of postpartum care visits within 2 weeks of delivery.	Proportion of facilities with at least one Pentavalent vaccine available
Under-Five Child Care:		
Number of one-year-old children fully immunized.	Number of under-one-year-olds fully immunized.	Proportion of facilities with at least one oral polio vaccine available
Number of children receiving Vitamin A supplementation.	Number of children age 6–59 months who received a dose vitamin A.	Proportion of facilities with at least one measles vaccine available
Care related to the HIV/AIDS prevention:		
Number of HIV-positive pregnant women who were initiated on ART.	Number of pregnant women started on ART in pregnancy.	Proportion of facilities with at least one rapid HIV test kit available
Number of infants born by HIV positive mothers tested for HIV	Number of clients tested with partner during VTC.	-/-
Number of couples tested for HIV during VTC services.	Number of new and old clients counseled for family planning.	-/-
Number of HIV/AIDS cases screened for tuberculosis.		
Care related to family planning:		
Number of clients counseled for family planning.		
Selected aspects of service organization and management:		
-/-		Proportion of facilities with at least one treatment of injectable contraceptive available
-/-		Proportion of facilities having received any external supervision within past 6 months.
-/-		Proportion of facilities having held any management meetings within past 6 months.
-/-		Proportion of facilities with client feedback system in place

ART = antiretroviral therapy; BCG = *Bacillus Calmette-Guérin*; DID = difference-in-differences; HMIS = Health Management Information System; IPTp = intermittent preventive treatment for malaria during pregnancy; ITS = interrupted time series; SPA = Service Provision Assessments; SSDI-PBI = Support for Service Delivery Integration Performance-Based Incentives; VTC = voluntary HIV testing and counseling.

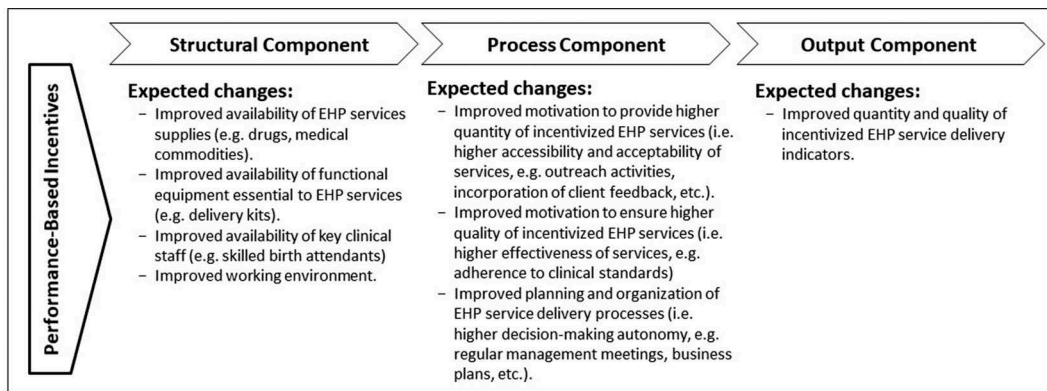


Figure 1. Schematic outline of the theory of change relating to the SSDI-PBI.

matches were preferred; out-of-district matches in nearby SSDI-districts were made when no comparable control facilities within the PBI district could be identified. The resulting control sample consisted of 12 health centers and five hospitals in eight districts (see [Table 2](#)).

Data

We extracted baseline data for selected quality indicators from the publicly available Malawi 2013/2014 SPA database. Endline data for each of the indicators were obtained by primary data collection conducted by our study team in March 2016 (i.e. 20 months after the official SSDI-PBI launch). Primary data were collected by trained research assistants using an abbreviated version of the SPA facility survey instruments.

Table 2. Facility sample characteristics of SSDI-PBI and matched controls.

	SSDI-PBI sample	Control sample
Total number of facilities	17	17
Facility type		
Health centers	12	12
Hospitals	5	5
Facility ownership		
Government (public)	17	17
Facility distribution by district:		
Chitipa	3	1
Mangochi	8	3
Nkhotakota	6	4
Karonga		4
Kasungu		2
Lilongwe		1
Salima		1
Zomba		1
Facility distribution by region:		
Northern Region	3	5
Central Region	6	8
Southern Region	8	4
Facility catchment population	525,153	574,033
(in persons) ^a		
SSDI project sites	17	17
Enrollment in PQI program	14	14

DiD = difference-in-differences; *ITS* = interrupted time series; *PQI* = Performance Quality Improvement; *SSDI-PBI* = Support for Service Delivery Integration Performance-Based Incentives.

^aEstimates for 2014 taken from SSDI project rosters.

Analysis

Data on quality indicators represented a fully balanced panel. Data were analyzed using a Difference-in-Differences (DiD) approach based on linear regression to estimate the impact of the PBI program on observed changes between SSDI-PBI and control facilities between baseline and endline, specified as:

$$y = \beta_0 + \beta_1 t + \beta_2 g + \beta_3 t^* g + \varepsilon,$$

with y representing the outcome, t the time point (0 = baseline, 1 = endline), and g the treatment group (0 = control, 1 = PBI). β_3 is the effect estimate attributable to the PBI. Models were further adjusted for fixed effects at the facility level.

Interrupted Time-Series Study Component

Outcome Variables

Service utilization measures represented by the 13 PBI key performance indicators served as outcome variables, as data on these indicators are reported monthly to the Malawi Health Management Information System (HMIS) accessed and extracted through the DHIS-2 platform. While HMIS data quality has been relatively poor, the data completeness and consistency for the selected indicators and time period compared to other HMIS reported indicators were deemed relatively high by the research team. Nevertheless, for two of these 13 indicators, data were largely missing, which reduced the number of outcome variables used in this component to 11 (see [Table 1](#)).

Sample

We used the same SSDI-PBI and control facilities as for the pre- and post-test component outlined above to facilitate the synthesis of results across the two methods. We extracted monthly data on the 11 outcome variables for each sampled facility for the period

August 2013 to February 2016 (i.e. a total of 31 consecutive months).

Data

After data extraction, missing data points for single months were identified across all indicators and facilities, and only facilities with less than 5% of missing values were included in the analysis of a given indicator. As a result, one facility for indicator 6 (number of births at facility attended by a skilled birth attendant) and four facilities had to be omitted for indicator 11 (number of clients attending family planning services counseled on contraceptive options).

HMIS routine data were only available as absolute numbers of services provided per month (e.g. total number of pregnant women completing the fourth antenatal care visit in a given month and facility) without information on respective reference populations (i.e. total number of pregnant women in the catchment area in a given month). We therefore computed pre-intervention means for each facility and each outcome variable by averaging the monthly counts for the period from September 2013 to July 2014. These means then served as denominators for each observed monthly count in the time series with each monthly ratio representing the relative change from the pre-intervention period (i.e. ratios above 1 indicating an increase in facility performance). This transformed data allowed us to compare each facility against its own performance baseline and across facilities.

Analysis

Data were analyzed using a multiple-group segmented linear regression,²⁵ comparing facility performance for selected outcome variables between SSDI-PBI and control facilities and between pre-intervention (August 2013 until July 2014) and post-intervention (August 2014 until February 2016) periods, specified as:

$$y_t = \beta_0 + \beta_1 T_t + \beta_2 g + \beta_3 g T_t + \beta_4 x_1 + \beta_5 x_1 T_t + \beta_6 g x_1 + \beta_7 g x_1 T_t + \varepsilon_t,$$

with y_t representing the outcome variable measured at each monthly time point t , variable T_t the number of months since the start of the time series, x the intervention periods ($0 =$ pre-intervention, $x_1 =$ post-intervention), and g the treatment group ($0 =$ control, $1 =$ PBI). In this model, β_2 and β_3 indicate the estimated differences in level (intercept) and slope (trend), respectively, in outcome y between treatment and control facilities prior to intervention start, and β_6 and β_7 the estimated differences in level and slope,

respectively, attributable to the PBI in the post-intervention period.

We used the PBI launch in August 2014 as an interruption time point. The presence of serial autocorrelation up to a lag of 1 month was demonstrated by the Cumby–Huizinga test for auto-correlation²⁶; we adjusted the model accordingly. The ITS model was further adjusted for seasonality based on the following seasonal categories: warm wet season (December to February), warm dry season (March to May), cool dry season (June to August), and hot dry season (September to November).²⁷

Results

The different sample sizes for hospitals and health centers resulted in the distortion of actual effect sizes and thus make a direct comparison of effect sizes between facility types somewhat meaningless. To overcome this challenge and to still allow direct comparison of effects between health centers and hospitals, we describe effects for each facility type in the form of effect patterns which we then compare and discuss later on.

DiD Analysis Results of Quality Indicators

Control and intervention facilities did not significantly differ at baseline for any of the 14 quality indicators. In the following, we present results along with typologies of effect patterns, as also shown in Table 3 (i.e. “saturated,” “protective,” “challenged,” “globally improved,” and “no substantial effect”), to aid comparison of effects in health centers versus hospitals.

“Saturated” effect patterns (defined as quality indicators achieved in at least 80% of facilities at baseline and endline in both intervention and control groups) were found for seven of the 14 quality indicators (4, 6, 7, 8, 9, 13, and 14) for health centers. For hospitals, 12 out of the 14 indicators showed a saturated pattern (1, 2, 4, 6, 7, 8, 9, 10, 11, 12, 13, and 14), including all of the indicators that presented a saturated pattern among health centers. For health centers, the estimated improvement produced by SSDI-PBI among health centers on indicator 8 (available stock of blank partograph forms) was statistically significant despite the saturated effect pattern due to declines among controls and increases among intervention facilities within this upper range of facility proportions.

“Protective” effect patterns (defined as quality indicator levels that declined from at least 80% of facilities at baseline to fewer than 80% of facilities at endline in both types of facilities with greater declines in the

Table 3. SSDI-PBI effects on dichotomous quality indicators. Computed proportions and estimated effect sizes based on difference-in-differences analysis.

Quality indicator	Health centers only				Hospitals only			
	SSDI-PBI (%)	Control (%)	DiD (SE) (%-points)	Effect pattern	SSDI-PBI (%)	Control (%)	DiD (SE) (%-points)	Effect pattern
1 Proportion of facilities having received any external supervision within past 6 months.	Before 75.0	58.3	-25.0 (15.7)	Globally improved	100	100	0.0 (-)	Saturated
	After 83.3	91.7			100	100		
2 Proportion of facilities having held any management meetings within past 6 months.	Before 75.0	58.3	-8.3 (25.9)	Globally improved	100	100	-20.0 (17.1)	Saturated
	After 83.3	75.0			80.0	100		
3 Proportion of facilities with client feedback system in place	Before 58.3	33.3	-25.0 (10.9)*	Globally improved	60.0	80.0	20.0 (38.8)	No effect
	After 75.0	75.0			60.0	60.0		
4 Proportion of facilities with at least one treatment of sulphadoxine-pyrimethamine available	Before 100	100	0.0 (-)	Saturated	100	100	0.0 (-)	Saturated
	After 100	100			100	100		
5 Proportion of facilities with at least one supplement of iron available	Before 100	100	16.7 (15.9)	Protective	100	100	80.0 (22.4)**	Protective
	After 41.7	25.0			80.0	0.0		
6 Proportion of facilities with at least one uterotonic treatment available	Before 91.7	100	8.3 (11.2)	Saturated	100	80.0	-20.0 (17.1)	Saturated
	After 83.3	83.3			80.0	80.0		
7 Proportion of facilities with at least one delivery pack available at maternity unit	Before 91.7	100	16.7 (10.9)	Saturated	100	100	0.0 (23.3)	Saturated
	After 100	91.7			80.0	80.0		
8 Proportion of facilities with at least one blank partograph form available at maternity unit	Before 83.3	100	25.0 (6.5)**	Saturated	100	100	0.0 (23.3)	Saturated
	After 100	91.7			80.0	80.0		
9 Proportion of facilities with at least one Pentavalent vaccine available	Before 91.7	91.7	8.3 (9.9)	Saturated	100	80.0	-20.0 (36.5)	Saturated
	After 91.7	91.7			80.0	80.0		
10 Proportion of facilities with at least one oral polio vaccine available	Before 75.0	75.0	8.3 (12.2)	Globally improved	80.0	100	20.0 (15.8)	Saturated
	After 91.7	83.3			80.0	80.0		
11 Proportion of facilities with at least one measles vaccine available	Before 100	100	16.7 (12.2)	Protective	100	100	0.0 (23.3)	Saturated
	After 91.7	75.0			80.0	80.0		
12 Proportion of facilities with at least one BCG vaccine available	Before 91.7	100	-16.7 (15.8)	Challenged	100	100	0.0 (23.3)	Saturated
	After 66.7	91.7			80.0	80.0		
13 Proportion of facilities with at least one rapid HIV test kit available	Before 83.3	91.7	8.3 (10.2)	Saturated	100	100	0.0 (-)	Saturated
	After 100	100			100	100		
14 Proportion of facilities with at least one treatment of injectable contraceptive available	Before 100	91.7	-16.7 (13.9)	Saturated	100	80.0	-40.0 (27.4)	Saturated
	After 91.7	100			80.0	100		

* $p < 0.1$; ** $p < 0.05$ BCG = *Bacillus Calmette-Guérin*; DiD = difference-in-differences; PMTCT = prevention of mother-to-child transmission; SE = standard error; SSDI-PBI = Support for Service Delivery Integration Performance-Based Incentives.

control compared to the intervention group) were observed for indicator 5 (available stock of iron supplements) in both health centers and hospitals and for indicator 11 (available stock of measles vaccine) in health centers only. For hospitals, the effect size produced by SSDI-PBI on indicator 5 was statistically significant.

“Challenged” effect patterns (defined as quality indicators achieved in at least 80% of both intervention and control facilities at baseline, but falling to fewer than 80% of facilities in only the intervention group at endline) included only quality indicator 12 (available stock of BCG vaccines) in health centers, indicating a negative (though not statistically significant) effect of the SSDI-PBI.

“Globally improved” effect patterns (defined as indicator levels that showed room for improvement at baseline and improved among both intervention and control facilities at endline) were observed for quality indicators 1, 2, 3, and 10 among health centers, with control facilities showing more improvement than intervention facilities in all but quality indicator 10 (available stock of oral polio vaccines), for which intervention facilities showed slightly more improvement.

We observed a “no substantial effect” pattern (defined as quality indicator levels that showed room for improvement at baseline but with little meaningful change in either intervention or control groups at endline) only for quality indicator 3 (having a client feedback system in place) among hospitals.

Baseline Comparability of Intervention and Control Facilities for ITS Analysis

In conducting the ITS analysis, we compared changes in quantity indicators before and after the PBI intervention (i.e. monthly performance versus mean baseline performance). Table 4 presents the baseline averages calculated for the entire pre-intervention period (Sept 2013–July 2014) by study arm for the two facility types and for the eleven quantity indicators used in the ITS analysis.

Within each arm of the study, large standard deviations indicate high variation around variable means. Average monthly numbers for all indicators were generally higher in the hospital compared to the health center group, reflecting the higher patient volume at hospital levels. Baseline group means for many variables had statistically significant differences between intervention and control arms within each facility group, with means generally higher for SSDI-PBI health centers and lower for SSDI-PBI hospitals compared to their respective controls.

Table 4. Comparison of non-zero averages during the pre-intervention period (Aug 2013–July 2014) by intervention arm and facility type for outcome variables used in interrupted time series analysis.

Quantity indicator	Health centers only						Hospitals only					
	SSDI-PBI			Control			SSDI-PBI			Control		
	Mean	SD	<i>p-value</i> ^a	Mean	SD	<i>p-value</i> ^a	Mean	SD	<i>p-value</i> ^a	Mean	SD	<i>p-value</i> ^a
1 Average monthly number of pregnant women attending first ANC service during their first trimester.	8.1	6.6	7.0	5.9	5.9	.21	21.5	13.4	13.4	21.6	12.6	.97
2 Average monthly number of pregnant women having attended at least four ANC visits during pregnancy.	18.9	13.1	14.2	12.5	12.5	.01	37.8	31.8	31.8	44.3	18.7	.20
3 Average monthly number of pregnant women attending ANC who received iron supplementation.	34.1	31.7	27.6	20.4	20.4	.07	93.3	67.8	67.8	77.6	45.0	.17
4 Average monthly number of pregnant women having received at least two doses of IPTp during pregnancy.	63.2	30.1	47.7	31.9	31.9	<.01	141.7	106.2	106.2	173.6	86.2	.09
5 Average monthly number of births at facility attended by a skilled birth attendant.	63.1	34.0	41.1	29.1	29.1	<.01	299.3	217.4	217.4	353.4	194.0	.15
6 Average monthly number of postpartum women receiving PNC from a skilled birth attendant within 14 days after delivery.	53.7	38.0	33.6	22.8	22.8	<.01	81.9	97.1	97.1	149.9	129.2	.01
7 Average monthly number of one-year-old children attending pediatric services fully immunized.	84.0	41.7	56.4	33.0	33.0	<.01	146.3	72.3	72.3	210.5	111.5	.03
8 Average monthly number of children attending pediatric services having received Vitamin A in past 6 months. ^b	52.2	45.6	105.2	393.4	393.4	.67	75.0	93.6	93.6	834.3	1758.7	.03
9 Average monthly number of HIV-positive pregnant women initiated on ART during pregnancy.	2.8	2.1	4.2	4.0	4.0	.01	9.2	8.3	8.3	8.4	6.1	.58
10 Average monthly number of couples tested for HIV during voluntary counseling and testing services.	29.8	48.5	27.8	33.2	33.2	.75	57.3	123.0	123.0	176.2	132.0	<.01
11 Average monthly number of clients attending family planning services counseled on contraceptive options.	385.6	1166.0	157.1	156.1	156.1	.06	559.5	478.4	478.4	619.2	373.8	.50

^a*p*-values based on *t*-test comparing group means for each indicator.

^b3-monthly (instead of monthly) time intervals, given periodic outreach activities.

ANC = antenatal care; ART = antiretroviral therapy; IPTp = intermittent preventive treatment for malaria during pregnancy; PNC = postnatal care; SSDI-PBI = Support for Service Delivery Integration Performance-Based Incentives.

ITS Estimated Time Trends in Quantity Indicators

Similar to using the DiD results to specifically observe heterogeneities of PBI effects on service quality indicators, we used the ITS results to identify effect heterogeneities in service quantity indicators. The graphs in [Figures 2 to 12](#) present the predicted time series trends adjusted by seasonality for each indicator by facility type. [Table 5](#) presents effect estimates adjusted for seasonality for each quantity indicator resulting from the ITS analysis by facility type. Again, we present effect pattern typologies (i.e. “no substantial effect,” “recovering,” “improved level,” “improved trend,” “intervention period alignment”) when describing our results along with these quantity indicators. Notice, however, that we do not have information about whether quantity indicators have reached saturation (whether there is room for improvement or not), so these typologies are determined primarily by whether notable trend (slope) changes or level (immediate effect) changes were estimated.

“No substantial effect” patterns (defined as quantity indicator values that change little relative to baseline mean values for both intervention and control facilities) were observed for indicators 5 and 7 in both health centers and hospitals, and indicator 11 in health centers.

We observed a “recovering” effect pattern (defined as quantity indicator values that demonstrate irregular patterns, often with pre-intervention declines that reverted during the post-intervention period leading to relatively stronger upward trends among intervention facilities than among controls) for indicators 3 and 8 in health centers and hospitals (statistically significant trend changes only estimated for indicator 8) and indicator 4 in hospitals (trend changes statistically significant). Similar to quality indicator 5 on available stocks of iron supplementation above, these patterns reinforce the finding of a protective SSDI-PBI effect.

An “improved level” effect pattern (defined as quantity indicator values manifesting an immediate and sustained improvement in intervention compared to control facilities) was observed for indicator 2 in hospitals and for indicator 9 in health centers, both with statistically significant level changes. The reverse “diminished level” effect was observed for indicator 11 among hospitals, where controls showed a greater level of improvement compared to interventions.

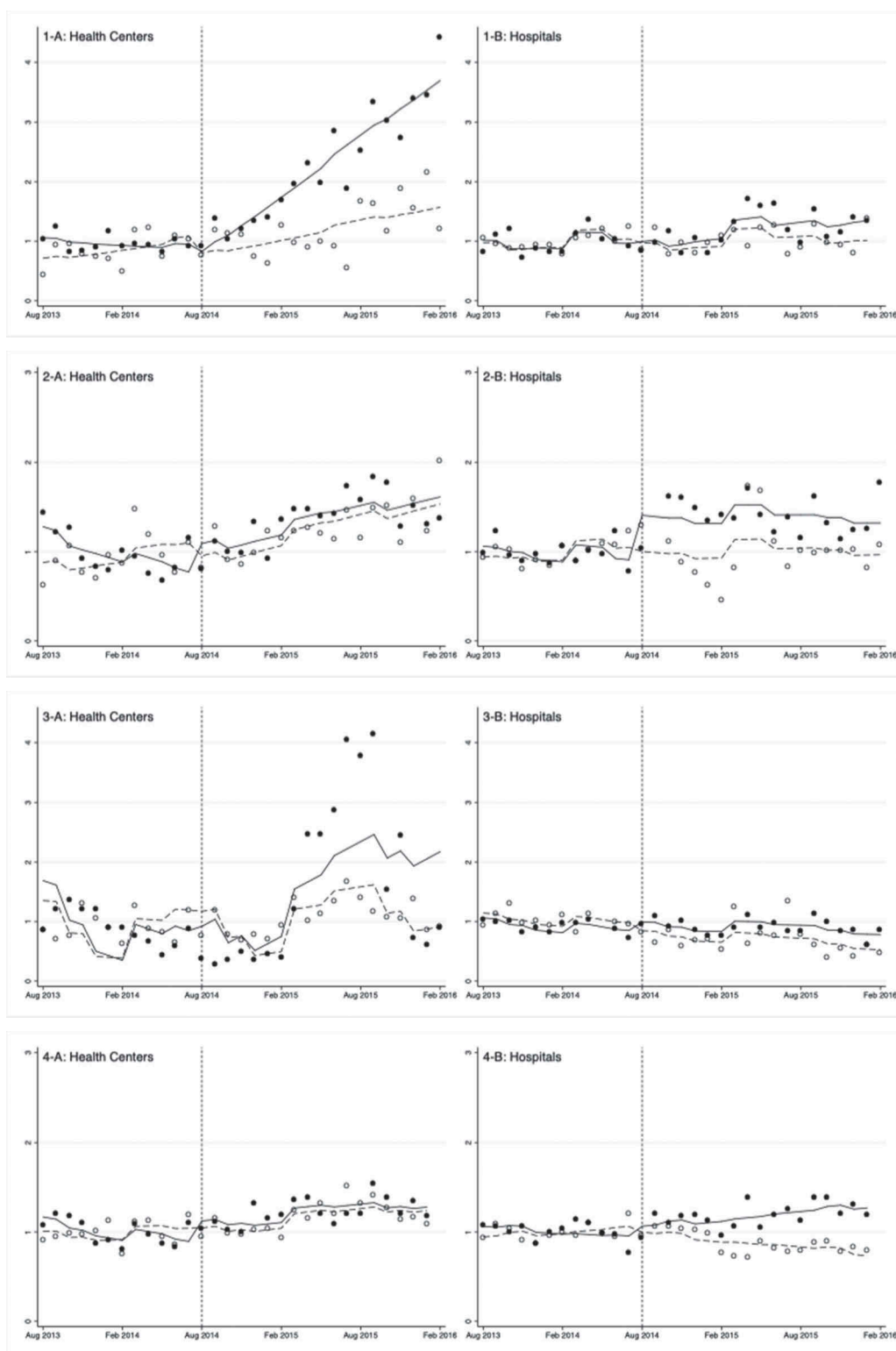
An “improved trend” effect pattern (defined as quantity indicator values that do not so much demonstrate an immediate improvement, but greater improvements in intervention trends compared to controls) was observed for indicators 1 and 10 among both

health centers and hospitals. These effect sizes of trend changes are statistically significant except for indicator 1 among hospitals. Indicator 10 controls hospital values show a level increase during the intervention period and then decline while intervention hospital values appear to rise more gradually, thus exceeding control values. Further, quantity indicator 6 shows opposing trend patterns for two facility types: trend improvements in intervention hospitals compared to controls, and diminished trend in intervention health centers compared to controls.

“Intervention period alignment” effect patterns (defined as intervention and control quantity indicator levels that are less comparable during the pre-intervention compared to the post-intervention period), such as in the case of indicators 2 and 4 among health centers where interventions, but not controls, show declining values at baseline. Statistically, this upward turn among interventions to catch up to and even exceed control trends during the intervention period resulted in a statistically significant level or trend effects. The underlying data, however, demonstrate that intervention and controls simply align their divergent trends with one another. A similar phenomenon with statistical significance in level change is seen among hospitals for indicator 9.

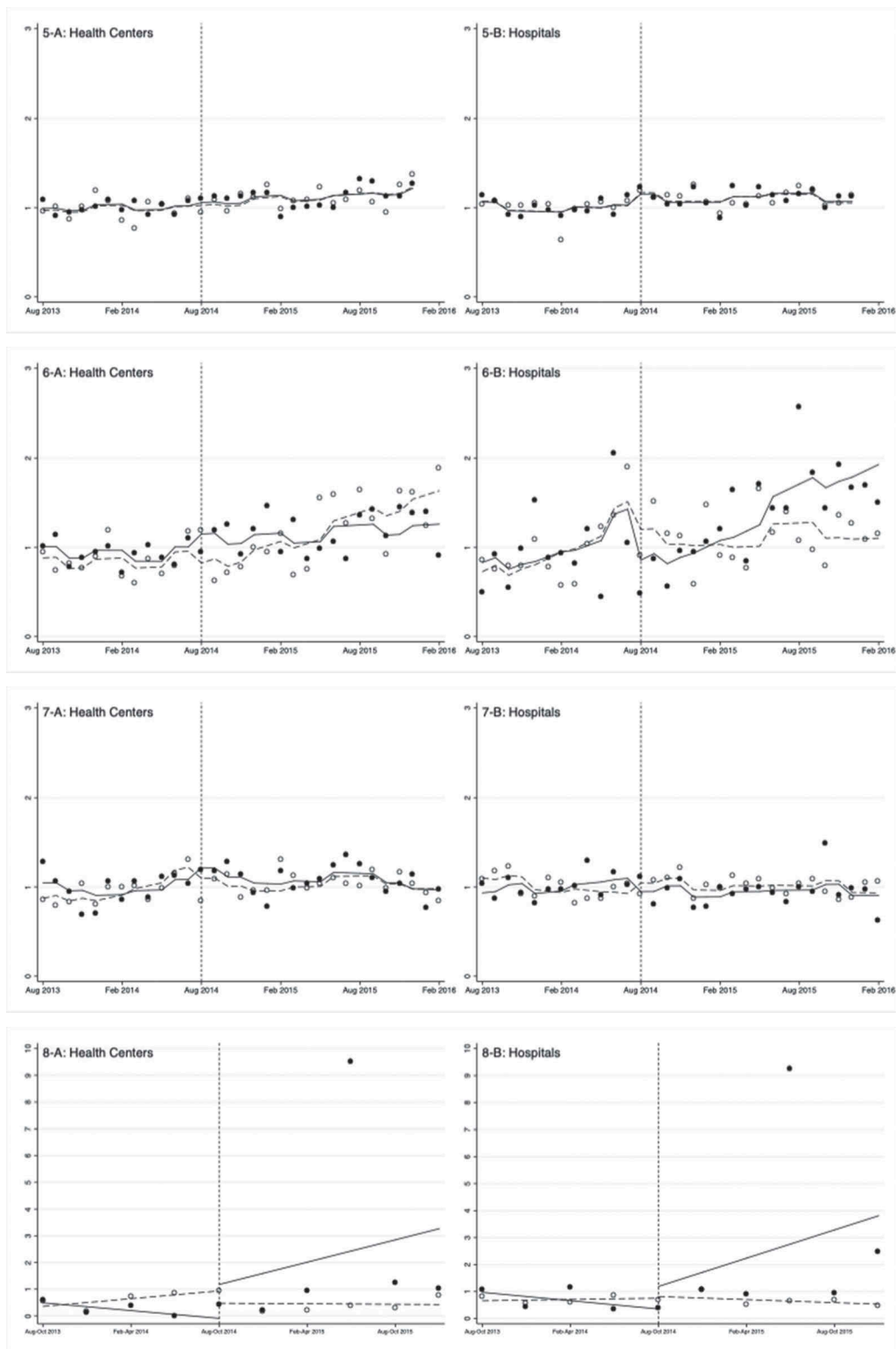
In summary, for six of the eleven quantity indicators, both health centers and hospitals follow similar patterns: no substantial effects in indicators 5 (number of births attended by a skilled provider) and 7 (number of one-year-olds fully immunized); trend improvements in indicators 1 (number of pregnant women with first ANC visit in the first trimester) and 10 (number of couples with voluntary HIV counseling and testing); and recovering effect patterns in indicators 3 (number of ANC-attending women who received iron supplements) and 8 (number of pediatric patients who received vitamin A in past 6 months).

The remaining five indicators show no obvious common pattern regarding their differences. For quantity indicator 2 (number of pregnant women with at least four ANC visits), while health centers in both study arms appear to be on a similar upward trend throughout the intervention period, hospitals diverge at the interruption time point and remain relatively steady at those levels throughout the intervention period. For indicator 4 (number of pregnant women with at least two IPTp doses), intervention health centers, control health centers, and intervention hospitals all seem to display similar trends with only control hospitals trending toward lower values throughout the post-intervention period. For quantity indicator 6 (number of postpartum women receiving PNC) control trends appear to exceed



Figures 2–12. Predicted time-series trends based on segmented linear regression for each quantity indicator by facility type. **Figure 2.** Number of pregnant women attending the first ANC service during their first trimester. **Figure 3.** Number of pregnant women having attended at least four ANC visits during pregnancy. **Figure 4.** Number of pregnant women attending ANC who received iron supplementation. **Figure 5.** Number of pregnant women having received at least two doses of IPTp during pregnancy. **Figure 6.** Number of births at the facility attended by a skilled birth attendant. **Figure 7.** Number of postpartum women receiving PNC from a skilled birth attendant within 14 days after delivery. **Figure 8.** Number of one-year-old children attending pediatric services fully immunized. **Figure 9.** Number of children attending pediatric services having received vitamin A in past 6 months. **Figure 10.** Number of HIV-positive pregnant women initiated on ART during pregnancy. **Figure 11.** Number of couples tested for HIV during voluntary counseling and testing services. **Figure 12.** Number of clients attending family planning services counseled on contraceptive options.

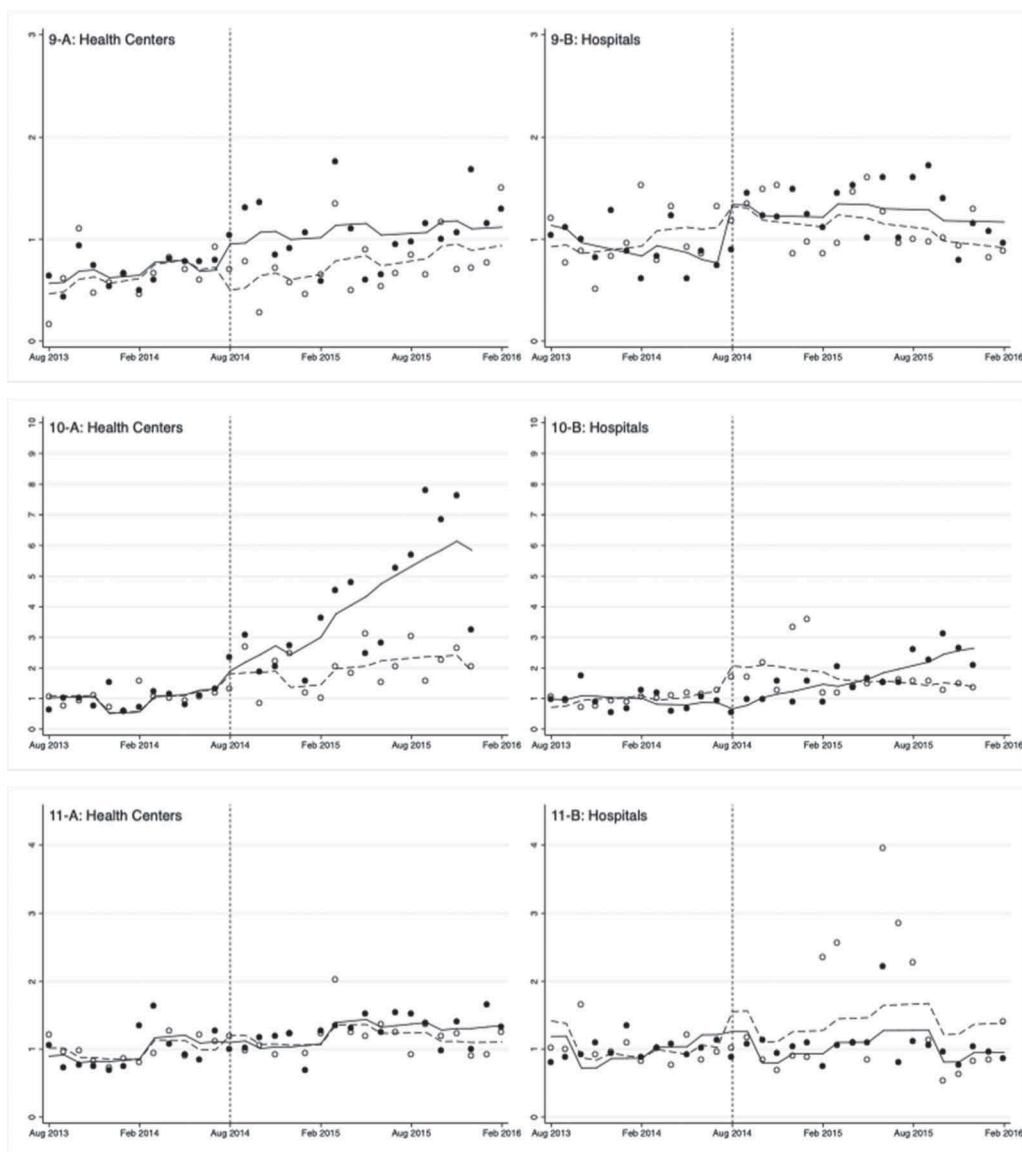
x-axis: months within time series, except for Indicator 8 which uses 3-month intervals; y-axis: change from baseline average, ratio of 1 = no change.



Figures 2. (Continued).

intervention trends in health centers, but intervention trends exceed control trends in hospitals. For quantity indicator 9 (number of HIV-positive pregnant women on ART), values started and stayed low among control health centers, but the similarly low intervention health

centers at baseline improved after the intervention. For quantity indicator 11 (number of clients counseled on contraceptive options) no effect was seen among health centers, but control hospitals appear to outperform intervention hospitals throughout the intervention period.



Figures 2. (Continued).

Discussion

The objective of this study was to assess SSDI-PBI effect patterns with respect to quality and quantity performance related to EHP service delivery in the specific context where identical PBI mechanisms were applied to both health centers and hospitals. To do so, we employed a multi-method evaluation design including two different analytical approaches to make the best use of available primary and secondary data. Along with this multi-methods approach, our study focus is unique in that we examined the heterogeneity of effects between facility types (i.e. health centers, hospitals). In this regard, the discussion and appraisal of our findings in light of existing quantitative evidence are relatively constrained, as the literature on PBI evaluations

exploring effect heterogeneity between different facility types or levels of care is currently limited, both in number and geographical focus.

The pre- and post-test comparison revealed a strong protective effect of the SSDI-PBI on stocks of iron supplements (quality indicator 5, more pronounced among hospitals) and to some degree on the stocks of delivery packs and partograph forms (quality indicators 7 and 8) among health centers during a period when nationwide supply shortages due to a central government-level corruption scandal were faced by both health centers and hospitals.²⁸ Evidence from another PBF program evaluation in Malawi—the RBF4MNH Initiative, implemented by the MoH in four different districts (Balaka, Dedza, Mchinji, Ntcheu) around the same time as the SSDI-PBI—indicated a similar

Table 5. SSDI-PBI effects on service utilization quantity indicators. Estimates based on interrupted time series analysis; coefficients retrieved by multiple-group segmented linear regression adjusted for seasonality.

Quantity indicator ^a	Coefficients	Estimated SSDI-PBI effects, health centers effect size (95%-CI)	Effect pattern	Estimated SSDI-PBI effects, hospitals effect size (95%-CI)	Effect pattern
1 Number of pregnant women attending first ANC service during their first trimester.	Pre-intervention period:				
	Control slope (β_1)	0.03 (0.00– 0.06) *		0.01 (–0.01– 0.03)	
	Difference PBI versus control in slopes (β_3)	–0.04 (–0.07 – –0.01) *		–0.01 (–0.04– 0.02)	
	Post-intervention period:				
2 Number of pregnant women having attended at least four ANC visits during pregnancy.	Difference PBI versus control in level change (β_6)	0.21 (–0.23– 0.64)	Trend improved	0.11 (–0.17– 0.39)	Trend improved
	Difference PBI versus control in slope change (β_7)	0.16 (0.11– 0.21) **		0.03 (–0.01– 0.06)	
	Pre-intervention period:				
	Control slope (β_1)	0.02 (–0.02– 0.06)		0.01 (–0.01– 0.03)	
3 Number of pregnant women attending ANC who received iron supplementation.	Difference PBI versus control in slopes (β_3)	–0.06 (–0.12 – –0.01) *		–0.02 (–0.05– 0.01)	
	Post-intervention period:				
	Difference PBI versus control in level change (β_6)	0.53 (–0.01– 1.07)	Alignment	0.70 (0.25– 1.14) **	Level improved
	Difference PBI versus control in slope change (β_7)	0.06 (0.01– 0.12) *		0.01 (–0.04– 0.06)	
4 Number of pregnant women having received at least two doses of IPTp during pregnancy.	Pre-intervention period:				
	Control slope (β_1)	–0.01 (–0.11– 0.03)		–0.01 (–0.04 – –0.01)	
	Difference PBI versus control in slopes (β_3)	–0.06 (–0.20 – –0.07)		0.00 (–0.03– 0.02)	
	Post-intervention period:				
5 Number of births at facility attended by a skilled birth attendant.	Difference PBI versus control in level change (β_6)	0.16 (–0.98– 1.30)	Recovering effect	0.28 (0.06– 0.49) *	Recovering effect
	Difference PBI versus control in slope change (β_7)	0.15 (–0.03– 0.32)		0.01 (–0.02– 0.04)	
	Pre-intervention period:				
	Control slope (β_1)	0.00 (–0.02– 0.03)		0.01 (–0.01– 0.03)	
6 Number of postpartum women receiving PNC from a skilled birth attendant within 14 days after delivery. ^b	Difference PBI versus control in slopes (β_3)	–0.03 (–0.08– 0.00)		–0.02 (–0.04– 0.00)	
	Post-intervention period:				
	Difference PBI versus control in level change (β_6)	0.29 (0.01– 0.59) *	Alignment	0.19 (–0.03– 0.42)	Recovering effect
	Difference PBI versus control in slope change (β_7)	0.02 (–0.02– 0.05)		0.05 (0.02– 0.08) **	
7 Number of one-year-old children attending pediatric services fully immunized.	Pre-intervention period:				
	Control slope (β_1)	0.00 (–0.01– 0.01)		0.00 (–0.01– 0.01)	
	Difference PBI versus control in slopes (β_3)	0.00 (–0.01– 0.01)		0.00 (–0.01– 0.02)	
	Post-intervention period:				
8 Number of children attending pediatric services having received Vitamin A in past 6 months. ^c	Difference PBI versus control in level change (β_6)	0.03 (–0.14– 0.21)	No substantial effect	–0.02 (–0.18– 0.13)	No substantial effect
	Difference PBI versus control in slope change (β_7)	0.00 (–0.02– 0.02)		0.00 (–0.02– 0.02)	
	Pre-intervention period:				
	Control slope (β_1)	0.01 (–0.01– 0.02)		0.07 (0.03– 0.11) **	
9 Number of children attending pediatric services having received Vitamin A in past 6 months. ^c	Difference PBI versus control in slopes (β_3)	–0.01 (–0.03– 0.02)		–0.02 (–0.08– 0.04)	
	Post-intervention period:				
	Difference PBI versus control in level change (β_6)	0.28 (–0.03– 0.59)	Trend diminished	–0.24 (–0.80– 0.32)	Trend improved
	Difference PBI versus control in slope change (β_7)	–0.03 (–0.07– 0.01)		0.08 (0.01– 0.15) *	
10 Number of one-year-old children attending pediatric services fully immunized.	Pre-intervention period:				
	Control slope (β_1)	0.03 (0.02– 0.05) **		–0.02 (–0.03– 0.00)	
	Difference PBI versus control in slopes (β_3)	–0.03 (–0.06– 0.00)		0.03 (0.01– 0.05) *	
	Post-intervention period:				
11 Number of children attending pediatric services having received Vitamin A in past 6 months. ^c	Difference PBI versus control in level change (β_6)	0.28 (0.01– 0.56) *	No substantial effect	–0.30 (–0.53 – –0.06) *	No substantial effect
	Difference PBI versus control in slope change (β_7)	0.02 (–0.01– 0.05)		–0.03 (–0.05– 0.00) *	
	Pre-intervention period:				
	Control slope (β_1)	0.15 (–0.03– 0.33)		0.02 (–0.12– 0.17)	
12 Number of children attending pediatric services having received Vitamin A in past 6 months. ^c	Difference PBI versus control in slopes (β_3)	–0.30 (–0.50 – –0.08) **		–0.18 (–0.45– 0.09)	
	Post-intervention period:				
	Difference PBI versus control in level change (β_6)	0.90 (0.06– 1.74) *	Recovering effect	0.03 (–0.91– 0.96)	Recovering effect
	Difference PBI versus control in slope change (β_7)	0.47 (0.14– 0.79) **		0.53 (0.17– 0.89) **	

(Continued)

Table 5. (Continued).

Quantity indicator ^a	Coefficients	Estimated SSDI-PBI effects, health centers effect size (95%-CI)	Effect pattern	Estimated SSDI-PBI effects, hospitals effect size (95%-CI)	Effect pattern
9 Number of HIV-positive pregnant women initiated on ART during pregnancy.	Pre-intervention period:				
	Control slope (β_1)	0.02 (-0.02– 0.06)		0.02 (-0.02– 0.05)	
	Difference PBI versus control in slopes (β_3)	-0.01 (-0.05– 0.04)		-0.05 (-0.09 – -0.01) **	
	Post-intervention period:		Level improved		Alignment
10 Number of couples tested for HIV during voluntary counseling and testing services.	Difference PBI versus control in level change (β_6)	0.49 (0.08– 0.91) *		0.42 (-0.02– 0.86)	
	Difference PBI versus control in slope change (β_7)	0.00 (-0.06– 0.05)		0.06 (0.02– 0.11) **	
	Pre-intervention period:		Trend improved		Trend improved
	Control slope (β_1)	0.02 (-0.04– 0.08)		0.05 (0.01– 0.09) *	
11 Number of clients attending family planning services counseled on contraceptive options. ^d	Difference PBI versus control in slopes (β_3)	0.01 (-0.06– 0.08)		-0.06 (-0.11– 0.00) *	
	Post-intervention period:		No substantial effect		Level diminished
	Difference PBI versus control in level change (β_6)	0.06 (-1.09– 1.20)		-0.99 (-1.83– 0.15) *	
	Difference PBI versus control in slope change (β_7)	0.23 (0.06– 0.40) **		0.22 (0.13– 0.30) **	
11 Number of clients attending family planning services counseled on contraceptive options. ^d	Pre-intervention period:				
	Control slope (β_1)	0.00 (-0.03– 0.02)		-0.04 (-0.10– 0.03)	
	Difference PBI versus control in slopes (β_3)	0.02 (-0.02– 0.06)		0.04 (0.05– 0.12)	
	Post-intervention period:		No substantial effect		Level diminished
Difference PBI versus control in level change (β_6)	-0.24 (-0.72– 0.23)		-0.52 (-1.31– 0.27)		
Difference PBI versus control in slope change (β_7)	0.00 (-0.05– 0.04)		-0.05 (-0.15– 0.06)		

* $p < .05$, ** $p < .01$ ^aProportional changes of averaged pre-intervention value of each indicator.^bOne intervention health center was omitted from this analysis due to having no non-zero baseline values.^cThree-monthly (instead of monthly) time intervals, given periodic outreach activities.^dTwo control facilities and two intervention facilities were omitted from the analysis of this indicator due to missing or poor data.

ANC = antenatal care; ART = antiretroviral therapy; IPTp = intermittent preventive treatment for malaria during pregnancy; PNC = postnatal care SSDI-PBI = Support for Service Delivery Integration Performance-Based Incentives.

protective PBF effect on the availability of selected essential obstetric equipment and supplies including delivery kits and uterotonic drugs.²⁹ Although the micro-context of the RBF4MNH program differed (e.g. that program had a limited focus on obstetric care service delivery only), both PBF programs seemed to have produced similar beneficial effects in making essential commodities available given the shared macro-context of central-level stock-outs.

Among quality indicators, we observed rather high saturation patterns for almost all indicators in at least one of the facility types. This unfortunately obscured the full appreciation of potential effect heterogeneities. The only exception to this seemed to be quality indicator 3 on the availability of client feedback systems. However, this heterogeneity in patterns appears not to be a result of the PBI, as the majority of control and PBI hospitals had such systems in place already prior to PBI implementation, while the proportion of control and PBI health centers with client feedback systems increased in a rather parallel way, thus unlikely to have occurred in response to the PBI intervention.

While performance indicators were identical for both facility types, we observed divergent effect

patterns between health centers and hospitals for a number of quantity indicators, the most divergent pattern was found for the number of post-partum women receiving PNC by a skilled provider (quantity indicator 6). For this indicator, the immediate effect on level differences was positive for SSDI-PBI health centers but negative for SSDI-PBI hospitals, and mainly resulting from the post-intervention trend developments observed in the respective controls. This pattern suggests a redirection of demand for PNC services from SSDI-PBI health centers to SSDI-PBI hospitals. PNC service quality in Malawi has been found to be higher in hospitals and private health facilities compared to government-owned health centers,³⁰ and this quality difference and resulting demand shifts between levels of care might have been intensified by the SSDI-PBI.

Especially for ANC-related quantity indicators (1, 2, 3, 4, 9), SSDI-PBI produced generally positive net effects of various magnitude and statistical significance among both intervention health centers and hospitals. For instance, the positive level difference in the number of pregnant women receiving iron supplements (indicator 3) might be partly linked to the fact that SSDI-PBI hospitals were able to maintain a higher availability of

iron supplements compared to their controls. SSDI-PBI health centers seemed to provide full stocks only later in time, which might explain the delayed upward trend (during the second payment cycle) contributing to the estimated overall level change. Similarly, the positive level differences observed for the number of pregnant women receiving at least two IPTp doses (indicator 4) and the number of HIV-positive pregnant women receiving ART (indicator 9) in SSDI-PBI health centers and hospitals might reflect the additional effect of PBF in the context of full stocks of sulphadoxine-pyrimethamine and HIV-testing kits in almost all intervention and control facilities of any type. In contrast, the SSDI-PBI did not seem to produce such an additional effect on the level differences for the number of couples tested for HIV (quantity indicator 10), despite the wide availability of testing kits.

While the SSDI-PBI has seemingly produced stronger positive effects on indicators related to ANC, we did not observe any substantial effects on the number of births attended by a skilled provider among health centers or hospitals (quantity indicator 5) and the number of one-year-olds fully immunized. For indicator 5, a possible explanation might be pregnant women's overall high use of skilled providers at birth in Malawi with 90%.¹¹ As our quantitative indicators represent count changes from pre-intervention averages, it might be likely that this indicator has been highly saturated with respect to the underlying target population of pregnant women; hence, the addition of SSDI-PBI could not create substantial further increases. The lack of substantial effects for indicator 7 might be explained by the fact that Malawi has generally high coverage for single vaccinations, but full immunization is less common due to frequent stock-outs of single vaccine types in facilities.³¹ This inconsistent availability of the full range of childhood vaccines might also be reflected to some degree by our findings on quality indicators 9–12.

Unlike indicator 7, which measures full immunization of one-year-olds attending facility child health services, indicator 8 on the number of vitamin A supplemented children also included community outreach activities, as reflected by the periodicity of observed values (hence the quarterly instead of monthly time points in our analysis for this indicator). The observed recovering effect patterns observed for both health centers and hospitals for this indicator might suggest a beneficial impact of the SSDI-PBI on preventive services that can be provided through outreach activities.

For the number of clients attending family planning counseling (indicator 11), the observed negative effects

remain less clear. While the overall use of contraceptive methods in Malawi is relatively low (46% among women of reproductive age), about half of modern contraception users (about 52%) tend to obtain their methods from government-owned health centers, with injectables and implants representing the predominant forms of long-term contraception.¹¹ One explanation could be that the exclusion of four facilities (two intervention health centers, two control health centers) for this indicator during our data cleaning process might have biased the observed effect pattern. Further research will be needed to understand the causes of this observed negative effect with respect to family planning service provision across intervention facilities.

Methodological Considerations

Our study has some limitations. First, among the methodological approaches applied in our study, we found that the most commonly used study design to examine PBF intervention effects (i.e. simple pre- and post-test design with controls in combination with DiD analysis⁷) produced almost no significant intervention effect estimates for the tested quality indicators. When compared to the effects on quantity indicators estimated by the ITS analysis, we hypothesize a lack of observable change or statistical significance on quality indicators may be due to the use of dichotomous indicators in a small sample, saturation of many of those indicators within each sub-sample, and probable month-to-month variations related to at least some of these indicators that are difficult to capture by a single indicator measurement. Further, using only two time-points to represent baseline and endline performance masks the nuanced trends that can be observed in an ITS analysis. Also, our selection of quality indicators was restricted to variables available in the SPA dataset, thus limiting our ability to identify additional indicators that might have been more appropriate in reflecting quality aspects incentivized by the SSDI-PBF.

Second, the ITS analysis, on the other hand, provided a picture of mostly positive and some statistically significant effects of the SSDI-PBI intervention with notable differences between health center and hospital effects. However, the nature of the underlying quantitative data did not allow us to determine the degree of saturation, as we had no information on respective reference target populations. Instead, we looked at overall counts with effect estimates measuring changes from pre-intervention averages. Hence, we could not make meaningful assessments with respect to actual changes in service coverage. For example, as outlined above, we anticipated greater effects on the number of

births attended by a skilled provider (quantity indicator 5). However, given the generally high use of skilled providers at birth in Malawi,¹¹ our findings might simply reflect maximum saturation for this indicator with respect to the respective target population.

Third, the nonrandom assignment of SSDI-PBI to selected facilities limited our design choice to a multi-method approach assessing causal interference based on two different quasi-experimental approaches. In the case of the before-and-after DiD analysis, with only one observation point available prior to intervention start, we could not test the parallel trend assumption underlying this method. While our choice of a matched control group based on facility-specific characteristics likely improved the accuracy of our DiD estimates, the differences in pre-intervention baseline averages in the quantitative data suggest that matching might not have worked ideally in all respects, so that we cannot fully exclude differences in pre-intervention trends on quality indicators.

Fourth, the study period of about 20 months in both the DiD and ITS analysis might have been too short to assess the full impact the SSDI-PBI might have eventually generated considering the general scope of PBF schemes to not only change reimbursement structures but to concomitantly introduce a set of new management and decision-making reforms. Fifth, the largely retrospective selection of outcome variables in both analytical components might have been less sensitive in capturing more detailed SSDI-PBI quality changes that could have been captured if indicator sets had been defined a priori.

Fifth, although the quality of DHIS-2-recorded indicators in Malawi is rather variable for some indicators, the routine indicator data we included in our analysis showed high levels of completeness and internal consistency. Still, in some instances, we excluded single time points or even single facilities due to poor data quality or extreme outliers, which might have biased some of our findings (e.g. indicators 6 and 11).

Lastly, we need to acknowledge the limited generalizability of our findings to other PBF programs or settings. As highlighted earlier, PBF programs operate with varied performance contracts and incentive structures, even within the same country as is the case for Malawi. Hence, we caution the reader when extrapolating our results in other PBF programs.

Conclusion

SSDI-PBI effects were more pronounced or statistically significant on quantity compared to quality indicators of

service provision. Given the underlying methodologies for each indicator set, it seems that multiple repeated or time-series data might be more favorable in exploring the effects of PBF programs compared to simple before-and-after designs. Although effect patterns for certain indicators were similar between facility types, the estimated effect magnitudes often still differed greatly between health centers and hospitals. Further, in the Malawi context performance incentives related to service quantity might be more effective if their focus is on relatively under-utilized services (e.g. ANC) compared to highly utilized services (e.g. skilled birth attendance). On the other hand, the currently applied incentives with respect to family planning services seem to produce unintended effects on contraception counseling of clients in both facility types. Regarding effect heterogeneity, incentives on PNC service provision seem currently to be more effective at hospitals compared to the health center level. If this should be an unintended effect, additional incentives increasing targeting PNC provision at health centers should be considered.

Disclosure of Potential Conflicts of Interest

No potential conflict of interest was reported by the authors.

Acknowledgments

The authors thank Takondwa Mwase for his support in reviewing and ensuring correct presentation of all sections pertaining to the SSDI-PBI program description.

Funding

This research project was made possible through Translating Research into Action, TRAction, which was funded by the United States Agency for International Development (USAID) under cooperative agreement number No. GHS-A-00-09-00015-00.

References

1. Eichler R, Levine R. Performance incentives for global health: potential and pitfalls. Washington (D. C): Center for Global Development; 2009.
2. Savedoff WD. Basic economics of results-based financing in health. Bath (Maine): Social Insight; 2010.
3. Soucat A, Dale E, Mathauer I, Kutzin J. Pay-for-performance debate: not seeing the forest for the trees. *Health Syst Reform.* 2017;3(2):74–79. doi:10.1080/23288604.2017.1302902.
4. Eichler R, Agarwal K, Askew I, Iriarte E, Morgan L, Watson J. Performance-based incentives to improve health status of mothers and newborns: what does

- the evidence show? *J Health Popul Nutr.* 2013;31:36–47.
5. Das A, Gopalan SS, Chandramohan D. Effect of pay for performance to improve quality of maternal and child care in low- and middle-income countries: a systematic review. *BMC Public Health.* 2016;16:321. doi:10.1186/s12889-016-2982-4.
 6. Fritsche GB, Soeters R, Meessen B. Performance-based financing toolkit. Washington (D.C.): World Bank; 2014.
 7. Witter S, Fretheim A, Kessy FL, Lindahl AK. Paying for performance to improve the delivery of health interventions in low- and middle-income countries. *Cochrane Database Syst Rev* 2012;2. Art. No.: CD007899. doi:10.1002/14651858.CD007899.pub2.
 8. World Health Organization. Malawi: WHO statistical profile. Malawi; 2015. <https://www.who.int/countries/mwi/en/>
 9. Kanyuka M, Ndawala J, Mleme T, Chisesa L, Makwemba M, Amouzou A, Borghi J, Daire J, Ferrabee R, Hazel E, et al. Malawi and millennium development goal 4: a countdown to 2015 country case study. *Lancet Glob Health.* 2016;4(3):e201–e214. doi:10.1016/S2214-109X(15)00294-6.
 10. *Tracking Progress towards Universal Coverage for Reproductive, Newborn and Child Health: The 2017 Report.* Washington (D.C.): United Nations Children’s Fund (UNICEF) and the World Health Organization (WHO); 2017.
 11. National Statistical Office Malawi. Malawi demographic and health survey 2015–16. Zomba (Malawi): National Statistical Office/Malawi and ICF; 2017.
 12. Ministry of Health Malawi. Malawi service provision assessment (MSPA) 2013–14. Lilongwe (Malawi, and Rockville, Maryland, USA): MoH and ICF International; 2014.
 13. Bradley S, McAuliffe E. Mid-level providers in emergency obstetric and newborn health care: factors affecting their performance and retention within the Malawian health system. *Hum Resour Health.* 2009;7(1):14. doi:10.1186/1478-4491-7-14.
 14. Wild L, Cammack D. The supply and distribution of essential medicines in Malawi. London (UK): Overseas Development Institute; 2013.
 15. Ministry of Health Malawi. Malawi health sector strategic plan 2011–2016. Moving towards equity and quality. Lilongwe (Malawi): Ministry of Health (Malawi); 2011.
 16. Mueller DH, Lungu D, Acharya A, Palmer N. Constraints to implementing the essential health package in Malawi. *PLoS One.* 2011;6(6):e20741. doi:10.1371/journal.pone.0020741.
 17. Global Health Expenditure Database. <http://apps.who.int/nha/database>.
 18. O’Neill T, Cammack D. Fragmented governance and local service delivery in Malawi. London (UK): Overseas Development Institute; 2014.
 19. McMahan SA, Muula AS, De Allegri M. “I wanted a skeleton ... they brought a prince”: A qualitative investigation of factors mediating the implementation of a performance based incentive program in Malawi. *SSM Popul Health.* 2018;5:64–72. doi:10.1016/j.ssmph.2018.04.006.
 20. Senlet P, Kachiza C, Katekaine J, Peters J. USAID/ Malawi support for service delivery – integration performance evaluation. Lilongwe (Malawi, and Vienna, VA, USA): IBTCI; 2014.
 21. Ministry of Health Malawi. Performance Based Incentives - Operational Manual. Lilongwe (Malawi): RBF4MNH, Options Office; 2014.
 22. Rawlins BJ, Kim Y-M, Rozario AM, Bazant E, Rashidi T, Bandazi SN, Kachale F, Sanghvi H, Noh JW. Reproductive health services in Malawi: an evaluation of a quality improvement intervention. *Midwifery.* 2013;29(1):53–59. doi:10.1016/j.midw.2011.10.005.
 23. Reeves BC, Wells GA, Waddington H. Quasi-experimental study designs series-paper 5: a checklist for classifying studies evaluating the effects on health interventions-a taxonomy without labels. *J Clin Epidemiol.* 2017;89:30–42. doi:10.1016/j.jclinepi.2017.02.016.
 24. Hesse-Biber SN, Johnson RB, eds. The Oxford handbook of multimethod and mixed methods research inquiry. New York (USA): Oxford University Press; 2015.
 25. Linden A. Conducting interrupted time-series analysis for single- and multiple-group comparisons. *Stata J.* 2015;15(2):480–500. doi:10.1177/1536867X1501500208.
 26. Cumby RE, Huizinga J. Testing the autocorrelation structure of disturbances in ordinary least squares and instrumental variables regressions. *Econometrica.* 1992;60(1):185–95. doi:10.2307/2951684.
 27. Chikhungu LC, Madise NJ. Seasonal variation of child under nutrition in Malawi: is seasonal food availability an important factor? Findings from a national level cross-sectional study. *BMC Public Health.* 2014;14:1146. doi:10.1186/1471-2458-14-1146.
 28. Mweninguwe R Health sector in coma. D+C Development and Cooperation. Published 2015 Dec 21 [Accessed 2019 Jul 9]. <https://www.dandc.eu/en/article/donors-bypass-malawi-government-support-health-sector>.
 29. Brenner S, Wilhelm DJ, Lohmann J, Kambala C, Chinkumba J, Muula AS, Allegri M. Implementation research to improve quality of maternal and newborn health care, Malawi. *Bull World Health Organ.* 2017;95(7):491–502. Article ID: BLT.16.178202. doi:10.2471/BLT.16.178202.
 30. Chimtembo LK, Maluwa A, Chimwaza A, Chirwa E, Pindani M. Assessment of quality of postnatal care services offered to mothers in Dedza district, Malawi. *OJN.* 2013;3(4):343–50. doi:10.4236/ojn.2013.34046.
 31. Tsega A, Hausi H, Chriwa G, Steinglass R, Smith D, Valle M. Vaccination coverage and timely vaccination with valid doses in Malawi. *Vaccine Rep.* 2016;6:8–12. doi:10.1016/j.vacrep.2016.06.001.