Master Thesis

A cost-effectiveness analysis of performance-based financing in Jimma, Ethiopia

Student name: Carmen Schakel
Student number: 624442
Master Thesis, Health Economics
School of Economics, Erasmus University Rotterdam
Supervisor: Owen O'Donnell
Date: 31 October 2023
Word count: 20,984 (excluding appendix and references)

Abstract

Background

Ministries of health, particularly in sub-Saharan Africa, are under pressure to achieve their country's health goals and work toward Universal Health Coverage within constrained budgets. Increasing competing health priorities in already challenged health systems, further exacerbated by COVID-19 setbacks, put ministries of health up for a daunting task. In several countries, including in Ethiopia, anticipations of a changing donor landscape require more dependence on domestic resources in the future. Within these constraints, it is therefore crucial to make evidence-based decisions on forms of health financing that are cost-effective in achieving these goals.

In Ethiopia, Performance-Based Financing (PBF) was first piloted in 2015 by Cordaid as an intervention to improve utilization and quality of health service delivery, enhance system governance and improve the health information system. After initial positive results, the program was expanded and the ministry of health in Ethiopia is currently also piloting its own national PBF design. Pursuant to this, the objective of this research is to assess the cost-effectiveness of PBF compared to status quo financing.

Methods

The cost-effectiveness of PBF is assessed in Jimma Zone, Ethiopia, for the period April 2019 to August 2022 at health center level using a health system and financial perspective. Costs considered were PBF program costs incurred by Cordaid. These costs were assumed as additional to status-quo financing as no cost of the alternative scenario was assessed. Annual service utilization data for September 2018 to August 2022 for Jimma Zone and a neighboring control Zone was retrospectively retrieved during a recently conducted end-term evaluation. Using health statistics, service utilization was converted to coverage rates. Considering the time trend of the control group, service coverages were converted to lives saved modelled in the Lives Saved Tool. Lastly, lives saved were converted to quality-adjusted life years (QALYs).

Results

Under PBF 3,211 deaths occurred while this was 3,262 in the control group translating to 51 lives saved. Raw PBF program costs were €0,96 per capita per year and after adjustments to the study population and services modelled, this translated to €0.2561 (EUR 2019) per capita per year. QALYs gained were 1,108 per million population resulting in an incremental cost-effectiveness ratio of €233 per QALY gained (EUR 2019).

Conclusion

Based on a cost-effectiveness threshold of 1.5 times Ethiopia’s 2019 GDP of €1,126, PBF in Jimma, Ethiopia appears cost-effective. When compared to similar cost-effectiveness studies in other sub-Saharan countries, PBF also appeared cost-effective with lower effects but at significantly lower costs.

Table of Contents

1. Introduction

Government health budgets in low and middle-income countries (LMICs), especially in Sub-Saharan Africa (SSA), are under pressure (WHO, 2022), while donor reliance in the health sector is expected to decrease, including in Ethiopia (MoH – Ethiopia, 2021). This pressure is exacerbated by COVID-19, leaving ministries of health challenged with trying to find ways to reach their health goals and move towards Universal Health Coverage, with constrained health budgets and increasingly competing health priorities in already challenged health systems. It is therefore becoming ever more important that evidence-based decisions are made towards effective interventions in improving health outcomes at the lowest possible cost.

One way in which governments and non-governmental organizations (NGOs) have tried to improve health outcomes towards these goals in resource-limited settings, is through Performance-Based Financing (PBF). In contrast with centrally allocated ex-ante input-based funding, under PBF health providers are paid ex-poste at the margin based on their performance (Musgrove, 2011). Performance is measured through a set of pre-defined indicators which measure the quantity and quality of care provided by healthcare providers.

In SSA, PBF first started in Rwanda in 2001 (Fritsche et al., 2014). Over the years, PBF gained traction and as of 2017 it had been implemented in 32 of 46 SSA countries, whereby the World Bank alone has spent approximately US$2.5 billion on PBF interventions since the late 2000s (De Walque et al., 2022). While PBF has been shown to be effective in certain countries (Basinga et al., 2011; Fichera et al., 2021), mixed results are found elsewhere (Diaconu et al., 2021). Moreover, PBF is often considered to be costly due to their extensive monitoring framework and, as a result, the (cost-)effectiveness of such programs is much debated (Binyaruka et al., 2020). While part of this debate questions the most appropriate form of health financing (De Walque et al., 2022; Witter et al., 2021), a major source of the debate can be linked to limited scientific evidence of the cost-effectiveness of such programs (Turcotte-Trembla et al., 2016), although evidence has been increasing over recent years (Shepard et al., 2020; Zeng et al., 2018; Zeng et al, 2022). Hence, the need for more scientific evidence of the cost-effectiveness of PBF. This will be explored for a PBF program implemented by Cordaid in Jimma Zone, Ethiopia.

Cordaid, a Dutch NGO, is a pioneer in PBF and has been involved in PBF programs in over 15 countries since the early 2000s. In 2015, Cordaid started a small pilot in the Borana Zone in the Oromia Region of Ethiopia to improve health service delivery by increasing equitable service coverage and better quality of care, as well as improving the health information system. The program expanded multiple times and is currently implemented in Borana Zone, Jimma Zone and North Achefer Woreda. With seemingly positive results, the Ministry of Health in Ethiopia has gained interest and included PBF as a potential approach under its health financing agenda in the Health Sector Transformation Plan II (MoH – Ethiopia, 2021). They are currently piloting their own national PBF design with technical assistance from Cordaid. Therefore, an economic evaluation of PBF has both practical relevance, for implementers, and policy relevance, for decision-makers.

1.1 Study Objective

This research aims to conduct a cost-effectiveness analysis (CEA) of the PBF program implemented by Cordaid in the Jimma Zone in Ethiopia, compared to status quo financing, to measure the long-term health consequences and related costs. The main research question is: What is the cost-effectiveness of PBF implementation within Jimma Zone, Ethiopia compared to status-quo financing?

Sub-questions are:

1. What are the incremental costs attributed to PBF?
2. What are the health effects attributed to PBF?
3. What are the health effects attributed to status quo financing?
4. What is the incremental cost-effectiveness of PBF?

These questions will be examined specifically at health center level. All health centers are examined together and also separately by program phase (Phase I vs Phase II)

2. Theoretical Background

2.1 Performance-Based Financing: What is it?

Within the health sector, Results-Based Financing (RBF) is an overarching term defined as "a cash payment or non-monetary transfer made to a national or sub-national government, manager, provider, payer or consumer of health services after predefined results have been attained and verified. Payment is conditional on measurable actions being undertaken" (Musgrove, 2011). PBF is a subset RBF, focused on the supply-side of a health system, more narrowly defined by three principles: (1) incentives are rewarded to individual health facilities with partial disbursement to health workers, (2) incentives are financial and (3) incentives are based on outputs (e.g., fee-for-service but do not reflect unit costs) and conditioned on a certain level of service quality. In practice, however, the terms PBF and RBF are often used interchangeably. To make things more complicated, pay-for-performance is considered synonymous with the umbrella term, RBF. For this study, however, we will be referring to the term PBF and its meaning as described here.

While the PBF definition as defined above is focussed on one dimension of the approach, specifically provider payments, PBF interventions usually adopt multiple performance frameworks for different health system levels, including for community-based organizations (CBOs), district/provincial health offices and ministries of health (Fritsche et al., 2014). Thus, multiple changes to the health system are introduced simultaneously, whereby PBF is sometimes referred to as a 'health system reform' (Musgrove, 2011). Payment to providers, but also to stakeholder in other levels of the health system, are intended to complement existing government input financing. It is not intended to completely replace existing input financing. Therefore, PBF is a (partial) shift from input-based to output-based financing (Fritsche et al., 2014).

From an economic perspective, the goal of PBF is to (1) enhance allocative efficiency by incentivizing high priority and cost-effective services (2) to improve technical efficiency within facilities through leveraging existing productive assets (3) improving effectiveness (4) improving coverage rates and equitable outcomes and (5) improving transparency and accountability among stakeholder within the system (Frische et al., 2014; Witter et al., 2013). More, broadly there is also an expectation that PBF can facilitate strategic purchasing (Soucat et al., 2017; Witter et al., 2019). Therefore, it is important to analyse and understand the cost-effectiveness of such programs in achieving these goals.

Primary evidence of PBF in Rwanda found that the pure 'incentive effect,' rather than increases in financial resources, had a large and significant effect on child health indicators (Gertler and Vermeersch, 2012). These findings were associated with improvements in utilization and quality of prenatal and postnatal care by reducing the know-do gap by 20%, improving provider technical efficiency. Authors also found the magnitude of the effects to be larger for services where financial incentives and marginal return to the effort were higher. Bonfrer et al. (2014) found that PBF improved use and quality of healthcare in Burundi, finding a 22, 5 and 17 percentage point increase in the utilization of institutional deliveries, family planning and higher technical quality of care, respectively. However, larger improvements were seen across groups with higher incomes suggesting that the approach may not have been equitable. Nkangu et al., (2023) found that PBF may facilitate increasing maternal service utilization through a reduction in out-of-pocket expenditure due to lowering operational barriers, such as hidden facility costs, in settings where services should be subsidized or free. However, this was not found in settings were PBF was a stand-alone supply-side approach, as other demand-side barriers hinder service utilization. Despite these positive results, a recent systematic study revealed PBF was institutionalized within a country's health system in only a few cases, and that evidence of PBF, particularly for its cost-effectiveness and large system-wide effects, remain limited and inconclusive (James et al., 2020).

2.2 Performance-Based Financing: Evidence of Cost-Effectiveness

While gaps remain, in recent years several studies have attempted to measure the cost-effectiveness of PBF across various countries (Zeng et al., 2018; Salehi et al., 2020; Shepard et al, 2020; Zeng et al., 2022). Three recently published CEAs of PBF programs in Zambia, Zimbabwe and Nigeria followed the methodology described in a World Bank toolkit (Shepard et al., 2015), which was also be applied in this study. In these three countries, PBF programs were compared to (1) status quo financing with a constant budget and (2) *additional* input-financing at an average level to that of PBF payments. All three studies found PBF to be cost-effective. In Zambia, authors noted that PBF was cost-effective in improving Maternal and Child Health (MCH) services, in comparison to both status quo financing and additional input-based financing (Zeng et al., 2018). In Zimbabwe, PBF was found to be cost-effective for MCH services at $636 per quality-adjusted life year (QALY) gained for a pilot program, and $479/QALY gained for an ongoing program when compared to additional input-based financing. In Nigeria, the cost-effectiveness of PBF was compared to direct financing facility (DFF) and the status quo (Zeng et al., 2022). While both forms of financing improved coverage, PBF was found to be more effective by saving more lives and DFF was found to be more efficient being half the cost of PBF. Both are cost-effective compared to the status quo.

A fourth, yet to be published, CEA of PBF in Borana Zone, Ethiopia similarly followed the same methodology laid out in the toolkit. However, this study only measured the impact of improved service utilization attributable to PBF on effectiveness, and the impact of improved quality of care was excluded from the analysis (M.A., Koricho, personal communication, May 26, 2023). According to the study abstract presented at the International Health Economics Association Congress in July, 2023, the authors were able to convert district-level health service volume to service coverage, use the Lives Saved Tool (LiST) to convert the change in service coverage to lives saved, and finally convert lives saved to QALYs (Koricho, M.A., 2023). PBF in Borana saved 261 lives compared to 194 lives saved with status quo financing, translating to 6118 and 4526 QALYs gained per one million people, respectively. With an annual PBF cost per capita of $1.8 USD, this resulted in an ICER of $1,441 USD, which was found to be cost effective.

In the Republic of Congo, a CEA was conducted using a different methodology which measured the effects of PBF by drawing from empirical data on intervention effectiveness, measured as disability-adjusted life years (DALYs) per unit of service delivery (World Bank, 2020). DALYs averted were measured through an increase in service delivery attributable to the intervention. When considering both direct and indirect costs, the incremental cost-effectiveness ratio results in $52.40 per DALY for the entirety of the PBF program. This was considered highly cost-effective based on findings from other studies. However, the study had several serious limitations including the lack of a robust control group, quality of care in the effectiveness analysis, (i.e., likely underestimating the impact) and empirical health service effectiveness data were used from other developing countries.

On the other hand, in Afghanistan, PBF was not found cost-effective from a provider-payer perspective compared with status quo financing, even though the quality of MCH services improved (Salehi et al, 2020). The cost of US$1,242 per DALY was above the threshold of US$349 GDP per capita. The lack of cost-effectiveness was due to systemic issues such as insufficient health workers, lack of essential drugs and supplies and referral system, in part due to the lack of health facility autonomy in the program. A high incentive for skilled birth attendance (US$37 per service) and an imbalance with low incentives for other services also appeared to reduce cost-effectiveness. Finally, transaction costs for administering periodic data verification were considered high. The latter has also been found in other PBF programs and has been criticised as making the financing mechanism costly (Borghi et al., 2015).

3. Ethiopia and Performance-Based Financing

This section first describes the general Ethiopian healthcare context. Next, a closer look is given to the context in Jimma before moving into what PBF entails in Jimma Zone.

3.1 Ethiopian context

Ethiopia is a low-income country in the Horn of Africa with an estimated population of 123 million people in 2022 (World Bank, 2023). It is the second most populous country on the African continent with 80% of people living in rural areas (African Development Bank Group, 2021). The country has a federal governance structure decentralized along ten regional states and two chartered cities. Regions are further divided into three levels: zones, woredas (districts) and kebeles (villages). Oromia Region, where Jimma Zone is located, is the largest and most populous region with 40 million people (Ethiopian Statistical Service, 2022). Ethiopia's per capita Gross Domestic Product (GDP) of USD$1,027 in (2022, current prices) is one of the lowest in the world, despite strong economic growth from 2015 to 2019 at an average 10% annually, one of the highest growth rates worldwide (World Bank, 2023). Economic growth slowed between 2019 to 2022 because of COVID-19. Lastly, poverty declined from 30% to 24% from 2011 to 2016 but COVID-19 may have hindered achieving additional gains (Africa Development Bank Group, 2021).

3.2 Ethiopian healthcare context

Ethiopia experienced improvements in its population health status over the last decades. Life expectancy at birth increased from 58 years to 65.5 years from 2007 to 2017, respectively (MoH-Ethiopia, 2021). Maternal mortality declined from 871 to 401 per 100,000 live births between 2000 and 2017. Neo-natal mortality decreased from 39 to 33 per live birth from 2000 to 2019. Moreover, infant mortality and under-five mortality declined to 47 and 59 (2019) per 1,000 live births. At the same time, coverage for maternal, neonatal, and child health services improved. Finally, the country's UHC service coverage index stood at 43% (2019) (Ministry of Health, Ethiopia 2021).

While health outcomes have improved, likewise the financial resources for healthcare have increased. According to 2019/2020 National Health Accounts, Ethiopia's total health expenditure (THE) increased from USD$3.1 billion to 127 billion between in 2016/17 and 2019/20 resulting in an increase in the share of THE to GDP from 4.7% to 6.3% (Ministry of Health – Ethiopia, 2022). Similarly, per capita health expenditure increased from USD$33.2 to USD$36.3 during this same period. However, this is below the WHO, in 2015, recommended per capita health expenditure of 86$USD. Of THE, the largest share went to health center and health post levels (30%), government health administration (20%), primary hospital level (14.5%) and the remainder to other service providers in 2019/20. During this period, THE was financed by the government (32%), donors (34%), out-of-pocket (30.5%), private employers (2.5%), and although coverage of community-based health insurance (CBHI) was being scaled throughout Ethiopia, it remained limited to a contribution of 1% (MoH – Ethiopia, 2022).

In Ethiopia, health services are delivered by both public and private providers with most users (80%) accessing healthcare through public facilities (Woldie et al., 2021). The public healthcare delivery in Ethiopia, as shown in Figure 1, is organized in a three-tiered system: primary, secondary, and tertiary level (MoH – Ethiopia, 2021). The primary level consists of health posts, health centres (HC) and primary hospitals which together form the Primary Health Care Unit. Health posts are situated in kebele's which are staffed with two are three health extension workers. Extension workers provide preventative care and health promotion services to the local community. On average, five satellite health posts are linked to one health center who provides health extension workers with technical support and serves as a referral center for them. A health center serves roughly 25,000 people and delivers basic preventative and curative care. Health centers can refer to primary hospitals which provide inpatient, ambulatory and emergency surgical services. At secondary level, there are the general hospitals which are referral centers for primary hospitals, and at tertiary level are specialized hospitals which are referral centers for general hospitals.

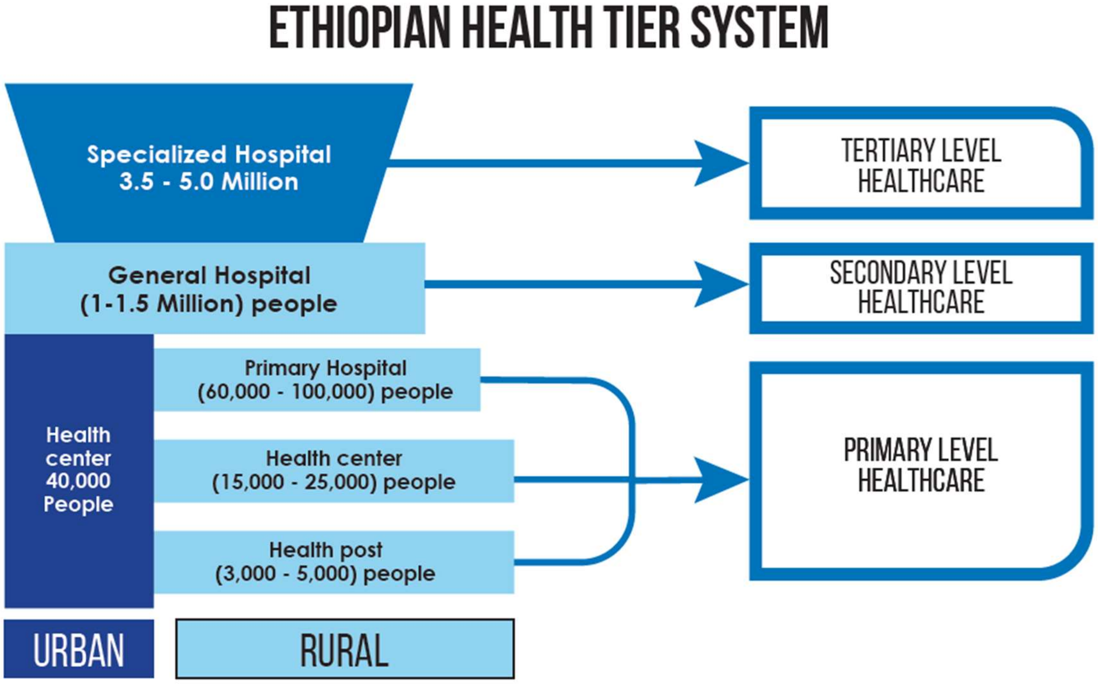


Figure 1. Ethiopian Health System Structure (Source: MoH - Ethiopia, 2021)

Current health financing practices in Ethiopia are informed by the Health Financing Strategy endorsed in 1998 by the Ministry of Health, and more recently the country's 2022-2023 health care financing strategy (Alebachew, 2015; EHIS, 2022). Key components of the 1998 reform include revenue retention and utilization, setting of user fees, fee waiver system for the poor, provision of exemption services, establishment of insurance, introduction of private wing in public hospitals, outsourcing of non-clinical services and establishing facility-level governing bodies (USAID, 2012). Currently, public facilities are allowed to retain unused user fees from non-exempted services for quality improvements. At the public primary level, this comprises revenue from direct user fees and fee-for-service claims from the rolled-out CBHI. User fees for primary government facilities are set by the Regional Health Bureaus and are subsidized through government budget allocation (EHIS, 2022). However, most public facility financing remains through stringent input-based line-item government budgeting to cover operational and pharmaceutical costs. These line-item budgets are based on historical expenditures and any unused funds are returned to the treasury (EHIS, 2022). The fee waiver system was introduced to make services free for the poor at the discretion of local authorities where local governments are expected to reimburse health providers for lost user fees (Alebachew, 2015). However, in practice targeting the poor is challenging and reimbursement is often delayed or non-existent. Exempted services are free of charge for all users. The revised 2019 Essential Health Service Package proposed 570 services to be exempt mainly related to maternal, family planning, immunization, and tuberculosis services (EHIS, 2022). However, there is no explicit funding to cover exempted services and government line-item budgets are insufficient to finance this whereby facilities resort to using their internally generated revenue to buy medicines for free services. Finally, CBHI has been rolled out for the informal sector which is managed at the woreda level. In the scheme, (voluntary) enrolees pay an annual premium whereafter service within the benefit package are free of charge. Deficits are common whereby government facilities often experience delays in receiving claim payments or only receive partial payments (EHIS, 2022). In general, revenue sources to public health facility are characterized by inflexibility and are insufficient to deliver quality services.

3.3 Performance-Based Financing in Jimma, Ethiopia

Jimma Zone lies in the Oromia Region of Ethiopia and is geographically divided into 21 woredas as shown in Figure 2 (Cordaid, 2019). Each woreda consists of a Woreda Health Office (WorHO) which supervises and supports the health centers within its respective geographical boundaries. The Zonal Health Department (ZHD) manages and oversees the WorHO and hospitals within Jimma Zone. In Ethiopian Fiscal Year (EFY) 2013 (July 2021 to June 2022), based on data from the Jimma ZHD, the zone had an estimated population of 3,538,266 and is one of the most densely populated zones in Oromia Region (Cordaid, 2019). The zone counts 521 health posts, 121 health centers and 7 hospitals in EFY 2013.

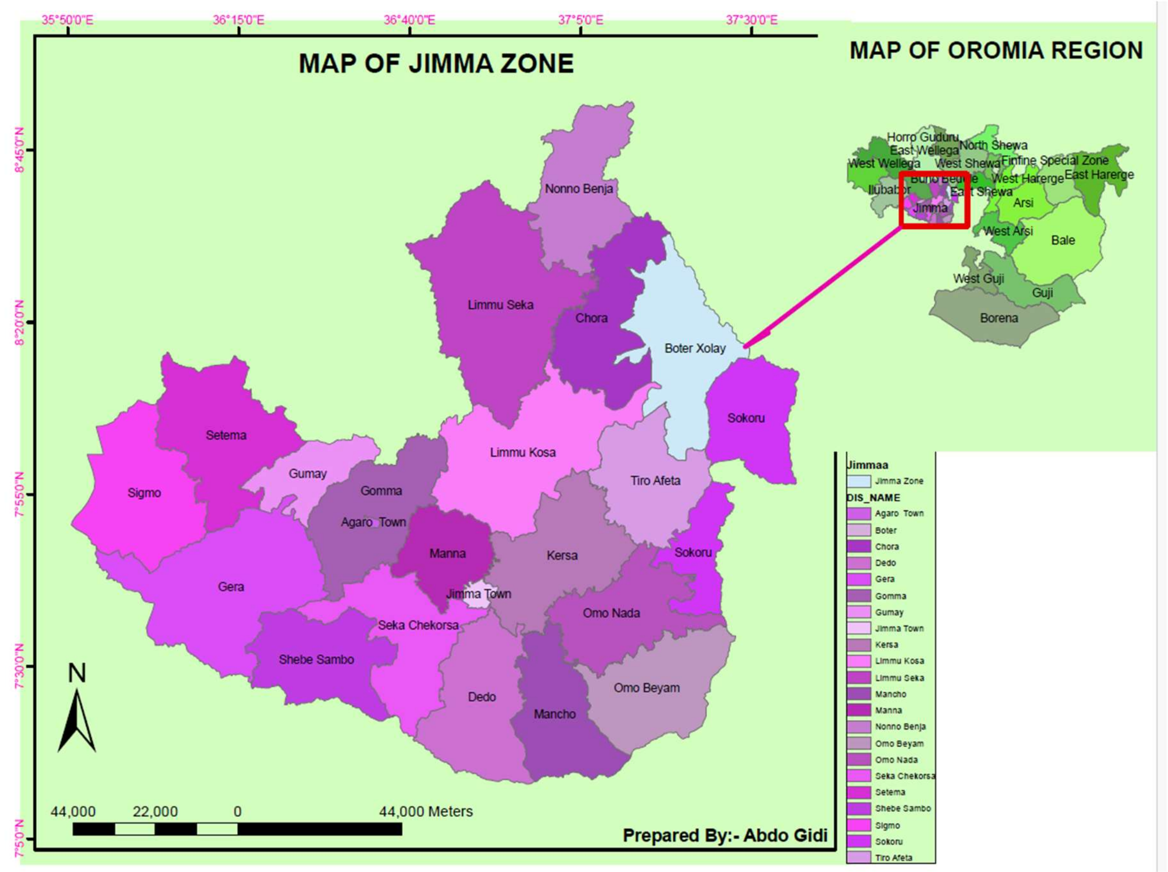


Figure 2. Map of Jimma Zone, Ethiopia (Source: Jimma Zone Health Department)

Implementation started in Jimma Zone as of April 2019 with financing from the Netherlands Embassy. After an initial six-month inception phase to design and set up program activities, PBF started implementation in October 2019 in 13 woredas which covered 64 health centers and 4 hospitals. Woredas were selected based on whether USAID was implementing its 'Transform' program and woredas were chosen where the program was not active. In January 2021, PBF was expanded to all 21 Woredas in the Jimma Zone covering all government-funded health posts, health centers and hospitals (Cordaid, 2020). Facilities starting PBF in October 2019 are known as Phase I facilities and those starting in January 2021 are Phase II facilities. Note, that health posts for the entire Zone, including for Phase I health centers, only started implementation towards the end of 2021.

The main objectives of the PBF program in Jimma Zone were to (Cordaid, 2019):

(1) improve *health service delivery* through increased utilization of good quality services and increased equity in access,

(2) improve *governance* of the service delivery through increased capacity at the WorHOs and the ZHD and institutionalisation of PBF in the Ethiopian health system, and

(3) enhance the *health information system* through increased data-based decision making at WorHOs, ZHD and Regional Health Bureau.

To achieve these objectives, Cordaid builds its PBF programs upon on six guiding principles which are expected to dynamically change the health system. These principles include: (1) *separation of functions*, (2) business planning and *contracting*, (3) linking *payments to results* (4) healthcare provider *autonomy*, (5) *equitable access* through remoteness bonus (6) and *community engagement* through patient satisfaction surveys (Cordaid, 2020). The assumption is that when moving from input-based to output-based financing through safeguarding these principles—paying health facilities and their staff for performance and allowing them to autonomously purchase inputs—will motivate them in their work. Additionally, it is assumed that it will lead to more respectful and caring staff, better infrastructure, and equipment to provide quality services, more efficient and entrepreneurial management of facilities, provision of services more in line with community needs, and improving health management information system (HMIS) data for informed decision making. In turn, this should lead to the three objectives. To operationalize these objectives and safeguard the guiding principles, thorough institutional arrangements and performance frameworks were in place.

3.3.1 Institutional Arrangement

As shown in Figure 3, there are multiple stakeholders with distinct roles in the PBF design (Cordaid, 2019). To start, the Federal Ministry of Health (FMoH), Oromia Regional Health Bureau, the Jimma ZHD and WorHOs form the health system regulatory bodies where the federal and regional levels are responsible for setting standards and playing a supervisory role, while the ZHD and WorHOs were contracted for their supportive supervision role towards health facilities. Hospitals and health centers are contracted for improving quantity and quality of services. CBOs are also contracted to conduct quarterly patients' satisfaction surveys and report feedback back to health providers. All contracts are signed with the performance purchasing agency (Cordaid field office in Jimma) on a bi-annual basis which stipulate what each entity is paid for and for how much. Regulator and health facility performance contracts have underlying business plans which are developed by them specifying priority areas and where earned PBF funds will be invested in. An external agency is contracted every few years to conduct a counter verification (Cordaid, 2019).

Rigorous verifications are in place to measure facility output: (1) monthly quantity of services and (2) quarterly quality performance (Cordaid, 2019). Health facilities submit their service delivery data monthly to the WorHOs or ZHD via Ethiopia's regular HMIS/DHIS2 system. This 'declared' data is the primary data source for quantitative PBF verification by the Cordaid field office. Quantity verifications are conducted based on rigorous verification guidelines for each indicator according to national standards and registration books (Cordaid, 2019). Any missing criteria for a specific case, for example client personal information, results in that case not being considered 'verified' nor paid. When, for a certain indicator, there is a difference between *declared* and *verified* data which exceeds the agreed upon 10% error margin, that indicator will not be paid for in that month. This is to incentivize data accuracy. The quarterly quality verification consists of two parts: the technical quality verification and the patient's satisfaction verification. Technical quality verification is conducted for the health centers and hospitals by the WorHOs and ZHD, respectively, along with a Cordaid field office representative. The patient satisfaction surveys are conducted by locally contracted CBOs which trace randomly selected patients, who received a health service in the past quarter, and measure their satisfaction. The 'verified' results of these three verifications make up the quarterly payment for a health facility. Finally, the performance of the WorHOs and ZHD in their supportive supervision role is also verified and paid quarterly. CBOs are paid based on the number of patients traced and production of a report with findings (Cordaid, 2019).

All data collected by the Cordaid field office is captured in its PBF DHIS2 system specific to the program. Quarterly invoices were validated and paid by the Cordaid country office (fundholder). Payments are made directly to the bank accounts of all contracted entities, and service providers and regulators can invest their subsidies according to their previously developed business plans. They were allowed to spend a maximum of 30 percent for staff incentives and a minimum of 70 percent for investments in the facility. Staff incentives were dispersed based on an indexes tool which measures the individual performance of staff. This cycle of reporting, verification and payments continued throughout the project and was also accompanied by regular training and coaching of facilities on PBF (Cordaid, 2019).

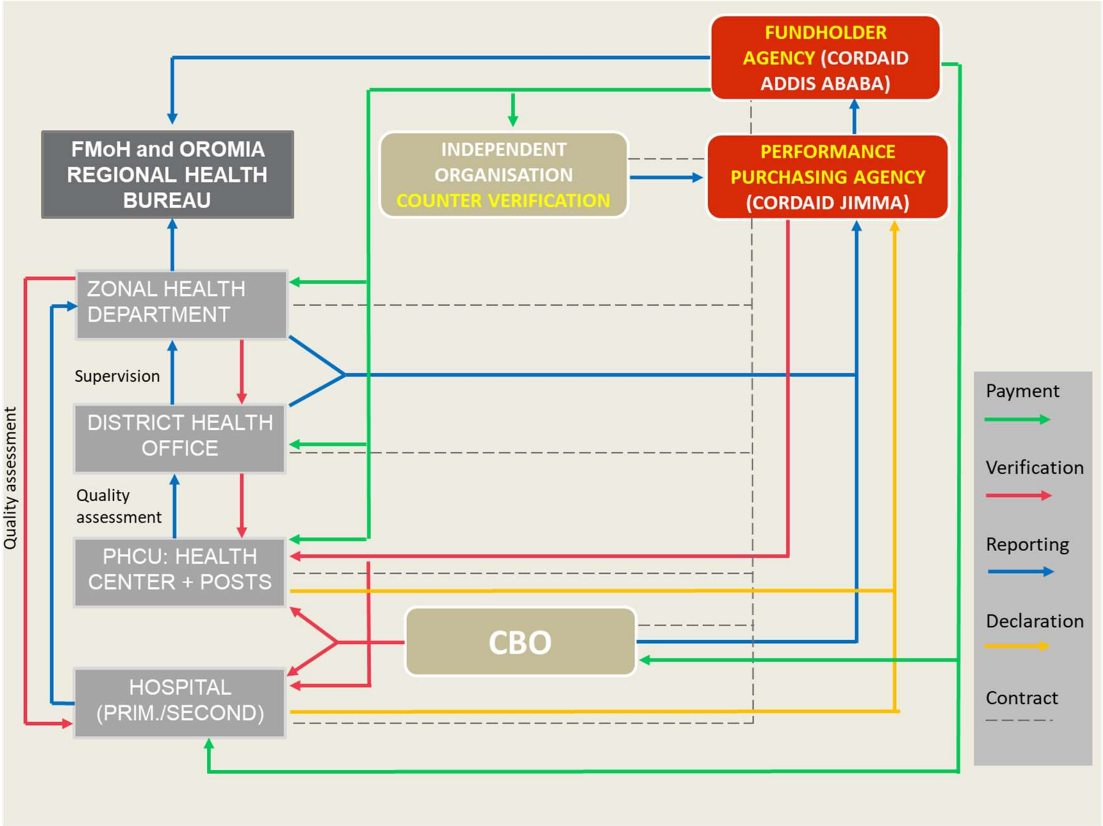


Figure 3. Institutional setup of PBF in Jimma, Ethiopia (Source: Cordaid, 2019)

3.3.2 Health Center Performance Frameworks

Performance frameworks exist for each contracted entity. However, given that the research question pertains specifically to health centers, the health center performance framework will only be described here. The performance framework determines the quarterly PBF payment to health centers and is made up of three main components: quantity, equity, and quality payments (both technical and perceived quality). The PBF payment formula for a health center is as follows (Cordaid, 2019):

$$PBF\ Payment = \{\sum_{i=1}^{n}(P_iQ_i)\} * (1+E) * (1+(0.5*(0.8*T+0.2*C)))$$

The first part in the formula, $\sum_{i=1}^{n}(P_iQ_i)$, represents the quantity payment where P_i is the fixed unit price for indicator *i* and Q_i is the quantity 'verified' for indicator *i*. Hence, the sum of quantity of services verified multiplied by the price for each indicator forms the quantity payment. All 24 incentivized quantity indicators and their prices can be found in Appendix 1. Most incentivized indicators are related to MCH, as has often been the case in PBF programs.

Note, that as of January 2021 there were several changes in the performance framework. First, a 24th indicator was added, post-abortion care, while there were only 23 indicators before this period. Second, all indicator prices were increased by 1.5 times as of January 2021 (Cordaid 2019, Cordaid 2020).

The second part of the formula, $(1 + E)$, considers an additional equity bonus, where E represents the equity score based on a set of five criteria (catchment size, nearest hospital, road condition, availability of public transport and availability of communication services). The higher the remoteness and difficult circumstances a facility operates in, the higher the equity bonus. The equity bonus is on top of the quantity payment and the equity score, E, was either 0%, 10% or 20% for a health center before 2021. As of 2021, the equity scores could be 0%, 10%, 20%, 30% or 40% (Cordaid 2019, Cordaid 2020).

The third part of the formula, $\left(1 + \left(0.5 * (0.8 * T + 0.2 * C)\right)\right)$, accounts for the quality payment component, where T is a score from 0 to 100% for the quarterly technical quality and C is a score from 0 to 100% for the patients perceived quality (i.e., satisfaction surveys). The addition of these two scores forms the total quality score. Hence, the technical quality weights 80 percent of the total quality score and patient satisfaction weights 20 percent. The total quality score is then multiplied by 50 percent to the quantity and equity payment components. The multiplication of 50 percent represents the trade-off between incentivizing the quantity versus quality of services. In this case, the quantity payment represents that largest part of the PBF payment meaning that increasing the demand or quantity of services is incentivized more compared to quality of care. The 16 categories and respective weights for which technical quality was measured can be found in Appendix 2. Patient satisfaction was measured based on six criteria: (1) staff attitude, (2) waiting time, (3) availability of medicines and equipment, (4) adequate infrastructure, (5) ambulance services and (6) cleanliness of facility (Cordaid 2019, Cordaid 2020). Finally, not shown in the formula is that sometimes, due to sanctions, a certain percentage of the payment is withheld. This only occurs on occasion.

4. Methodology and Data

This CEA aimed to align with methodology laid out in a paper, developed based on experiences in Zambia, for assessing cost-effectiveness of PBF programs and to facilitate cross-country comparison (Shepard et al., 2015). Where alignment was not feasible, due to lack of data or for other reasons, other ways of performing the analysis were considered and explained. This methodology was chosen for several reasons:

1. It was the only comprehensive guideline available on assessing the cost-effectiveness of PBF programs.
2. Most recent CEA studies of PBF found in the literature review were conducted according to these guidelines allowing for cross-country comparison (Zeng et al., 2018; Salehi et al., 2020; Shepard et al, 2020; Zeng et al., 2022; Koricho, M.A., 2023).
3. Methodologies used in other CEA studies of PBF, such as in the Republic of Congo (World Bank, 2020) and Afghanistan (Salehi et al, 2020), were not appropriate in the context of rural Ethiopia and were not feasible due to data unavailability.

According to the guideline, the measurement of cost-effectiveness is largely linked to results from an impact evaluation conducted on a PBF program (Shepard et al., 2015). An impact evaluation of PBF generally contains information on the changes in the quantity and quality of health services attributable to the program, for which guidelines have been written elsewhere (Vermeersch et al., 2012). Hence, the intention of this cost-effectiveness study was to build on information collected during the end-term impact evaluation of the PBF program in Ethiopia, which was recently conducted by evaluators external to Cordaid (Icos Consulting Plc., 2023).

4.1 Study Scope

The study employed a quasi-experimental design to assess the costs and effects of a sample of health centers in Jimma Zone, Ethiopia before and after PBF implementation compared to a control group, neighbouring Bedele Zone. The implementation period within the study ran from April 2019 to August 2022 or 3.42 years. This included an initial 6-month inception phase and costs were assessed for this entire period. Phase I health centers (64) started implementation in October 2019 (2.92 years of implementation) and Phase II health centers (57) in January 2021 (1.67 years of implementation). Hence, effects were studied during these periods. Effectiveness data on service utilization were captured for the period September 2018 to August 2022, which includes pre-intervention data. Differences were estimated by comparing changes to the neighbouring, Bedele Zone.

4.1.1 Geographical Area

While the end-term evaluation was conducted in all zones where Cordaid has implemented PBF, this CEA solely zoomed in on PBF implementation in the Jimma Zone and non-implementing Bedele Zone (Icos Consulting Plc., 2023). Other geographical locations where PBF was implemented and evaluated were excluded from the analysis because there were no robust control groups. Bedele Zone was chosen as the control group for Jimma Zone because it was also used in the mid-term evaluation in December 2020, at which time it was concluded that out of all zones adjacent to Jimma Zone, Bedele Zone appeared to have the most socio-demographical similarities (K. Woldemariam, personal communication, June 6, 2023).

4.1.2 Facility Type

While PBF is implemented at hospital, health center and health post levels of the health system, this study only measured the cost-effectiveness of PBF at health centers. Health posts were excluded because the necessary data were not collected during the end-term evaluation. Second, hospitals were excluded because of their small sample size; specifically, four intervention and one control hospitals were sampled. Given that hospitals had a distinct performance framework from health centers, it is counterintuitive to combine their data with that of health centers as one outcome in the analysis. Therefore, hospitals are excluded.

4.1.3 Sampling technique and sample size

Facility sampling was conducted during the end-term evaluation because the intervention and control groups were not randomly assigned prior to program implementation (Icos Consulting Plc., 2023). The evaluators used a multi-stage sampling technique. First, woredas were stratified by program phase. Next, 20% of the woredas entering in each phase were sampled using a systematic sampling technique, after which health facilities were sampled based on the lottery method (Icos Consulting Plc, 2023). The end-term evaluation does not describe in further detail the type of systematic sampling technique applied.

Table 1 shows the final number of health centers sampled. There were 14 health centers sampled in the intervention group and 10 health centers sampled in the control group. The total catchment population of the sampled facilities was 279,146 and 197,338 for Phase I and Phase II facilities, respectively, during the four-year study period. The catchment population represents an estimation of the population within a geographically outlined area that a specific health center serves. Catchment population data was retrieved from the sampled facilities during the end-term evaluation only for the last quarter of the study, Q3 2022. Based on this catchment data, populations were estimated for the control group during the entire study period. Similarly, this was done for PBF facilities, although Cordaid had additional catchment data available for these facilities. Refer to Annex 3 for a full list of sampled facilities and explanation on population estimates.

Table 1. Sampled health centers of intervention and control groups

	Intervention Group (Jimma Zone)	**Control Group (Bedele Zone)**
Woredas Phase I	3	
Woredas Phase II	2	
Total	5	4
Health Centers Phase I	8	
Health Centers Phase II	6	
Total	14	10
Catchment Population[a]	476,484	252,132

**Note*. The HC catchment population refers to the average catchment population during the four-year study period from September 2018 to August 2022. Estimations are described in Appendix 3.

Sample (size) representativeness

Given that available data were health center or program level data and not at individual level, no assessment could be performed to understand whether the sample is representative of the population in Jimma or beyond. Hence, we do not know whether the demographics of the sample, such as age and level of education, are representative to that of Jimma. It is only possible to show how the sample size compares to the total population in Jimma Zone. Table 2 shows that the sampled woredas, health centers and catchment populations represent 23.81%, 11.57% and 13.44% of the Jimma Zone, respectively. For Bedele Zone, it is even more difficult to discuss the representativeness of its sample size because this information was not available. However, given that during the end-term evaluation the same sampling technique was used for both the intervention and control zones, it is expected that the sample representativeness is similar to that of Jimma. Finally, it should be noted that the number of health centers sampled is small.

Table 2. Sample size of PBF woredas, health center and catchment populations compared to Jimma Zone

Group	Sample	Jimma Zone	Sample as % of population
		Woredas	
Phase I	3	13	23.08%
Phase II	2	8	25.00%
Total	5	21	23.81%
		Health Centers	
Phase I	8	64	12.50%
Phase II	6	57	10.53%
Total	14	121	11.57%
		Catchment Population*	
Phase I	279,146	1,862,860	14.98%
Phase II	197,338	1,681,132	11.74%
Total	476,484	3,543,992	13.44%

**Note*. The HC catchment population refers to the average catchment population during the four-year study period from September 2018 to August 2022. Estimations are described in Appendix 3.

4.1.4 Effectiveness Measurement

PBF aims to improve, among others, service utilization and quality of healthcare, therefore, both measurements should be incorporated into the effectiveness analysis (Shepard et al., 2020). However, given the lack of appropriate baseline data available on the *quality* of services for the control group, the quality component was excluded from the effectiveness analysis. This is assumed to underestimate the cost-effectiveness of the program, as the quality of care appeared to have significantly improved based on an average increase in the technical quality score for all Phase I and II PBF health centers from 18% during baseline to 81% during endline, while control facilities increased their technical quality score from 17% to 44% (Icos Consulting Plc., 2023). However, it should be noted that baseline time periods were not consequent between the various groups. That is, Phase I health center baseline values were from October 2019, while the baseline values for Phase II and the control group were from December 2020 (mid-term). Moreover, the control sample during the mid-term were a distinct set of facilities from the end-term. As a result, *service utilization* is the only output measure used to estimate program effectiveness in terms of QALYs gained. This still allows for comparison with other CEA results in sub-Saharan Africa.

4.1.5 Health Services

This cost-effectiveness analysis modelled effectiveness of five services, while the program incentivized health centers for 24 services. First, health services were excluded if they could not be modelled using the LiST, which is required for this analysis. LiST focusses on modelling MCH interventions (Shepard et al., 2015). Second, services were excluded if the volume of service utilization was low or had many missing observations. As a result, Table 3 shows the indicators included in the analysis. Appendix 1 provides an overview of all PBF indicators, including reason for exclusion.

Table 3. Definitions of Health Services Included in Effectiveness Analysis

Indicator	Definition	Price (ETB) *before* January 2021	Price (ETB) *from* January 2021
First Antenatal Care Visit (ANC1)	*Number* of pregnant women that received antenatal care services at least 1 time before 16 weeks (i.e., up to and including 15 weeks)	80	120
Four Antenatal Care Visits (ANC4)	*Number* of pregnant women received antenatal care services at least 4 visits	80	120
Skilled Birth Attendance (SBA)	*Number* of births attended by skilled health personnel (health officer/nurse/midwife)	160	240
Immunization of Children < 1 year (Full Immunization)	*Number* of surviving infants who received all required vaccines doses before their first birthday. Vaccines include: -BCG (1 dose) -OPV (3 doses) -DPT-Hep-Hib (Pentavalent) (3 doses) -PCV (3 doses) -Rota (2 doses) -IPV (1 dose) -Measles (1 dose)	64	96
Vitamin A supplementation	*Number* of children between 6 and 59 months supplemented with Vitamin A	8	12

4.2 Comparator

Given that the control group did not implement PBF, nor did it receive any other form of additional financing, the comparator for this study is status quo input-based financing, as described in section 3.2. Therefore, this study does not follow the standard CEA application which is used to identify the largest health gain between *interventions* within a budget constrained health system (Drummond et al., 2015). Rather, this CEA tries to identify health gains by comparing two ways of *allocating the budget* to local service providers. While PBF incentives are not funded from the government health budget, a health opportunity cost remains as the external funds for PBF could have been used elsewhere within the system. Hence, an overall constrained budget remains, and a CEA should still provide useful insights. However, given that PBF is an additional form of financing for the intervention group, and the comparator received no additional form of financing, it is not possible to separate out the *incentive* effect from the *resource* effect. This is a limitation.

4.3 Perspective

A healthcare perspective, or more specifically, an implementer perspective was used to assess costs and effects. This perspective excludes costs and benefits incurred outside the healthcare system related to changes in productivity, informal care, and patient costs (Drummond et al., 2015). Given this analysis intends to inform public-oriented parties including implementers (e.g., NGOs), policymakers (e.g., governments), and funders (e.g., donors), which have constrained health program budgets for a limited period of time within which to allocate funds, the healthcare perspective is most appropriate (Shepard et al., 2020). Moreover, a financial perspective is used rather than an economic perspective for two reasons. First, there are no data available for the impact of PBF on productivity and other costs and benefits incurred outside of the health system. Second, the financial perspective considers the notion that decisions are made within a time-limited period, as is the case in the PBF program (Shepard et al., 2015). Thus, only subsidies and program monetary costs related to the PBF intervention are considered, while future medical expenses due to lives saved are excluded in the financial perspective.

4.4 Time Horizon

The time horizon utilized for a CEA should be sufficient to account for all differences in costs and effects between the intervention and control group (Drummond et al., 2015). When changes in mortality are expected, a lifetime horizon is the appropriate time period. This requires both costs and effects to be estimated within the same timeframe. For this study, the program time span between April 2019 and August 2022 is used to measure effects, more detail was provided in section 4.1. LiST is utilized to extrapolate these effects to longer-term impacts beyond the evaluation period (Walker et al., 2022). This means program benefits are extrapolated for a longer time horizon than costs, as future (medical) costs due to reduction in mortality is not considered. This may overestimate results. However, the methodology followed suggests utilizing a timespan that corresponds with the period in which decisions are made in relation to funding commitments by donors (Shepard et al., 2015) and other similar studies have followed suit (Zeng et al., 2018; Shepard et al, 2020; Zeng et al., 2022).

4.5 Measuring Costs

When conducting a cost analysis, as part of a CEA, one wants to find the incremental cost between the intervention and the alternative, in this case the additional cost of PBF. Because no reliable cost data was collected during the end-term evaluation at the facility level, all costs considered for this analysis are costs incurred by Cordaid for implementing PBF. It is assumed that status quo, government input financing, did not change because of PBF in implementing facilities. Hence, the incremental costs are the additional financial costs related to implementing PBF.

All financial program cost data were obtained from Cordaid's reporting system, PowerBI, except for subsidies paid to sampled health centers, woreda administrations and the zonal administration. Subsidies were obtained from the Cordaid PBF DHIS2 database, to have segregate subsidy costs for the sampled facilities and administrations. From here on, subsidies are interchangeably referred to as 'direct' costs and all other implementation expenditures as 'indirect' costs (Shepard et al., 2015).

To start the analysis, initially all costs from the start of the PBF program, April 2019 to August 2022, were considered. This also included costs related to a six-month inception phase from April to October 2019. However, because effectiveness is only measured for a sampled number of facilities and five services, the attributable PBF costs used in the CEA were adjusted accordingly. The following steps were taken:

Step 1: Excluding unrelated costs
Step 2: Adjusting costs to catchment sample
Step 3: Adjusting costs to five services modelled
Step 4: Deflating and discounting costs

Each step is explained below. The first step is to exclude any unrelated costs to this study. Unrelated costs excluded from the analysis were:

- Costs for a blockchain component piloted within the program,
- Subsidies for health posts and hospitals
- Subsidies outside the sampled health centers
- Subsidies for non-sampled Woreda Health Offices (WorHOs). Additionally, for the last quarterly incentive included in the study (Q3 2022) only two-thirds of the payment was included since the study period ends August 2023.

- For the Zonal Health Department (ZHD), subsidies for indicators solely related to hospital level support were excluded. Since the ZHD mostly rely on payments for supporting and monitoring hospitals, only four indicators were considered; namely, (1) semi-annual ZHD action plan, (2) monthly reports for WorHO, (3) joint review meetings and (4) number of facilities that participated in the review meeting. Like WorHO subsidies, only two-thirds of the last quarterly subsidy payment were considered.

After excluding unrelated costs, the second step is to adjust the remaining, mainly *indirect* implementation, expenditures to the sample size. Note, this step is not relevant to health center incentives as it was already considered in step one. This step is also distinctly treated for regulator incentives and CBO costs. For indirect costs, the proportion of sampled catchment population to total catchment population in Jimma per study year is used. Refer to Appendix 3 for an explanation on catchment population estimates and refer to Appendix 4 and Table 2 for statistics used in step 2. This method is most appropriate because the catchment population also underlies service coverage measures, and costs need to be measured per capita. Furthermore, a large part of the indirect costs related to PBF programs are the periodic verification costs. A higher catchment population would likely mean a higher patient flow and more time needed to complete data verification. Therefore, the catchment population may be more representative than, for example, the number of health facilities. The only exceptions are for costs related to WorHO incentives, ZHD incentives and Community Based CBO costs. The sampled WorHO costs are adjusted by the share of sampled health centers to total health centers within their jurisdiction. This method is chosen rather than the sampled catchment population in the woreda because most of the work by WorHOs and indicators incentivized are specifically to provide support to individual health centers in their catchment area. This is under the assumption that the administration provides a similar level of support and effort to each health center. For ZHD incentives, the incentives of the four relevant indicators are adjusted to the share of sampled woredas in the Zone as this entity provides more direct support to WorHOs rather than individual health centers. CBO costs are adjusted to sample size by the proportion of sampled health centers in Jimma Zone. In the PBF program, one CBO was contracted per facility to conduct community verifications. Their workload and related costs are independent of a facility's catchment population.

In the third step, the relevant program costs remaining were adjusted to five services modelled. During the study period, there were 23 and 24 health services incentivized from October 2019 to December 2020 and from January 2021 to August 2022, respectively. Because program costs may vary based on the number of incentivized services, these costs were allocated, hence reduced, to the five services in the CEA. Quantity payments to health facilities are paid per service output and can be directly tied to the five services, and incentives for services not in the CEA were removed. Equity payments can also be tied directly to services as it is a percental top-up to each quantity payment per indicator for specific health centers. The quality payment was not allocated to the five services, but rather the total quality payment paid to sampled facilities is included. This is because the quality payment is not linked to specific indicator output, and it is very difficult to decipher how much of the quality payment facilities used to improve the quality of individual services. While it is intuitive to assume this may overestimate costs, realistically the potential bias is unclear because it is unknown what portion of this payment was used for the five analysed services. For joint indirect program costs, it is not possible to pinpoint which costs are related to the five services, and a pragmatic top-down approach was taken (Shepard et al., 2015). In this case, program costs were attributed to the five services based on their *proportion* of quantity and equity payment relative to the total quantity and equity payment earned by sampled health centers for all 23 or 24 services. The quantity payment (q*p) considers both the volume (q) and price (p) of each indicator, therefore providing a direct link to all five indicators. The volume accounts for the variability in time spent on indicators, for example, during verification exercises inducing higher costs. The price considers the priority and complexity of indicators which is also assumed to translate to more time and program resources (van Keulen, 2017). The equity incentive considers the remoteness and difficult contextual circumstances of facilities, which translates into higher travel costs and potentially other program costs, for example, coaching time. Therefore, this method is chosen as it assumed to be more accurate than, for example, allocating equal shares of joint indirect costs to each service. For costs during the inception phase, when no subsidy payments were made, indirect costs were allocated based on data from study year 2, September 2019 to August 2020.

Finally, related PBF costs adjusted for catchment size and five serviced modelled should be deflated and discounted (Shepard et al., 2015). Costs were deflated using a GDP deflater, as an overall index of inflation and expressed in EUR 2019. The GDP deflators were sourced from the World Bank Open Databank and were 18,25%, 21,76% and 34,68% for the years 2020, 2021 and 2022, respectively. (2023). These rates are quite high given Ethiopia has been experiencing significant inflation in the last few years. Costs were discounted monthly using the recommended 3% yearly rate (Shepard et al., 2015).

After each step, the costs remaining are an estimation of the subsidies and indirect costs which are assumed to be attributed to the five services modelled. Note, however that there are a few cost considerations to be mentioned. Second, the increase in consumable costs due to an increase in service volume is excluded to avoid double counting (Zeng et al., 2022). With an increase in service volume comes an increase in PBF incentives, whereby facilities use the earned funds to purchase additional consumables. Thus, the increase in consumables is already partially subsidizes via PBF funding. Thirdly, donor monitoring program costs are not considered as they are not available. Fourthly, all costs were segregated during each step by program phase. For costs made before the expansion as of January 2021 all costs are allocated to Phase I. Finally, incremental costs are estimated per capita to standardize for population size (Shepard et al., 2015).

4.6 Measuring Effects

Based on the CEA methodology followed, several steps need to be taken to move from results at the output level to generic health impact measures such as Lives Saved and QALYs (Shepard et al., 2015). Based on the scope of the study described earlier, the steps include:

Step 1: Converting service utilization statistics to service coverage
Step 2: Estimating the impact of PBF on service coverage
Step 3: Converting health service coverage to the number of lives saved using LiST
Step 4: Converting the number of lives saved to QALYs

These steps are slightly adapted from the toolkit to provide additional clarity, due to differences in data availability and study scope. First, this study utilizes health service utilization data captured at health facility level, while the toolkit builds on household surveys to directly capture service coverage. Hence, step one is an additional step required in this analysis. Second, the scope of this study is limited to measuring program effectiveness through changes in service utilization and does not include changes in quality of care as a measure of effectiveness, as described in section 4.1.4. Each step is further described below, including how it was practically applied to this CEA.

4.6.1 Step 1: Converting service utilization statistics to service coverage

The first step in the analysis was to convert service utilization to service coverage. Before conversion, it is important to understand the definition of service coverage for proper estimation. Service coverage is measured as a percentage and requires an appropriate definition and valid information of both the numerator and denominator. First, the numerator is the number of individuals who receive the service. Second, the denominator is the number of individuals who need the service, that is, the target population. Hence, service coverage means that those individuals who need an intervention receive it (Tanahashi, 1978). This equates to the following notation (Murray et al., 2003):

$$C_j = \frac{N_j}{P_j}$$

where N_j is the number of individuals who received service j, and P_j is the target population. It should be noted that this measure equates to only a partial measure of service coverage because coverage also depends on a sufficient level of quality of care, more commonly known as 'effective coverage' (Tanahashi, 1978).

The numerator

For the numerator, service utilization data for each service were retrospectively collected during the end-term evaluation and was received from external consultants. The data was sourced from the national District Health Information System (DHIS2) for both the intervention and control groups (Icos Consulting Plc, 2023). Hence, the data consists of *declared* data. The data is annualized data from September to August each year, in accordance with the Ethiopian calendar, and was collected for the period September 2018 to August 2022, i.e., 4 years.

Further, it is important to note on the reliability of the data. The service utilization data in DHIS2 is based on routine monthly health facility reports, which facilities *declare* themselves for services delivered over the previous month. Facilities either report into DHIS2 electronically at the facility level, or in case of lack of electricity or internet, they report on paper to their respective Woreda Health Office, where it is uploaded electronically (Adana et al., 2021a). While this routinely collected health information system data should be timely, complete, and accurate for proper evidence-based decision making, the reality looks different and this continues to be a challenge in Ethiopia (Adana et al., 2021b). For example, health facilities may be tempted to over or under-report the number of services provided due to pressure in meeting health targets (Adana et al., 2021b). The reliability of declared data, therefore, pose a challenge especially for the control group.

Another challenge with the data is the difficulty in comparing *declared* data between the intervention group implementing PBF and the control group. As described earlier in the section 3.3.1, facilities implementing PBF are strictly verified monthly whereby the declared data is compared to the original source data, namely facility registers. Registers are checked for completeness and accuracy and each unit of service is only considered 'verified' and paid, if strict guidelines are met according to national standards (Cordaid, 2021). Further, only those indicators are paid for during a specific month when the number of *verified* cases is within a 10% error margin of the number of *declared* cases. Consequently, health facility data for the intervention group may be more reliable and accurate and less likely to over or under-report the true number of services provided. However, the control group did not undergo such verification during the time of the study and may be less reliable. Therefore, comparing declared data between the intervention and control group may lead to bias. Considering that service utilization data is the only data available to conduct this study, this challenge is considered in the sensitivity analysis whereby *declared* data of the control group are adjusted according to potential over or underreporting. This is described in more detail in section 4.8.

The denominator

Estimating the denominator of service coverage rates requires understanding the target population that should receive the service. These calculations are shown in Table 4.2. Note, the target for each of these services is to reach 100% of that specific population. More concretely, 100% of pregnant women are targeted to receive antenatal care services. Hence, the target population is equivalent to the expected number of pregnancies. Given the availability of data, the target population becomes the facility catchment population multiplied by the percentage of crude live births. This same calculation is applied to the skilled birth attendance indicator. Of course, not all pregnancies end in live births therefore this may lead to a small underestimation of the target population and consequently a slight overestimation of coverage. This mostly only affect the ANC1 estimation to a certain extent as there is the highest probability that a pregnancy may not end up in a live birth. The target population is also set to 100% for infant immunizations and Vitamin A supplementation. Finally, it should be noted that the target population used does not consider the level of service delivery within the health system that an individual is expected to receive the care. For example, skilled birth attendance may be provided at both health center and hospital level, as more complicated deliveries may need to be referred to a higher level. The other services are also provided at multiple levels from health posts up to hospitals. However, considering the difficulty in finding what percentage of the population is expected to deliver at a certain level of care, this is not included in the target population estimation. Given the scope of this study is limited to health center level, it is expected that 100% service coverage at this level is not realistic.

Table 4. Description of service coverage definition and target population

Indicator	Target Population (Denominator)
ANC1	Catchment population * % of live births
ANC4	Catchment population * % of live births
Skilled deliveries	Catchment population * % of live births
Fully immunized infants	Catchment population * % of population that are surviving infants at 1 year of age
Vitamin A supplementation	Catchment population * % of population in aged group 6-59 months * 2 doses per year

Note. Source (Cordaid 2021; MoH - Ethiopia, 2017)

The target population for each service was calculated in Excel for each year using population level statistics for the Oromia Region as shown in Table 5. These statistics are currently used by Ethiopian health authorities for service coverage calculations within the Oromia Region and are estimates based on the latest census in Ethiopia in 2007 (Central Statistical Agency, 2010; M.A., Koricho, personal communication, May 26, 2023). To be in line with the ministry of health in estimating service coverage, these statistics are used here. It is not clear, however, whether statistics related to the Oromia Region are also representative specifically to the Jimma and Bedele Zone. However, such data are lacking at Zonal level.

Table 5. Population statistics for Oromia Region, Ethiopia (estimates based on Census in 2007)

No	Statistics	Oromia Region
1	Estimated live births (crude)	3.47%
2	Total number of surviving infants at 1 year of age	3.22%
3	6 – 59 months age group	15.00%

Finally, the largest component of the denominator is the facility catchment population. This data was only available for PBF facilities as of 2021 and for the control facilities for only one point in time, Q3 2022, the final quarter of the study. Hence, there was a need to estimate catchment populations for each of the four study years as explained in section 4.1.3 and Appendix 1. These catchment population estimates, and population statistics were used to estimate annual service coverage rates for each indicator. However, the unreliability of this data is a limitation Adana et al., 2021a).

4.6.2 Step 2: Estimating the impact of PBF on service coverage

To estimate the impact of PBF on service coverage, a difference-in-difference (DiD) method was employed to compare changes in service coverage over time between the treatment and control zones. This method has been commonly used in evaluating PBF programs (Basinga et al., 2011, Bonfrer et al., 2014; Friedman et al., 2016; World Bank, 2016).

The DiD method captures the change in pre- and post-intervention outcomes of the program by comparing it to a counterfactual (Gertler et al., 2016). The counterfactual is the change in pre-post intervention outcomes that would have occurred in the Jimma Zone in the absence of the PBF program. Given that it is not possible to know this, a control group that is not exposed to the program but with the same environmental conditions should be used (Gertler et al., 2016). Based on these presumptions, the difference is taken of the pre-post intervention outcomes of the treatment group, the "first difference." Next, the difference in the pre-post intervention outcomes of the control group, representing the counterfactual, is referred to as the "second difference." Finally, the difference in the pre-post intervention outcomes between the treatment and control groups represents the outcome of the DiD estimation.

For the DiD estimation to be valid, the control group must accurately represent the counterfactual (Gertler et al., 2016). While DiD is able to control for time-invariant factors, it is unable to control for time-varying factors. If trends, time-varying characteristics other than PBF, between the intervention and the control group are not equal, this results in an over or underestimation of the effect of PBF. To estimate the true effect of PBF on service coverage, it must therefore be assumed that no time-varying characteristics (other than PBF) exist, whereby trends in the intervention and control group would move in tandem in the absence of the intervention known as the parallel trend assumption—the main assumption underlying the DiD method (Gertler et al., 2016). While there is no way to prove that time-varying factors do not exist and so, it is impossible to test the validity of the parallel trend assumption, the plausibility of the assumption can be assessed by comparing trends in the pre-intervention period.

The parallel trends assumption can only be assessed for Phase II PBF facilities in the intervention zone, which started the program January 2021. This cannot be examined for the Phase I PBF facilities, starting implementation October 2019, because it requires data for at least two pre-intervention time periods. The annualized data for four years between September 2018 to August 2022 do not allow for this. Given that the parallel trends assumption is the main determinant for validity of a DiD, it is a limitation not being able to assess it for the Phase I facilities.

The parallel trends assumption will be assessed separately for all five service coverage outcomes between the Phase II PBF facilities and the control facilities. Two pre-intervention data points are used, including annual data from September 2018 to August 2019 and September 2019 to August 2020. This results in the following model to evaluate the trends between the intervention and control group of the service coverage data:

$$Y_{it} = \beta_0 + \beta_{1,2} * PBFII_i * Year_{pre1,2} + \beta_3 * HC_i + Year_{pre} + \varepsilon_{it}$$

Where,

$PBFII_i$ = Phase II treatment facility fixed effects

$Year_{pre}$ represents pre-intervention year dummies, one dummy for study year 1 (September 2018 to August 2019) and one dummy for year 2 (September 2019 to August 2020)

HC_i are health center indicators (fixed effects) to account for unobserved time-invariant differences

$PBFII_i * Year_{pre}$ pertains to an interaction between the pre-intervention year and treatment group

The interaction terms, β_1 and β_2, are the coefficients of interest to determine whether the parallel assumption appears to hold. An F-test is conducted to test this jointly on the coefficient of interest (World Bank. (2016). The null hypothesis is that the intervention-year interaction term jointly equal zero. The higher the p-value of the interaction term coefficient, the more likely we fail to reject the null hypothesis and the more plausibility for the parallel trends assumption to hold.

Difference-in-Difference Model

To estimate the impact of PBF on service coverage, ten different linear DiD models are estimated. Specifically, two models are assessed for each of the five services. The first model estimated the overall effect of PBF on coverage rates for Phase I and Phase II facilities together, and the second model assessed this separately for each phase of facilities. The latter is to account for several changes in the program that occurred during the introduction of Phase II facilities, most notably that the indicator prices increased by 1.5 times. The specification of Model 1 is as follows:

$$Y_{it} = \beta_0 + \beta_1 PBF_i * Post_{it} + \beta_2 * HC_i + Year_t + \varepsilon_{it}$$

Where,

PBF_i = 1 if treatment facility (Phase I or II)
$Post_{it}$ = 1 if after PBF implementation in treatment groups (as of study year two for Phase I and year three for Phase II facilities),
HC_i are health center indicators
$Year_t$ are year dummies to account for time-specific confounders
β_1 is the interaction term and coefficient of interest which measures the treatment effect

Model 2 is as follows:

$$Y_{it} = \beta_0 + \beta_1 * PBFI_i * PostI_t + \beta_2 * PBFII_i * PostII_t + HC_i + Year_t + \varepsilon_{it}$$

Where,

$PBFI_i$ = 1 if Phase I treatment facility
$PBFII_i$ = 1 if Phase II treatment facility
$PostI_{it}$ = 1 if after PBF implementation Phase I (as of year 2) and Phase I facility
$PostII_{it}$ = 1 if after PBF implementation Phase II (as of year 3) and Phase II facility
HC_i are health center indicators
β_1 and β_2 are the coefficients of interest measuring the treatment effects of Phase I and Phase II, respectively.

All DiD models were performed in STATA 17.

4.6.3 Step 3: Converting service coverage to the number of lives saved using LiST

LiST is used to convert changes in service coverage to the number of lives saved. This modelling tool is the most comprehensive tool to implement this task and is widely used (Shepard et al., 2015). By taking the results from the DiD analysis, the change in service coverage attributable to PBF can be fed into LiST, which uses multiple scientifically backed inputs to mathematically calculate the number of Lives Saved based on this change in service coverage. More specifically, this requires developing a scenario in LiST that reflects what occurred to service coverage due to PBF, and it requires developing a scenario what would have occurred in the treatment area without PBF (Shepard et al., 2015). In the second scenario, the 'counterfactual' would, in that case, have started with the actual *pre-intervention* service coverage value, and it would evolve from year-to-year based on the coverage changes observed in the control group. Hence, the 'counterfactual' scenario was adjusted to (1) the PBF baseline sample population set equal to year 1 for Phase I (264,324) and year 2 for Phase II (197,071), (2) baseline coverage rates were set equal to year 1 or 2 PBF coverage rates depending on phase I or II, and (3) for subsequent years the coverage rates are based on the baseline coverage, as just described, plus the change in the control area from the first year to the year of interest (Shepard, et al. 2015). For the latter, the 'counterfactual' is constructed using the year effects from the DiD estimates. The lives saved under the 'counterfactual' scenario then reflects the number of lives saved in the intervention area exclusively resulting from the coverage trends observed in the control area. In this research, this requires developing four scenarios for each DiD model: one for Phase I facilities, one for Phase II facilities, one for counterfactual for Phase I facilities and one for the counterfactual for Phase II facilities. The differences in Lives Saved between the two PBF scenarios and the two counterfactual scenarios can then be associated to PBF. Life years gained are discounted at 3% annually back to April 2019, similar to costs.

Lives Saved Tool (LiST) Summary

LiST, developed by John Hopkins University, is a deterministic multi-cause modelling tool that measures the impact of expanding the service coverage of MNCH interventions on mortality in LMICs (Winfrey et al., 2011). LiST is part of the Spectrum software, which links it with other applications such as DemProj, FamPlan and an HIV/AIDS tool. The model uses a linear function to estimate a fixed relationship between inputs and outputs. One output, i.e., impact measurement, the model estimates is the number of Lives Saved (Walker et al., 2013). To measure Lives Saved, the model draws on multiple country-specific inputs. These inputs include demographic and socio-economic estimates and projections, mortality rates, deaths by cause, intervention effectiveness, intervention coverage and health status information (Boschi-Pinto et al., 2010). LiST uses preloaded data for its inputs from international literature such as Demographic and Health Surveys (DHS), Multiple Indicator Cluster Surveys, WHO databases and the Child Health Epidemiology Reference Group (Winfrey et al., 2011). In this case, the Spectrum sub-national wizard was utilized meaning that wherever available default inputs from the Oromia Region informed the projections, specifically from DHS 2019. If certain sub-national default statistics were unavailable, national, or sometimes international default inputs are used to model Lives Saved.

4.6.4 Step 4: Converting number of lives saved to QALYs

The last step of the effectiveness analysis is to convert the number of lives saved to QALYs gained. According to the CEA guideline for PBF (Shepard et al., 2015), this can be done by applying the formula for fatal cases separately for the lives saved of pregnant women and of children less than five years (Sassi, 2006):

$$QALYs\ gained = Q * \frac{1 - e^{-r(L-a)}}{r}$$

Where Q is the average quality of life of a surviving individual, r represents a 3% discount rate, and $(L - a)$ is the remaining life expectancy of an individual whose death was averted at age a. QALYs gained then represent the number of QALYs per life saved. Replicating similar PBF cost-effectiveness studies (Zeng et al., 2018; Shepard et al., 2020; Zeng et al., 2022), the formula was applied using Ethiopia's life expectancy from life tables (WHO, 2020) and its disease burden from the Global Burden of Disease Study (GDB 2016 DALYs and HALE Collaborators, 2017) to estimate Q. While there is no explicit explanation for calculating Q, the similar studies appeared to use healthy life expectancy at birth in proportion to overall life expectancy at birth. To verify this, it was attempted to replicate estimates of the number of QALYs gained per life saved for these studies. For one study this estimate could be replicated exactly but for two other studies the QALYs gained estimate was on average 0.71 higher for a children <5 life saved while it was on average 1.45 lower for pregnant women life saved. Therefore, to be conservative, the lower bound of healthy life expectancy at birth from the GBD study to total life expectancy at birth was used for estimating QALYs for children, while the mid-point was used for estimating QALYs of pregnant women. For life expectancy of pregnant women, the average life expectancy of all reproductive years (15-49) was taken, while the average life expectancy of ages 0-5 was taken for children. Consequently, QALYs gained per one life saved were estimated at 24.43 for children <5 and 21.86 for pregnant women. These statistics were multiplied by the number of lives saved to estimate the QALYs gained.

4.7 Measuring Costs-Effectiveness

Based on the outcomes of the cost and effectiveness analyses, the incremental cost-effectiveness (ICER) per capita was generated for three scenarios. One scenario produces an ICER for all phase I and II facilities to measure the overall cost-effectiveness of PBF. A second and third scenario captures the ICER for phase I and II facilities separately to understand whether there are differences due to changes in program design or maturity. The ICER per capita is measured as follows (Shepard et al., 2015):

$$\text{ICER} = = \frac{PBF\ costs\ per\ capita - Status\ quo\ costs\ per\ capita\ (zero)}{PBF\ QALYs\ per\ capita - Status\ quo\ QALYs\ per\ capita}$$

As mentioned, the cost of status quo financing is assumed not to change and is set to zero. An overall ICER is generated rather than for each individual modelled service because PBF programs do not incentivize single services but purchase a package of services. Therefore, it is more insightful to understand the cost-effectiveness of a package which may be more representative of the cost-effectiveness of the program holistically.

To assess the cost-effectiveness of an intervention, the ICER needs to be compared to a threshold representing the opportunity cost in a health system (Drummond et al., 2015). While the GDP per capita threshold is recommended by the CEA guidelines followed (Shepard et al, 2015), this threshold has been criticized (Marseille et al., 2014). Still to provide some level of insight, a cost-effectiveness threshold is considered but along with comparisons to other relevant interventions and sensitivity analysis. A study on MCH interventions valued a healthy life year at 1.5 times a country's GDP per capita (Stenberg et al., 2014). This was used as the threshold for recent CEAs of PBF programs (Zeng et al., 2018; Zeng et al., 2022) and is also used here. The GDP per capita in 2019 for Ethiopia was US$840 or €751 based on an average 2019 official exchange rate from USD to EUR (World Bank Open Databank, 2023). The threshold is therefore €1,126.

4.8 Sensitivity Analysis

Due to several types of uncertainty in economic evaluation, it is important to assess how this impacts the results (Drummond et al., 2015). As LiST is a deterministic model, a probabilistic sensitivity analysis was not possible. Instead, a one-way sensitivity analysis was conducted to individually adjust key parameters to test the robustness of the model (Shepard et al., 2015). Given the assumption that the largest uncertainties pertain to estimating effectiveness, two one-way sensitivity analysis were performed by adjusting effectiveness parameters.

First, a one-way sensitivity analysis was performed by adjusting parameters related to the number of services provided in the control group. This sensitivity analysis is chosen because HMIS data is known to be of sub-standard quality (Shepard et al., 2015). Therefore, it appears difficult to compare *declared* data between intervention and control group, as the PBF facilities undergo rigorous monthly data verification that the control facilities do not. This means that PBF facilities are more likely to report the true number of services provided, while control facilities may be tempted for several reasons to over or underreport service provision (Adana et al., 2021b; Endriyas et al., 2019; Kebede et al., 2020). Therefore, it is important to know in how far the results are sensitive to adjustments in service provision parameters in the control facilities.

Several sources were searched for the level of over or under reporting by indicator within Ethiopia, of which two were used to construct a 'correction factor' (Endriyas 2019, Solomon 2021). Given that within PBF a 10% data error margin was allowed for indicator payment, the correction factor only adjusts for over/under reporting beyond this tolerance level. Further, only a 'correction factor' is applied when information on data accuracy could be sourced specific for each indicator. This resulted in correction factors as shown in Table 6. These correction factors were multiplied to the number of services for each year during the study period per indicator. No correction is applied to ANC1 or Vitamin A Supplementation due to unavailability of data. Service utilization for the three other indicators is reduced in the control group due to consistent over reporting beyond 10% precision (Endriyas 2019, Solomon 2021).

Table 6. Correction factor applied to control group service utilization data

Indicator	Correction factor for over/under reporting >10%
ANC1	1
ANC4	0.84
Skilled Delivery	0.96
Full Immunizations <1	0.93
Vitamin A Supplementation	1

A second sensitivity analysis was conducted which removed the most uncertain indicator estimates, based on the DiD estimates, from the effectiveness analysis. That is, within LiST the service coverage rates between the PBF and control groups were equalized to not allow the model to detect changes in mortality due to those services.

5. Results

Cost and effects are first analysed separately. Next, cost-effectiveness is analysed. Finally, to check the robustness of these findings, results from a sensitivity analysis are shown.

5.1 Costs

From the healthcare, or more specifically the implementer perspective, and with a focus on financial costs, the starting point of the cost analysis was the total cost of the PBF program. From April 2019 to August 2022, the total program costs were €9,5 million (€6.8 million in 2019 EUR), as shown in Table 7, and is based on 9,548 individual financial transactions retrieved from Cordaid's reporting system. The amounts between actual costs and costs in constant 2019 EUR vary widely due to high inflation rates in recent years (World Bank, 2023). The largest share of costs went to health facility and regulator incentives constituting €4.5 million (€3 million in 2019 EUR) or 47% of total costs. The next highest costs were related to human resources at all program levels including at field, country and global level which amounted to €2.3 million (€1.7 million in 2019 EUR) or 25% of total costs. The total costs also include expenditures during the inception phase from April 2019 and September 202. This amounted to €0.39 million.

Table 7. Total unadjusted PBF program costs for all health centers in Jimma, April 2019 to August 2022

Cost Category	EUR (Actual)	EUR 2019	Share of Cost (EUR Actual)
Incentive payments (Health Facilities)	€ 4,170,294	€ 2,779,079	44%
Incentive payments (Regulators)	€ 300,636	€ 213,864	3%
Incentive payments (COVID-related)	€ 25,003	€ 21,143	0%
Human Resources	€ 2,326,472	€ 1,729,080	25%
Verification Costs (incl. CBOs)	€ 298,098	€ 210,701	3%
Trainings, meetings, workshops	€ 256,332	€ 195,839	3%
DHIS2 Data System	€ 208,202	€ 174,200	2%
Blockchain Pilot	€ 367,022	€ 326,192	4%
Monitoring and Evaluation	€ 189,235	€ 154,731	2%
Investments, office, and travel costs	€ 783,294	€ 589,846	8%
Overhead %	€ 568,089	€ 402,994	6%
Total Costs	€ 9,492,678	€ 6,797,669	100%

As explained in the methodology in section 4.5, several steps were taken to estimate the costs relevant to the sampled PBF health centers for the five services modelled. The estimation of costs during each step is shown in Table 7. Of the total program costs, €4.9 million (52%) were found to be relevant to this study. The substantial reduction in costs was attributed to excluding incentives for non-sampled facilities and incentives at hospital and health post level. In step 2, after reducing costs to sample size, the total cost remaining was €1.18 million, 24% of relevant costs. During step 3, costs were adjusted to the five services modelled based on the share of quantity/equity payment for the five services to all 23 and 24 indicators. It is found that the quantity/equity subsidy for the five services constituted €134,907 or 36% of the total quantity/equity subsidy of €376,559, as shown in Figure 4. In particular, 21% (€80,335) of the total quantity/equity payment were incentives for skilled deliveries services. This is expected because this service is incentivized the highest. Together with the quality bonus and small retentions of €131,091, this amounts to €265,998 in health center subsidies included in the CEA before deflation and discounting. Hence, the quality payment makes up almost half of the health center subsidy calculation. In Appendix 5, these estimates are provided for each individual study year and segregated by phase. The share of quantity/quality subsidies for five service to all PBF indicators, which is used to allocate indirect costs to the modelled services, is also shown. It means that 36% of indirect costs are allocated to the five services of which approximately 25.5% to Phase I health centers and 10.5% to Phase II health centers over the entire study period. This makes sense as Phase II health centers received subsidies for five quarters less compared to phase I.

Table 7. Estimated costs during each step of the cost analysis (in EUR)

Cost Category	Step 1: Relevant Costs	Step 2: Adjust to sample size	Step 3: Adjust to 5 services	Step 4: Deflate (EUR 2019)	Step 4: Discount (EUR 2019)	Share of Costs (EUR 2019)
Incentive payments (HC)	504,143	504,143	265,998	172,691	161,037	47%
Incentive payments (Regulators)	75,637	32,693	32,693	21,546	20,937	6%
Incentive payments (COVID-related)	25,003	3,725	1,246	1,054	1,005	0%
Human Resources	2,326,472	329,201	112,966	82,924	78,581	23%
Verification Costs (incl. CBOs)	298,098	37,676	13,344	9,364	8,786	3%
Trainings, meetings, workshops	256,294	36,309	12,612	9,546	9,051	3%
DHIS2 Data System	208,202	30,015	9,344	7,741	7,457	2%
Monitoring and Evaluation	189,235	27,645	10,112	8,293	7,898	2%
Investments, office, and travel costs	783,290	110,303	37,203	27,653	26,279	8%
Overhead %	270,014	69,751	32,000	21,768	20,476	6%
Total Costs	4,936,389	1,181,461	527,516	362,580	341,507	100%

Note. In step 1 to 3 all costs are presented as actual costs. As of step 4, costs are expressed in 2019 EUR using a discount rate of 3%. Step 2 does not apply to health center subsidies because only the subsidies for sampled health centers were already taken in Step 1. Step 3 was not applied to cost category 'Incentive Payments (regulators)' because those incentives would remain irrespective of the number of services targeted. COVID-related incentives are related to both health centers and regulator entities. These incentives were temporary and not recorded in the DHIS2 system, but rather aggregate figures were taken from Cordaid's reporting system. Hence, these incentives were treated as all other indirect implementation costs and not as regular PBF incentives for this cost analysis.

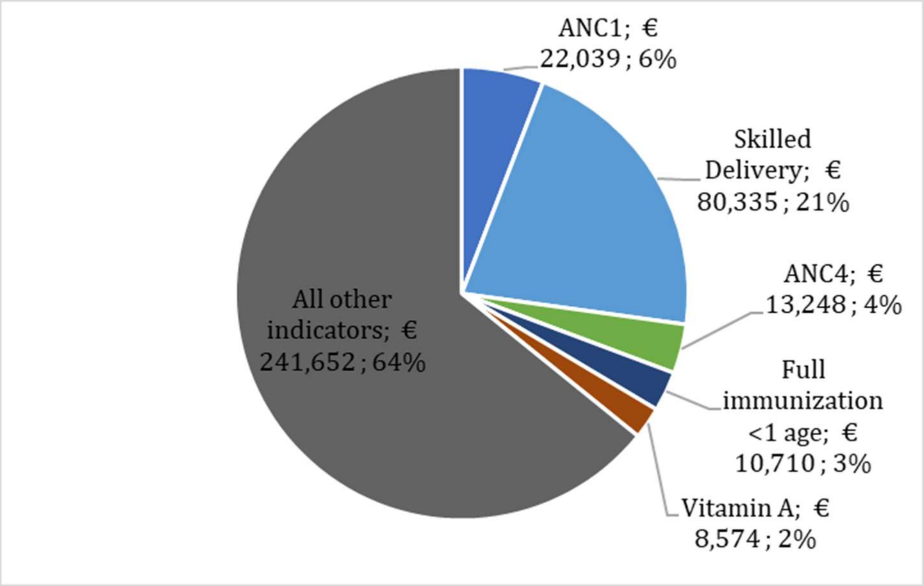


Figure 4. Composition of quantity + equity subsidy payments by service for sampled facilities (Actual EUR)

The final cost estimation of implementing PBF in the 14 sample health centers for the five services after deflating cand discounting costs is €341,507 in 2019 EUR for the entire study period. Of the total costs, 77% (€262,203 EUR 2019) is allocated to Phase I and 23% (€79,304 EUR 2019) to Phase II as shown in Table 8. With an average catchment population of 279,146 this translates to €0.2749 (in EUR 2019) per capita per year for the 3.42 years (41 months) of implementation for Phase I facilities. For Phase II, with an average population of 200,520 during 2021 and 2022, this results in a cost per capita of €0.2373 (in EUR 2019) for the 1.67-year (20 months) implementation period. The cost of the Phase II is likely lower per capita because no inception phase costs were allocated to it. Additionally, Phase II facilities are less mature, hence, they may have earned less incentives due to a lower performance during startup in the first quarters of 2021. Finally, averaging the costs of the two phases leads to an average cost per capita per year of €0.2561 (in EUR 2019).

Table 8. Final cost estimations by implementation phase (in EUR 2019)

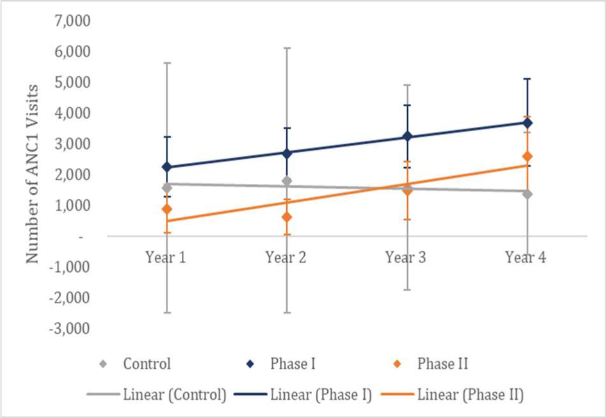
Group	2019*	2020	2021	2022*	Total Study Period	Share of Costs
Phase I Health Center	36,043	85,867	94,950	45,344	262,203	77%
Phase II Health Center	-	-	45,328	33,975	79,304	23%
Total Final Costs	36,043	85,867	140,278	79,319	341,507	100%

**Note*. Costs in 2019 are from April onwards. Costs in 2022 are only to the end of August.

5.2 Effects

5.2.1 Descriptive Statistics

Figures 5-9 show the total service utilization and raw trends over the four-year study period. Note, 'year 1' refers to the period September 2018 to August 2019 and so forth. Phase I started implementation as of Year 2 and Phase II during Year 3. In general, trends of both intervention group phases appear to move in tandem with one another. For four services, the trends of the PBF facilities in service utilization were (slightly) positive, while the number of ANC4 services appears to have declined over the years as shown in Figure 6. However, the number of ANC4 services for the control group appeared to also decline but at a steeper rate. The control group also experienced declining trends in the number of services for ANC1, Immunizations, and had a rather stable trend for Skilled Deliveries and Vitamin A. Note, for the Vitamin A indicator a health center in the control group was dropped due to an extreme outlier that appeared to be erroneous. Including this health center, the trend for the control group was positive for Vitamin A service utilization due to an extreme peak and when dropping this health center, the trend in number of Vitamin A services slightly declines. For the remainder of this analysis, this health center is dropped from the Vitamin A analysis. Besides having an extreme outlier, it can also be seen that the control group has large 95% confidence intervals (CI), much larger than for the intervention groups. This is the case for all five services, but is particularly evident for ANC1, where the CI lower bound is below zero which is realistically not possible. The control group has high variability in the number of each service provided according to the declared data used, resulting in large standard deviations and confidence intervals. Additionally, the control group has numerous missing observations, especially for ANC1, while this is not the case for the intervention group. There are only three control facilities with ANC1 data for all four years—one with over 1,000 ANC1 services annually and the others closer to 100 services annually—which shows the high variability. The difference in confidence intervals between the intervention and control groups could also allude to the difference in data quality across these groups given that accurate data reporting is incentivized for PBF facilities. Finally, it should be noted that the number of ANC4 visits were higher than for ANC1 likely because for the latter only those visits are included which occur before 16 weeks of pregnancy, while the ANC4 indicator is irrespective of the time of the visit. Appendix 5 provides aggregate utilization figures for each service.



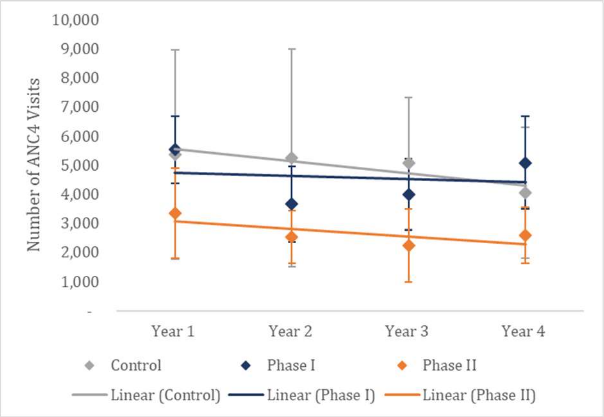
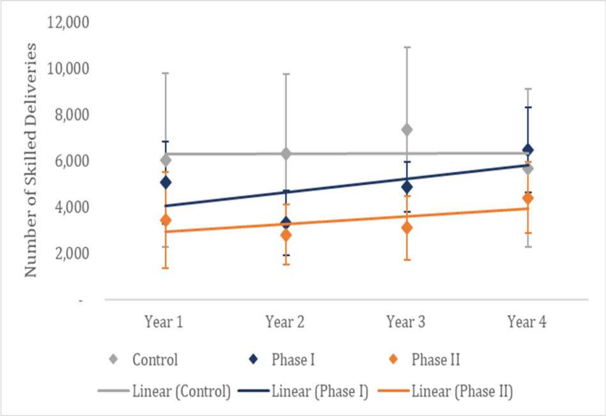


Figure 5. Service utilization: total number and trend of ANC1 visits before 16 weeks for each group during the study period (95% CI)

Figure 6. Service utilization: total number and trend of ANC4 visits for each group during the study period (95% CI)



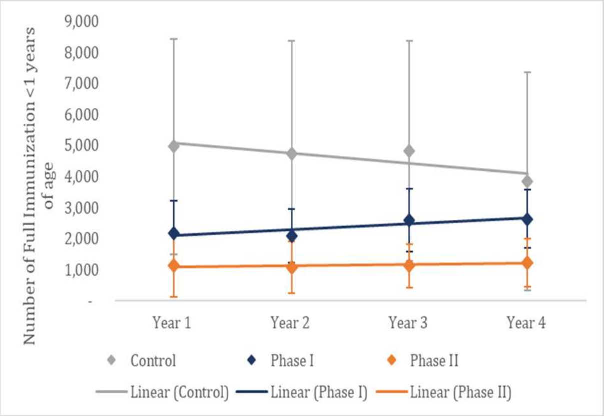


Figure 7. Service utilization: total number and trend of Skilled Deliveries for each group during the study period (95% CI)

Figure 8. Service utilization: total number and trend of Full Immunizations before 1 years of age for each group during the study period (95% CI)

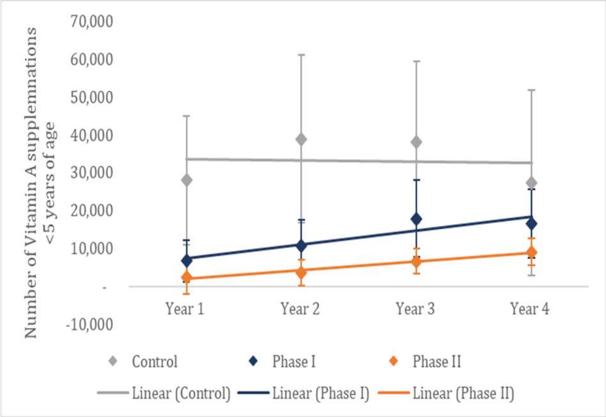


Figure 9. Service utilization: total number and trend of Vitamin A supplementations for children <5 years of age for each group during the study period (95% CI)

5.2.2 Service Coverage

Once target populations were estimated for each service, the service coverage was measured as shown in Figures 10-14. These figures represent the average service coverages for each group during the study period. Not surprisingly, the direction of service coverage trends remained similar to the trends seen for service utilization of each service. In some cases, however, the rate of change of the trend lines does differ between service coverage and service utilization. This is most clear for skilled deliveries in the control group where the downward trend is steeper for service coverage compared to service utilization where it appeared horizontal. This suggests that the target population may have grown at a faster rate than the number of skilled deliveries. Additionally, it is noticeable that the control group started with the highest service coverage in the first year of the study period for all services. By study year four, the control group no longer had the highest service coverage for both antenatal care indicators. On the other hand, Phase II PBF facilities had the lowest starting coverage rates for each service but by year four had a higher coverage rate for ANC1 and skilled delivery services compared to Phase I facilities. Hence, Phase II facilities appears to have increased service coverage rates at a faster pace than Phase I for these services. While it's difficult to say why Phase II facilities appear to have improved these two service coverage rates at a faster pace, it could be that they received better support and mentoring on PBF by Cordaid staff, as the program was more mature and knowledge was already gained during Phase I. Additionally, as Phase II facilities started implementation, incentives for these services were increased by 1.5 times, hence, they may have been more motivated from the start to improve certain services. However, this does not explain why this same trend is not seen for the other three services. In general, there appears to be a convergence in coverage rates between PBF and control health centers in favour for the intervention. Finally, the large confidence intervals for all indicators within the control group are still very evident. This suggests that the large confidence intervals are not simply due to significant differences in catchment size of individual health centers resulting in significant differences in service utilization. If that were the case the confidence intervals should have become smaller for the service coverage estimates. The large confidence intervals in the control group data makes getting precise DiD estimates more difficult. Appendix 7 provides average service coverage rates for each indicator.

Finally, while convergence in service coverage trends between the PBF and control group is favourable for PBF, this non-parallel trend should not be apparent before implementation started. Pre-intervention trends can be visually assessed by comparing the point estimates for year 1 and year 2 between the Phase II and control groups. This shows that ANC1 coverage between Phase II and control facilities may have been moving in opposite directions. During this time, control facilities experienced an upward trend in ANC1 coverage while Phase II facilities experienced the opposite. This suggests that the parallel trends assumption may not hold for this service. However, these diverging trends change to converging trends after Phase II implementation meaning that at least prior to implementation it does not seem that there are confounding factors that improved ANC1 coverage in PBF facilities. For the four other indicators, the trends between the Phase II and control facilities do appear to be moving in the same direction during the pre-intervention period suggesting that the parallel trends assumption may be plausible. Pre-intervention trends cannot be assessed for Phase I facilities.

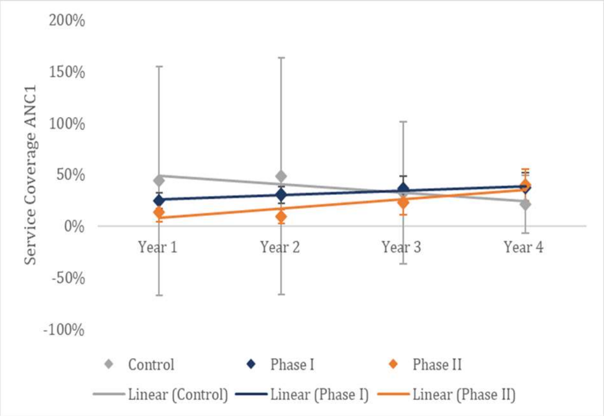


Figure 10. Trends in percent of pregnant women that received ANC1 before 16 weeks during study period (95% CI)

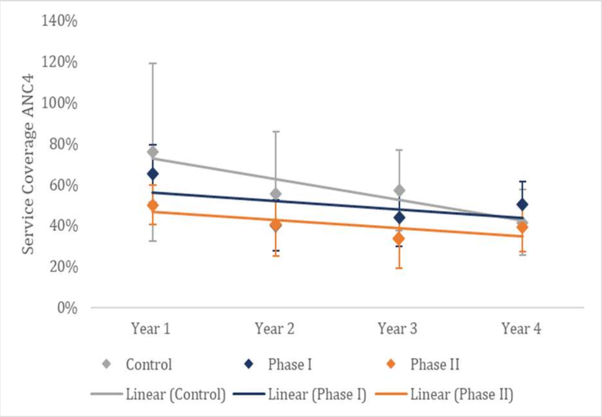


Figure 11. Service coverage: Trends in percent of pregnant women that received ANC4 during study period (95% CI)

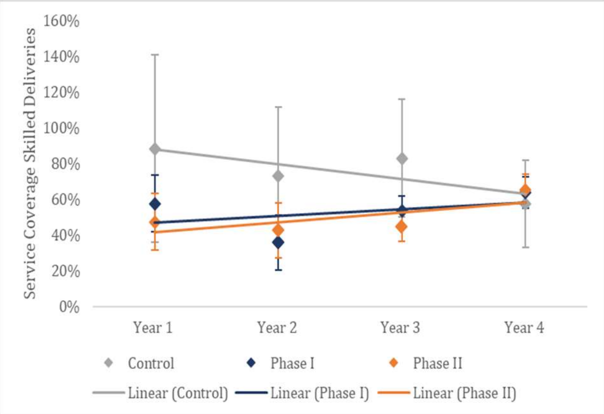


Figure 12. Service coverage: Trends in percent of births attended during delivery by skilled health personnel at a health facility (CI 95%)

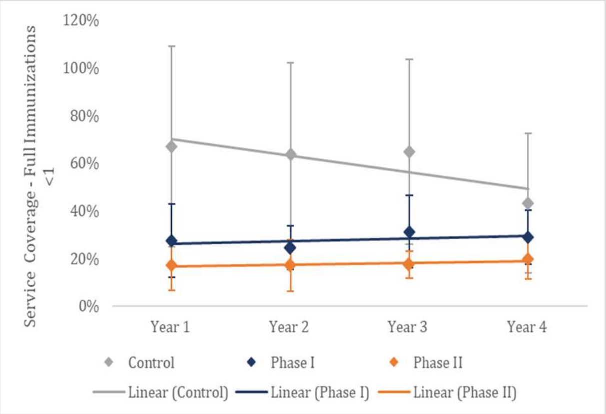


Figure 13. Service coverage: Trends in percent of surviving infant who received all vaccine doses before 1 year of age (CI 95%)

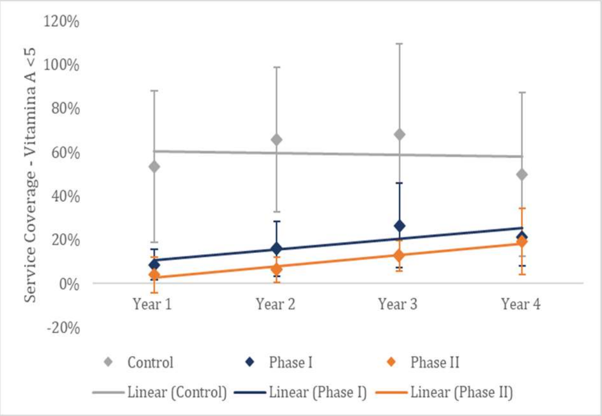


Figure 14. Service coverage: Trends in percent of children aged 6-59 months who received Vitamin A supplementation (every 6 months) (CI 95%)

5.2.3 Difference-in-Difference Model

Table 10 shows the statistical results of estimating the pre-intervention trends of each service for Phase II to assess whether there may be parallel trends. Not surprisingly, when jointly assessing the significance of the interaction terms using the F-test, as described in section 4.6.2, the coefficients are not jointly equal to zero for ANC1 service coverage at a 95% significance level (p = .029). Hence, we reject the null hypothesis that pre-intervention parallel trends exist and brings the plausibility of the parallel trends assumption into question. However, it should be noted that ANC1 service variable was missing quite some observations, specifically in the control group to the point that only nine observations remained. A single health center could have changed the result drastically. Therefore, the result could be a data issue rather than a confounding factor causing differences in trends. The coefficients of interest for each of the other four service were jointly equal to zero. It appears that the parallel assumption is plausible for these services providing credibility for performing the DiD analysis. However, this credibility is limited to Phase II facility results.

Table 9. Estimates of parallel trends of service coverage rates for Phase II health centers, Sept 18 - Aug 20

Indicator	Interaction Study Year 1	Interaction Study Year 2	p-value for F-test
ANC1	-0.121	-0.214	0.031**
ANC4	-0.078	0.039	0.602
Skilled Deliveries	-0.140	-0.037	0.504
Full Immunization <1 years of age	-0.088	-0.036	0.503
Vitamin A <5 years of age	-0.048	-0.104	0.644

Note. *p<0.1 **p<0.05 ***p<0.01. Both models are linear difference-in-difference model, including health center fixed effects.

Two DiD models were conducted for each of the five indicators of interest. The first model produced a DiD model which combined the attribution of Phase I and II together to form one overall estimate. Model 2 provides separate estimates for Phase I and II facilities to understand whether there may be any differences between the two. The results of Models 1 and 2 are shown in Table 5.5. The full output of each model can be found in Appendix 8.

Table 10. Difference-in-Difference estimates of effects of PBF on service coverage rates

Indicators	Model 1				Facility Phase	Model 2				N
	DiD estimate	P-value	95% Confidence Intervals			DiD estimate	P-value	95% Confidence Intervals		
ANC1	0.1878***	0.004	0.0641	0.3116	Phase I	0.1434*	0.073	-0.0139	0.3007	76
					Phase II	0.2188***	0.003	0.0776	0.3600	
ANC4	-0.0016	0.983	-0.1456	0.1424	Phase I	-0.0200	0.834	-0.2100	0.1699	94
					Phase II	0.0134	0.880	-0.1625	0.1893	
Skilled Deliveries	0.0605	0.468	-0.1050	0.2260	Phase I	0.0209	0.848	-0.1970	0.2389	94
					Phase II	0.0926	0.363	-0.1091	0.2944	
Full Immunization of infants <1	0.0979*	0.050	0.0001	0.1958	Phase I	0.1111*	0.091	-0.0180	0.2402	93
					Phase II	0.0870	0.155	-0.0337	0.2076	
Vitamin A for children <5	0.1081	0.159	-0.1050	0.2260	Phase I	0.1105	0.272	-0.0887	0.3097	90
					Phase II	0.1061	0.257	-0.0794	0.2917	

Note. * p<0.1 ** p<0.05 *** p<0.01. Both models are estimated using OLS regression with linear difference-in-difference specifications as described in section 3.5.2. Both models include health fixed effects.

PBF appeared to increase ANC1 coverage significantly and positively by 18.78 percentage points (p = 0.004). Both Phase I and II facilities separately improved ANC coverage by 14.34 and 21.88 percentage points, respectively. This is statistically significant for Phase II facilities (p = 0.003) but is only statistically significant for Phase I facilities at a 90% significance level (p = 0.073). The direction of the estimates aligns with what was initially suggested by Figure 10 where ANC1 coverage appeared to drastically increase for PBF facilities while decreasing for control facilities. When looking at the initial pre-intervention coverage, this means that Phase I facilities increased ANC1 coverage from 23.31% (Year 1 of study) to 37.65%. Hence, it suggests that 61.52% more pregnant women are seeking their first antenatal care visit before 16 weeks. For Phase II, this would translate to an increase in ANC1 coverage from 9.45% (Year 2 of study) to 31.33% resulting in 2.3 times more pregnant women getting their first ANC1 visit before 16 weeks. This magnitude is large. However, parallel trends were not assessed for Phase I meaning that we do not know whether this increase is due to PBF or some other confounding factor. On the other hand, pre-intervention trends for Phase II were decreasing and diverging away from the control group suggesting that the DiD estimate may underestimate the effect of PBF on ANC1 service coverage for this group.

The percentage of infants receiving all required immunizations before their first birthday increased by 9.79 percentage points likely due to PBF (p = 0.050). The estimate was slightly higher for Phase I facilities with a 11.11 percentage point increase in immunization coverage (p = 0.091), while the increase was estimated at 8.70 percentage points for Phase II facilities (p = 0.155). However, this was not statistically significant for Phase II facilities and only significant for Phase I facilities at the 90% significance level. Hence, the estimates of the individual facility phases are less precise.

DiD estimates of the effect of PBF on coverage rates for the three other indicators, were not statistically significant. In fact, when it comes to the ANC4 estimate, there is no precision at all (p =0.983). Looking at Figure 11, the simple service coverage trends, ANC4 coverage appeared to have been declining at a steeper rate for the control group compared to both PBF groups. However, this is not supported by the near-zero DiD estimate, suggesting that PBF did not influence ANC4 coverage. The DiD estimates for Skilled Delivery and Vitamin A coverage were 6.05 and 10.81 percentage points, respectively, but neither were again statistically significant. While the trends in coverage in Figures 12 and 14, suggest a relatively strong convergence in trends in favour of PBF, particularly, for Skilled Delivery, the low precision in estimates may be due to the large confidence intervals of the control group coverage rates. With consecutive large confidence intervals in coverage rates by the control group, it becomes difficult to estimate a precise DiD estimate when the effects are modest. Only large effects of PBF on the dependent variable, such as seen with ANC1 where coverage trends between PBF and control group cross, estimates are precise. Therefore, it is difficult to know the true estimate of PBF on Skilled Delivery and Vitamin A coverage rates.

5.2.4 Lives Saved and QALYs Gained

To estimate effectiveness, robust results on service coverage from the DiD analysis are needed to model lives saved. Based on the statistical outcomes and lack of precise estimates from the DiD analysis, it may not be sensible to use all these figures for modelling the number of Lives Saved using LiST. However, for the purpose of this research, this step will still be conducted using all five services given that costs have also been allocated to these five services. As discussed in the section 4.6.3, this required estimating four separate scenarios for the service coverage inputs into LiST: (1) PBF Phase I, (2) Phase I 'counterfactual' (3) PBF Phase II and (4) Phase II 'counterfactual'. The year fixed effects from DiD model 1 used to construct the 'counterfactual' scenarios, and the service coverage estimates for all four scenarios can be found in Appendix 9. Model 1 LiST output was aggregated to show the overall impact of PBF on the number of lives saved. Model 2 scenarios were constructed similarly and is used to show the separate effects of Phase I and Phase II facilities.

Table 12 shows the number of lives saved resulting from PBF compared to status quo financing for the five maternal and child services of interest. If service coverage inputs into LiST were robust, 48 children <5 were saved in the PBF area and 4 pregnant women lives saved based on DiD Model 1 estimates. Most lives saved occurred within the last study year explained by the largest convergence in service coverage rates between the PBF and control scenarios in favour of PBF. During this year coverage rates for PBF areas mainly increased while the opposite occurred in the control area. Of course, in Year 3 and 4, implementation had also commenced in Phase II facilities. Further, most lives saved were of children under five years likely due to two reasons. First, each of the five services reduce risk factors either related to causes of deaths of neonates or children 1-59 months, while only antenatal care visits and institutional deliveries include several risk factors related to maternal mortality. Secondly, ANC4 and skilled deliveries, which can reduce most risk factors related to maternal mortality out of the study indicators, appeared to be least impacted by PBF.

Table 13 and Table 14 show the results separately for Phase I and Phase II facilities, respectively. Phase I facilities had saved lives of 30 children under five and no pregnant women, while Phase II facilities saved lives of 15 children less than five and two pregnant women. Note, that the number of lives saved are less in the separate outcomes for each facility phase compared to the total PBF lives saved in Table 12 due to year fixed effects from DiD Model 1 which showed a stronger negative trend for the control group compared to Model 2 estimates. Nevertheless, the findings in Tables 13 and 14 translates to 38 lives saved per year per million population within Phase I facility areas based on 35 months of PBF implementation and similarly to 49 lives saved per year per million people for Phase II based on 20 months of implementation. This suggests Phase II appeared to be more effective in reducing mortality. However, there should be caution in interpreting these results as not all DiD estimates were statistically significant and the coverage rate estimates used for modelling the number of lives saved using LiST may not be precise.

Table 11. Number of deaths and lives saved under PBF vs status quo financing (control)

Study Year	Deaths		Lives Saved
	PBF	Control	PBF vs. Control
Children <5			
Year 1 (Sept 18 - Aug 19)	497	497	0
Year 2 (Sept 19 - Aug 20)	877	872	-5
Year 3 (Sept 20 - Aug 21)	834	841	7
Year 4 (Sept 21 - Aug 22)	787	833	46
Subtotal	2,995	3,042	48
Maternal			
Year 1 (Sept 18 - Aug 19)	34	34	0
Year 2 (Sept 19 - Aug 20)	66	65	-1
Year 3 (Sept 20 - Aug 21)	61	60	-1
Year 4 (Sept 21 - Aug 22)	55	60	5
Subtotal	216	220	4
Total	3,211	3,262	51

Note. LiST modelling inputs were based on DiD Model 1 outputs. Deaths and lives saved were discounted at 3% back to April 2019, similar to costs, and includes a mid-year discounting correction. Year 1 is baseline year for Phase I facilities and Year 2 is baseline year for Phase II facilities. Baseline population size: 461,395

Table 12. Number of deaths and lives saved under PBF Phase I facilities vs status quo financing (control)

Study Year	Deaths		Lives Saved
	PBF Phase I	Control	PBF vs. Control
	Children <5		
Year 1 (Sept 18 - Aug 19)	497	497	0
Year 2 (Sept 19 - Aug 20)	514	507	-7
Year 3 (Sept 20 - Aug 21)	481	491	10
Year 4 (Sept 21 - Aug 22)	459	486	27
Subtotal	1,951	1,981	30
	Maternal		
Year 1 (Sept 18 - Aug 19)	34	34	0
Year 2 (Sept 19 - Aug 20)	39	37	-2
Year 3 (Sept 20 - Aug 21)	35	35	0
Year 4 (Sept 21 - Aug 22)	32	34	2
Subtotal	140	140	0
Total	2,090	2,120	30

Note. LiST modelling inputs were based on DiD Model 2 outputs. Deaths and lives saved were discounted at 3% back to April 2019, similar to costs, and includes a mid-year discounting correction. Year 1 is baseline. Population size: 264,324.

Table 13. Number of deaths and lives saved under PBF Phase II facilities vs status quo financing (control)

Study Year	Deaths		Lives Saved
	PBF Phase II	Control	PBF vs. Control
	Children <5		
Year 2 (Sept 19 - Aug 20)	363	363	0
Year 3 (Sept 20 - Aug 21)	353	350	-4
Year 4 (Sept 21 - Aug 22)	328	346	18
Subtotal	1,044	1,059	15
	Maternal		
Year 2 (Sept 19 - Aug 20)	27	27	0
Year 3 (Sept 20 - Aug 21)	26	25	-1
Year 4 (Sept 21 - Aug 22)	23	26	3
Subtotal	76	78	2
Total	1,121	1,137	16

Note. LiST modelling inputs were based on DiD Model 2 outputs. Deaths and lives saved were discounted at 3% back to April 2019, similar to costs, and includes a mid-year discounting correction. Year 2 is baseline. Population size: 197,071. Difference in subtotals and total is due to rounding.

The final effectiveness measure, QALYs, can be found in Table 15. This represents the incremental QALYs gained throughout PBF implementation in comparison to status quo financing. Total QALYs gained for overall PBF implementation during October 2019 to August 2022 was 1,239 of which 1,160 for children and 78 for pregnant women. While most lives saved were seen for children, this is even more pronounced for QALYs gained as more life years can be gained for this younger population. Again, caution should remain in interpreting these results.

Table 14. The number of QALYs gained among pregnant women and children in PBF versus status quo financing, October 2019 to August 2022

Population	QALYs gained		
	PBF Overall (Model 1)	PBF Phase I (Model 2)	PBF Phase II (Model 2)
Children <5 years	1,160	736	355
Pregnant women	78	-2	39
All	1,239	733	394

Note. Lower bound healthy life expectancy was used for calculating QALYs for children <5 years, resulting in an estimation of 24.43 QALYs gained per life saved for children <5. Middle point healthy life expectancy figures were used for pregnant women resulting in an estimation of 21.86 QALYs gained per life saved.

5.3 Cost-effectiveness

Table 16 shows the ICER of PBF implementation in Jimma, Ethiopia compared with status quo financing. The overall PBF program ICER was € 5,649 per life saved (per capita per year) and €233 per QALY gained. For Phase I facilities the ICER was higher at €7,461 per life saved and €305 per QALY gained compared to €4,956 per life saved and €201 per QALY gained. The ICER for Phase I facilities was higher due to incurring both more costs and lower effectiveness compared to Phase II. It should be noted that all inception phase costs were allocated to Phase I, which may slightly overestimate the ICER for Phase I facilities while underestimating the ICER for Phase II.

Table 15. ICER of PBF in Jimma, Ethiopia vs status quo financing (EUR 2019)

Parameters	PBF	PBF Phase I	PBF Phase II
Costs			
Incremental costs (EUR 2019 per year per capita)	€ 0.26	€ 0.27	€ 0.24
Lives saved ICER			
Incremental lives saved	51	30	16
Incremental lives saved (per year per million people)	45	36	47
ICER cost/life saved (per capita per year)	€ 5,649	€ 7,461	€ 4,956
QALYS gained ICER			
Incremental QALYs	1,238	733	394
Incremental QALYs (per year per million people)	1,108	900	1,178
ICER cost/QALY gained (per capita per year)	€ 233	€ 305	€ 201

Note. Lives saved and QALYs rounded down to nearest whole number. In computing the overall PBF ICER, health effects were estimated using DiD model 1 outcomes for Phase I and II separately and later aggregated, separate costs related to each phase were used. Per capita statistics were calculated based on the average population for each phase over its implementing period this was 279,146 for Phase I and 200,520 for Phase II, similar to estimating per capita costs. Implementing years were 2.92 for Phase I and 1.67 for Phase II.

As described in section 4.8, a sensitivity analysis was conducted by adjusting service utilization parameters downward for potential overreporting in the control group. This resulted in an approximate 10% increase in the overall PBF ICER to € 256 per QALY gained. This is because the control group experienced negative trends in service utilization during the study period, hence, reducing this parameter by the same percentage each year reduced the control group's negative time trend. Thereby, convergence in PBF and control group coverage rates, in favour of PBF, was less strong resulting in a higher ICER. In particular, the ICER for Phase I increased to € 336 while it remained the same for Phase II, as it did not result in additional lives saved.

To further check robustness of results, indicators of which effectiveness were insignificant or uncertain based on the DiD estimates were removed from modelling lives saved. That is, lives saved were only modelled for ANC1 and Full Immunization, while costs were maintained. This reduced overall lives saved and QALYs gained to 18 and 461 per one million people, respectively. Consequently, the ICER increased to € 14,331 per life saved and € 587 per QALY gained. This translated to € 468 per QALY gained for PBF Phase I facilities and to € 708 for Phase II. Here, the ICER of Phase II facilities is higher because they were less effective in improving full immunization coverage. Finally, when comparing all ICERs per QALY gained to the GDP per capita threshold of € 1,126, as explained in section 4.7, PBF is potentially cost-effective.

6. Discussion and conclusion

The aim of this research was to assess the cost-effectiveness of PBF in Jimma, Ethiopia compared to status-quo financing from April 2019 to August 2022. In particular, the incremental effects and program costs were evaluated from a health system and financial perspective for five services incentivized through PBF: ANC1, ANC4, skilled deliveries, full immunization of children <1, and Vitamin A supplementation for children 6-59 months. It was found that of the total program costs of €9,492,678 (EUR actual), an amount of €341,507 (EUR 2019) could be attributed to the study population. With costs of status quo financing assumed to remain constant, i.e., zero, the incremental costs of implementing PBF was €0.2561 per capita per year. Incremental costs were €0.2373 and €0.2749 for Phase I and II facilities, respectively. In terms of health effects, 3,211 deaths of children <5 and pregnant women were estimated under PBF while this was 3,262 for status-quo financing. It was found that 51 incremental lives saved translates to 1,239 QALYs gained, or 1,108 QALYs per million people, and an ICER of €233 per additional QALY per capita/year. Phase I facilities within the PBF program appeared less cost-effective (€305/QALY) than Phase II facilities (€201/QALY) due in part by the additional inception phase costs incurred and lower effectiveness during Phase I. Higher effectiveness for Phase II could be attributed to higher incentives from the start of the program as incentives had increased by 1.5 times when Phase II commenced implementation. It could also be that since Phase II started with lower coverage rates compared to Phase I, it was easier for them to encourage women and children to seek care, which becomes more difficult as coverage increases. It could also be a combination of the two. Lower costs were attributed to no inception phase costs allocated to Phase II and there lower starting coverage rates, meaning less incentives were initially paid to those facilities per person per year.

Indifferent of the ICER between Phase I and Phase II, their cost-effectiveness can be judged against a threshold. Stenberg et al. (2014) valued a health life at 1.5 times a country's GDP based on a study on MCH services. With a cost-effectiveness threshold in Ethiopia of €1,126 (1.5 times the 2019 GDP), it suggests PBF was cost-effective. Sensitivity analysis revealed that the ICER remained under the cost-effectiveness threshold at €256/QALY gained when correcting for potential overreporting of service utilization in the control group and at €587/gained when only modelling the effectiveness of ANC1 and Full Immunization being that these services had the most robust DiD estimates. Further, Full Immunizations only contributed to 3% of incentives payments while ANC1 contributed 6%. Given the low coverage of Full Immunization in Ethiopia (EPHI, 2019), incentivizing this service would appear to be a cost-effective within PBF. On the other hand, skilled deliveries are highly incentivized, as is often the case in PBF programs, contributing to 21% of the incentive payment while the findings were not significant. Given the small number of maternal lives saved, this suggests that incentivizing skilled deliveries may not have been as cost-effective in this context as perhaps found elsewhere such as in Nigeria (Zeng et al., 2022).

Besides comparing the results to a threshold, it is important to compare these findings to other cost-effectiveness studies. In Zambia and Nigeria, it was found that PBF was cost-effective (Shepard et al., 2020; Zeng et al., 2018; Zeng et al., 2022). For example, in Zambia it was found that PBF cost $999 per QALY gained when compared to status quo financing and, similarly, in Nigeria PBF cost $787 per QALY gained. While the difference in currency and years make the ICER not directly comparable, PBF in Ethiopia appears cost-effective when compared to these programs. Taking a deeper dive, it appears that both countries gained more QALYs per million people due to PBF (Zambia = 4,107; Nigeria = 6,985) compared to 1,108 in Ethiopia, but both PBF programs were also more costly. That is, in Zambia the PBF program cost approximately $3,52 per capita per year and in Nigeria close to $5,50. This is significantly higher than €0.2561 per capita per year for PBF in Jimma, Ethiopia. While multiple assumptions were made on what costs to include in this CEA and which costs could be attributed to the five services of interest, even when looking at the total unadjusted program costs of 9.5 million EUR, the cost per capita per year would be close to €0,96. This is still significantly lower than these two programs and the benchmark of $3 - $5 USD (World Bank, 2014). In Zimbabwe, PBF was also cost-effective with cost/QALY gained of $1,166, but this is in comparison to additional input-based financing and, hence, is likely lower when compared to status quo financing.

6.1 Limitations

There are several limitations to this study design. First, it is not possible to isolate the incentive effect from the resource effect—the effect of simply providing facilities with additional financial resources with no change in incentives—as some previous studies have done (Basinga, 2011; Zeng et al., 2018; Shepard et al, 2020; Zeng et al., 2022). This is not possible as the control facilities did not receive additional financial resource to match the average PBF payments. Therefore, it makes it difficult to determine whether the outcomes are associated to the incentive effect, the resource effect or both (Vermeersch et al., 2012). However, findings in Nigeria and Zambia comparing PBF to both status quo and additional input-based financing, suggest that ICER will increase when compared to input-based financing, but that PBF remains costs effective (Zeng, et al., 2018; Zeng, et al., 2022).

Second, it is not possible to measure effects on non-incentivised indicators, for example, due to multitasking (Chalkley et al., 2020). However, given that there were a broad range of at least 23 services incentivized throughout the study, including MCH services but also curative services, this may be limited.

Third, spillover effects, such as patients shifting to care at PBF facilities from private clinics or non-participating facilities in bordering woredas, cannot be measured. However, given that PBF was implemented in all government facilities as of 2021, patients would have needed to come from neighbouring Zones to seek care in the implementing Zone. Further, the mid-term review did not find any inter-zonal movement (Mhlanga-Gunda, R., 2021). This is unlikely given that most people seek care by foot. Further, most patients seek care at public facilities (Woldie et al., 2021).

Fourth, no reliable cost estimates were available for control facilities and an assumption was made that control group costs remained unchanged. While some evidence has been found in another country that PBF may divert status quo funding away from intervention facilities and towards those not implementing the program, hence substitute rather than complement government funding, (Friedman et al., 2016) there are no indications of this for the PBF program in Jimma, Ethiopia (E., Ahmed, personal communication, July, 2023). This is assumed because the Jimma Zonal Health Department, the local Ethiopian health authority, have limited power to allocate government funding. Further, PBF program costs were also the only costs considered in the most recent CEA of PBF in Nigeria (Zeng et al., 2022).

Fifth, quite a few assumptions have been made for the target population estimates needed to measure service coverage. However, it has been found that estimated catchment data used as the denominator to calculate target populations may be unreliable and it is unclear in which direction a bias may tend toward (Adana et al., 2021a). Moreover, it was found in Ethiopia that service coverage estimates based on routine HMIS data were not always externally consistent with survey data from the Ethiopian Demographic and Health Survey in 2016 (Adana et al., 2021a). The authors concluded that a large part of the discrepancy may have been due to the uncertainty of population estimates used within the ministry of health, and hence, used in this analysis. Although costly, in the future it is recommended to use household data to measure coverage as this would add robustness to the findings.

Sixth, only five out of all 23/24 indicators are modelled. Hence, costs needed to be attributed to those five services. However, it is not known how exactly facilities distribute the PBF payments towards improving certain services. To be conservate, the total quality payment was considered rather than only a proportion to ensure these costs are not underestimated. At the same time, the incentives for the five modelled services contribute to 26% of the RBF payment, proportionally more than the average per indicator. This is because skilled deliveries contribute significantly (21%) to PBF payments with a high incentive of 260 ETB or approximately 5 EUR per unit. Further, since a financial perspective is taken, there is no consideration of the economic cost related to the level of effort of health workers. It could be that health workers are providing more or less effort to the services not modelled.

Seventh, LiST uses regional and national default inputs and it is unclear whether this is representative of the Jimma Zone.

Eighth, no future medical costs are considered while a lifetime horizon is considered when estimating QALYs or life years. While ideally lifetime costs are considered, it was not included as part of the 'protocol' followed. Additionally, similar studies also did not include an estimate of future healthcare costs (Shepard et al., 2020; Zeng et al., 2018; Zeng et al., 2020), which allow for comparability. Future healthcare costs could also not be modelled more specific than a country's healthcare costs per capita.

Finally, this CEA does not adjust service coverage for the change in quality of care, while improving the quality of services is a large component of the PBF program. Given the quality of care appeared to have improved based on an average increase in the technical quality score for all Phase I and II PBF health centers from 18% during baseline to 81% during endline, while control facilities increased their technical quality score from 17% to 44%, this would underestimate the cost-effectiveness of PBF (Icos Consulting Plc., 2023). For example, in Zambia, Zimbabwe and Nigeria, estimating cost per QALY by adjusting services for improvements in quality of care reduced the ICER by 19%, 45% and 62%, respectively. Therefore, this should be considered when interpreting the results.

It should also be noted that there are two known programs, other than PBF, which were active within the Jimma Zone at the time of the study. The USAID "Transform" program was active during the entire study period and ran from 2017 to September 2022, which enhanced capacity through training at the woreda level and of health facility workers (USAID, 2023). This program was active in the Phase II woredas of this study, as the PBF woreda's were initially strategically chosen during the program inception phase to work in woredas where USAID was not active (Cordaid, 2019). However, it can be found from USAID project reports that Bedele (control)

Zone also appears to be included in the program, and it was mentioned to be one of the highest ranked Zones in the program within the Oromia Region (USAID, 2021). It is not clear from the report, however, which woredas are included. Nevertheless, since the Transform program was active during the entire study period, and if treatment was the same in the Jimma and Bedele Zone, this should have been picked up by the DiD analysis and parallel trends assumption. The second program being implemented was the CBHI, which was being rolled out at woreda level throughout Ethiopia at the time (Cordaid, 2019). While there is data available for the PBF sample of the percentage of households enrolled in CBHI per woreda during the intervention period, this data is not available for the pre-intervention period, nor the control group. Only one study was found that assessed CBHI in the Bedele Zone in January 2021 (Dibaba et al., 2021) and found that at the time six out of the ten woredas in the Zone were implementing CBHI. However, it does not mention which woredas these are. Therefore, it is not clear in how far CBHI was already rolled out in all the sample facilities and how that may have influenced the results. However, as part of the external evaluation of PBF in Jimma Zone, patient exit interviews were conducted and it was found that 77% of the PBF patient sample were covered by public health insurance (CBHI), while this was 84% for the control patient sample (Icos Consulting Plc, 2023). The control patient sample was also found to have had insurance for a longer period compared to the PBF patient sample (30 vs 23 months). While these were patients visiting the sampled facilities, they are not linked to the facility level data used in the DiD analysis. Still, it provides at least some level of insight that the PBF facilities may not have been influenced by CBHI more than the control facilities. Additionally, according to local field staff in Jimma, there were a few other programs ongoing, however, these would not have impacted service utilization at large (K. Woldemariam, personal communication, June 6, 2023). In the control zone there were only a few health interventions. The main interventions were related to training of health workers.

6.2 Policy implications

To conclude, this research contributes to existing evidence on the cost-effectiveness of PBF. It showed that PBF in Jimma, Ethiopia alongside existing provider financing mechanisms were cost-effective, although results should be interpreted cautiously given the multitude of assumptions and limitations mentioned. This provides useful insights as the Ministry of Health in Ethiopia is currently piloting its own national PBF design in other parts of the country. These findings will allow for future comparison in cost-effectiveness of other designs in Ethiopia and bring insights into how the program can be further strengthened. However, while PBF appeared to be cost-effective during the study period, budget constraints require continuous ways of achieving efficiency and effectiveness. Given this, it is important to see whether PBF remains cost-effective once the program matures. For example, in Zimbabwe it was found that incentivizing ANC with existing high coverage left little room for improvement while it contributed to a substantial portion of incentives payments (World Bank, 2016). Hence, it was suggested that incentivizes should be targeted towards low coverage interventions to improve cost-effectiveness (Shepard et al., 2020). This requires continuous progress of monitoring of coverage to strategically update the selection and prices of indicators. Further, given the high inflation in recent years in Ethiopia, it is recommended to periodically review the level of incentives to ensure that it appropriately incentivizes individual health workers and institutions amongst rising costs.

6.3 Further research

Future research on the cost-effectiveness of PBF in Ethiopia could consider adding a quality component to measure 'effective coverage,' if the right data becomes available. As Ethiopia continues to search for the most efficient and effective provider payment mechanisms, it is important to continue measuring cost-effectiveness of other health financing pilots being tested. For example, capitation payments are currently being tested in various parts of the country to finance primary healthcare. It may be interesting to see how the cost-effectiveness of such blended payment models compare to solely implementing PBF on top of status-quo financing (Cattle, D., et al. 2020; Lancet Global Health Commission, 2023).

References

Adane, A., Adege, T. M., Ahmed, M. M., Anteneh, H., Ayalew, E. S., Berhanu, D., Berhanu, N., Beyene, M. G., Bhattacharya, A., Bishaw, T., Cherinet, E., Dereje, M., Desta, T. H., Dibabe, A., Firew, H. S., Gebrehiwot, F., Gebreyohannes, E., Gella, Z., Girma, A., . . . Janson, A. (2021a). Routine health management information system data in Ethiopia: consistency, trends, and challenges. *Global Health Action*, *14*(1). https://doi.org/10.1080/16549716.2020.1868961

Adane, A., Adege, T. M., Ahmed, M. M., Anteneh, H., Ayalew, E. S., Berhanu, D., Berhanu, N., Getnet, M., Bishaw, T., Busza, J., Cherinet, E., Dereje, M., Desta, T. H., Dibabe, A., Firew, H. S., Gebrehiwot, F., Gebreyohannes, E., Gella, Z., Girma, A., . . . Lemma, S. (2021b). Exploring data quality and use of the routine health information system in Ethiopia: a mixed-methods study. *BMJ Open*, *11*(12), e050356. https://doi.org/10.1136/bmjopen-2021-050356

African Development Bank Group. (2021, October 8). *Ethiopia - Country Profiles - 2021*. Retrieved October 20, 2023, from https://www.afdb.org/en/documents/ethiopia-country-profiles-2021

African Union. (2001). *Abuja Declaration on HIV/AIDS, Tuberculosis, and other Related Infectious Diseases.* African Summit on HIV/AIDS, Tuberculosis, and Other Related Infectious Diseases, Abuja, Nigeria. https://au.int/sites/default/files/pages/32894-file-2001-abuja-declaration.pdf

Alebachew, A., Yusuf, Y., Mann, C., Berman, P., Federal Ministry of Health. (2015). *Ethiopia's Progress in Health Financing and the Contribution of the 1998 Health Care and Financing Strategy in Ethiopia. Resource Tracking and Management Project.* Harvard T.H. Chan School of Public Health; Breakthrough International Consultancy, PLC; and Ethiopian Federal Ministry of Health: Boston, Massachusetts and Addis Ababa, Ethiopia. https://www.fpfinancingroadmap.org/data/files/resource-documents/pdf/hcfs_review_final_report_september_2015.pdf

Basinga, P., P.J. Gertler, A. Binagwaho, A.L.B. Soucat, J. Sturdy & C.M.J. Vermeesch. 2011. Effect on maternal and child health services in Rwanda of payment to primary health-care providers for performance: an impact evaluation. *The Lancet 377* (9775):1421-1428.

Binyaruka, P., Lohmann, J., & De Allegri, M. (2020). Evaluating performance-based financing in low-income and middle-income countries: the need to look beyond average effect*. BMJ Global Health, 5*(8), e003136. https://doi.org/10.1136/bmjgh-2020-003136

Boschi-Pinto, C., Young, M. J., & Black, R. E. (2010). The child health epidemiology reference group reviews of the effectiveness of interventions to reduce maternal, neonatal and child mortality. *International Journal of Epidemiology, 39*(Supplement 1), i3–i6. https://doi.org/10.1093/ije/dyq018

Cattel, D., Eijkenaar, F., & Schut, E. (2018). Value-based provider payment: towards a theoretically preferred design. *Health Economics, Policy and Law*, *15*(1), 94–112. https://doi.org/10.1017/s1744133118000397

Central Statistical Agency. (2010). *Ethiopia - Population and Housing Census 2007 - IPUMS Subset.* https://microdata.worldbank.org/index.php/catalog/2747/related-materials

Chalkley, M., Mirelman, A. J., Siciliani, L., & Suhrcke, M. (2020). Paying for Performance for Health Care in Low- and Middle-Income Countries: An Economic Perspective. *Global Health Economics*, 157–190. https://doi.org/10.1142/9789813272378_0006

Cordaid. (2019). *Project Implementation Manual (PIM) for Performance-Based Financing (PBF) in Health (Version 1.0. Valid as of August 2019).* Internal Cordaid Report: unpublished.

Cordaid (2020). *Implementation Manual for PBF Health in Oromia Region, Ethiopia (Version 2.0. Valid as of January 2021).* Internal Cordaid Report: unpublished.

Cordaid. (2021). *Implementation Manual for PBF Health in Oromia Region, Ethiopia (Version 2.2. Valid as of January 2022).* Internal Cordaid Report: unpublished.

De Walque, D., & Kandpal, E. (2022). Reviewing the evidence on health financing for effective coverage: do financial incentives work? *BMJ Global Health, 7*(9), e009932. https://doi.org/10.1136/bmjgh-2022-009932

Diaconu, K., Falconer, J., Verbel, A., Fretheim, A., & Witter, S. (2021). Paying for performance to improve the delivery of health interventions in low- and middle-income countries. *Cochrane Database of Systematic Reviews*, 2021(5). https://doi.org/10.1002/14651858.cd007899.pub3

Dibaba, T. J., Terefa, D. R., & Hailu, B. W. (2021). Willingness to join community-based health insurance and its predictors among informal sectors in Buno Bedele zone, south west Ethiopia. *International Journal of Health Economics and Policy*, *6*(4), 118. https://doi.org/10.11648/j.hep.20210604.12

Drummond, M. F., Sculpher, M. J., Claxton, K., Stoddart, G. L., & Torrance, G. W. (2015). *Methods for the economic evaluation of health care programmes (Fourth).* Oxford University Press.

Endriyas, M., Alano, A., Mekonnen, E., Ayele, S., Kelaye, T., Shiferaw, M. S., Misganaw, T., Samuel, T., Hailemariam, T., & Hailu, S. (2019). Understanding performance data: health management information system data accuracy in Southern Nations Nationalities and People's Region, Ethiopia. *BMC Health Services Research*, *19*(1). https://doi.org/10.1186/s12913-019-3991-7

Ethiopian Public Health Institute (EPHI) and ICF. 2021. *Ethiopia Mini Demographic and Health Survey 2019: Final Report*. Rockville, Maryland, USA: EPHI and ICF.

Ethiopia Health Insurance Service (EHIS). (2022). *Strategic health purchasing in Ethiopia: an assessment and strategic actions to improve purchasing*. Ethiopia Health Insurance Service. https://sparc.africa/download/strategic-health-purchasing-in-ethiopia-an-assessment-and-strategic-actions-to-improve-purchasing/

Ethiopian Statistical Service. (2022). *Population Size by Sex, Area and Density by Region, Zone and Wereda: July 2022.* Retrieved October 20, 2023, from http://www.statsethiopia.gov.et/population-projection/

Fichera, E., Anselmi, L., Gwati, G., Brown, G. W., Kovacs, R. J., & Borghi, J. (2021). Can Results-Based Financing improve health outcomes in resource poor settings? Evidence from Zimbabwe. *Social Science & Medicine, 279*, 113959. https://doi.org/10.1016/j.socscimed.2021.113959

Friedman J, Qamruddin J, Chansa C, Das A. (2016). *Impact evaluation of Zambia's health results-based financing pilot project.* Washington, DC: World Bank.

Fritsche, György Bèla; Soeters, Robert; Meessen, Bruno. 2014. Performance-Based Financing Toolkit. World Bank Training. Washington, DC: World Bank. World Bank. https://openknowledge.worldbank.org/handle/10986/17194

GDB 2016 DALYs and HALE Collaborators (2017). Global, regional, and national disability-adjusted life-years (DALYs) for 333 diseases and injuries and healthy life expectancy (HALE) for 195 countries and territories, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. *The Lancet, 390*(10100), 1260–1344. https://doi.org/10.1016/s0140-6736(17)32130-x

Gertler, P., Martinez, S., Premand, P., Rawlings, L. B., & Vermeersch, C. M. J. (2016). Impact Evaluation in Practice, second edition. In *Washington, DC: Inter-American Development Bank and World Bank eBooks*. https://doi.org/10.1596/978-1-4648-0779-4

Icos Consulting Plc. (2023). *End of program evaluation of the performance-based financing for health program in Oromia and Amhara Regions, Ethiopia. (Final Report).* Internal Cordaid report: unpublished.

James, N., Lawson, K., & Acharya, Y. (2020). Evidence on result-based financing in maternal and child health in low- and middle-income countries: a systematic review. *Global Health Research and Policy*, *5*(1). https://doi.org/10.1186/s41256-020-00158-z

Kebede, M., Adeba, E., & Chego, M. (2020). Evaluation of quality and use of health management information system in primary health care units of east Wollega zone, Oromia regional state, Ethiopia: *BMC Medical Informatics and Decision Making*, *20*(1). https://doi.org/10.1186/s12911-020-01148-4

van Keulen, L., (2017). *A Cost-Effectiveness Analysis of Result-Based Financing: From Theory to Practice.* [Master's thesis, Erasmus University of Rotterdam, Institute of Health Policy & Management]

Koricho, M. A. (2023). Cost-Effectiveness of Implementing Performance-Based Financing in Improving Maternal and Child Health Services in Ethiopia [abstract]. IHEA - 2023 World Congress on Health Economics. https://healtheconomics.confex.com/healtheconomics/2023/meetingapp.cgi/Paper/8201

Lancet Global Health Commission. (2023). Financing PHC in Ethiopia: overview of briefs. *London School of Hygiene and Tropical Medicine*. Retrieved October 27, 2023, from https://www.lshtm.ac.uk/media/71731

Marseille, E., Larson, B. A., Kazi, D. S., Kahn, J. G., & Rosen, S. (2014). Thresholds for the cost–effectiveness of interventions: alternative approaches. *Bulletin of the World Health Organization*, *93*(2), 118–124. https://doi.org/10.2471/blt.14.138206

Ministry of Health - Ethiopia. (2017). *HMIS INDICATOR REFERENCE GUIDE*. Policy and Planning Directorate. http://repository.iifphc.org/handle/123456789/392

Ministry of Health – Ethiopia. (2019). *Essential Health Services Package of Ethiopia*. https://www.hsph.harvard.edu/stephane-verguet/wp-content/uploads/sites/1834/2020/09/Essential_Health_Service_Package_Ethiopia_2019_0.pdf

Ministry of Health – Ethiopia. (2021). *Health Sector Transformation Plan II 2020/21 – 2024/25*. https://e-library.moh.gov.et/library/wp-content/uploads/2021/07/HSTP-II.pdf

Mhlanga-Gunda, R., & Rusakaniko, S. (2021). *EXTERNAL MID-TERM REIVEW REPORT 2020 STRENGTHENING THE HEALTH SYSTEM IN JIMMA& BORANA ZONE (OROMIA REGION, ETHIOPIA) THROUGH PERFORMANCE BASED FINANCING (2019-2023)*. Internal Cordaid report: unpublished.

Murray, C. J. L., Evans, D. B. (2003). Health Systems Performance assessment: Debates, Methods and Empiricism. World Health Organization.

Musgrove, P. 2011. "Financial and Other Rewards for Good Performance or Results: A Guided Tour of Concepts and Terms and a Short Glossary." World Bank, Washington, DC. http://www.rbfhealth.org/system/fi les/RBF%20glossary%20long%20revised.pdf.Health

Nkangu, M., Little, J., Omonaiye, O., Pongou, R., Deonandan, R., Geneau, R., & Yaya, S. (2023). A systematic review of the effect of performance-based financing interventions on out-of-pocket expenses to improve access to, and the utilization of, maternal health services across health sectors in sub-Saharan Africa. *Journal of Global Health*, *13*. https://doi.org/10.7189/jogh.13.04035

Sassi, F. (2006). Calculating QALYs, comparing QALY and DALY calculations. *Health Policy and Planning*, *21*(5), 402–408. https://doi.org/10.1093/heapol/czl018

Salehi, A. S., Borghi, J., Blanchet, K., & Vassall, A. (2020). The cost-effectiveness of using performance-based financing to deliver the basic package of health services in Afghanistan. *BMJ Global Health, 5*(9), e002381. https://doi.org/10.1136/bmjgh-2020-002381

Soucat, A., Dale, E., Mathauer, I., & Kutzin, J. (2017). Pay-for-Performance debate: Not seeing the forest for the trees. *Health Systems and Reform*, *3*(2), 74–79. https://doi.org/10.1080/23288604.2017.1302902

Shepard, D., W. Zeng & H.T.H. Nguyen. 2015. *Cost-Effectiveness Analysis of Result-Based Financing Programs: A Toolkit.* Washington: The World Bank group.

Shepard, D. S., Zeng, W., Mutasa, R., Das, A. K., Sisimayi, C. N., Shamu, S., Banda, S., & Friedman, J. (2020). Cost-effectiveness of results-based financing of maternal and child health services in Zimbabwe: a controlled pre-post study. *Journal of Hospital Management and Health Policy, 4*, 32–32. https://doi.org/10.21037/jhmhp-20-84

Stenberg, K., Axelson, H., Sheehan, P., Anderson, I., Gülmezoglu, A. M., Temmerman, M., Mason, E., Friedman, H. S., Bhutta, Z. A., Lawn, J. E., Sweeny, K., Tulloch, J., Hansen, P. M., Chopra, M., Gupta, A., Vogel, J. P., Østergren, M., Rasmussen, B., Levin, C., . . . Bustreo, F. (2014). Advancing social and economic development by investing in women's and children's health: a new Global Investment Framework. *The Lancet*, *383*(9925), 1333–1354. https://doi.org/10.1016/s0140-6736(13)62231-x

Tanahashi, T. (1978). Health service coverage and its evaluation. *PubMed*, *56*(2), 295–303. https://pubmed.ncbi.nlm.nih.gov/96953

Turcotte-Tremblay, A., Spagnolo, J., De Allegri, M., & Ridde, V. (2016). Does performance-based financing increase value for money in low- and middle- income countries? A systematic review. *Health Economics Review*, *6*(1). https://doi.org/10.1186/s13561-016-0103-9

USAID. (2012). *Health Care Financing Reform in Ethiopia: Improving Quality and Equity*. Retrieved October 21, 2023, from https://www.hfgproject.org/wp-content/uploads/2015/02/Ethiopia_Health_Care_Reform_Brief1.pdf

USAID. (2019). *USAID Transform: Primary Health Care Activity Year V Annual (October 2020 – September 2021) and Quarter IV (July -Sept 2021) Reports.* https://pdf.usaid.gov/pdf_docs/PA00Z654.pdf

USAID. (2023). *USAID Transform: Primary Health Care End of Project Report 2017 – 2022.* https://pdf.usaid.gov/pdf_docs/PA00ZSMH.pdf

Walker, N., Tam, Y., & Friberg, I. K. (2013). Overview of the Lives Saved Tool (LiST). *BMC Public Health, 13*(Suppl 3), S1. https://doi.org/10.1186/1471-2458-13-s3-s1

Walker, N., & Tam, Y. (2022). The role of modeling in evaluation of maternal and child health programs: using the lives saved tool to help answer core evaluation questions. *Global Health Action*, *15*(sup1). https://doi.org/10.1080/16549716.2021.2006421

Winfrey, W., McKinnon, R., & Stover, J. (2011). Methods used in the Lives Saved Tool (LiST). *BMC Public Health, 11*(Suppl 3), S32. https://doi.org/10.1186/1471-2458-11-s3-s32

Witter, S., Bertone, M. P., Namakula, J., Chandiwana, P., Chirwa, Y., Ssennyonjo, A., & Ssengooba, F. (2019). (How) does RBF strengthen strategic purchasing of health care? Comparing the experience of Uganda, Zimbabwe and the Democratic Republic of the Congo. *Global Health Research and Policy*, *4*(1). https://doi.org/10.1186/s41256-019-0094-2

Witter, S., Bertone, M. P., Diaconu, K., & Bornemisza, O. (2021). Performance-based Financing versus "Unconditional" Direct Facility Financing - False Dichotomy? *Health Systems and Reform, 7*(1). https://doi.org/10.1080/23288604.2021.2006121

Woldie, M., Dinsa, G., & Yitbarek, K. (2021). Scoping Report: Financing Primary Health Care in Ethiopia. *London School of Hygiene and Tropical Medicine: Media*. The Lancet Global Health Commission on Financing Primary Health Care. Retrieved October 22, 2023, from https://www.lshtm.ac.uk/media/59786#:~:text=Mobilisation%20and%20allocation%20of%20PHC,and%20government%20contributions%20(9%25).

World Bank. (2016). *Rewarding Provider Performance to Improve Quality and Coverage of Maternal and Child Health Outcomes: Zimbabwe Results-Based Financing Pilot Program Evidence to Inform Policy and Management Decisions*. World Bank.

World Bank. (2020). *Cost-Effectiveness Analysis of Performance-Based Financing for the Delivery of a Health Benefits Package in The Republic of Congo - HRBF Impact Evaluation.* World Bank. http://documents.worldbank.org/curated/en/238291593572868686/Cost-Effectiveness-Analysis-of-Performance-Based-Financing-for-the-Delivery-of-a-Health-Benefits-Package-in-The-Republic-of-Congo-HRBF-Impact-Evaluation

World Bank Open Data. (2023). GDP per capita (current US$) - Ethiopia [Dataset]. In *World Bank Open Data*. https://data.worldbank.org/indicator/NY.GDP.PCAP.CD?locations=ET

World Bank Open Data. (2023). Official exchange rate (LCU per US$, period average) - Netherlands [Dataset]. In *World Bank Open Data*. https://data.worldbank.org/indicator/SP.POP.GROW?locations=ET

World Bank Open Data. (2023). Population growth (annual %) - Ethiopia [Dataset]. In *World Bank Open Data*. https://data.worldbank.org/indicator/SP.POP.GROW?locations=ET

World Bank Open Data. (2023). Inflation, GDP deflator (annual %) – Ethiopia [Dataset]. *World Bank Open Data*. https://data.worldbank.org/indicator/NY.GDP.DEFL.KD.ZG?locations=ET

World Health Organization (WHO) (2011). *Guideline: Vitamin A supplementation in infants and children 6-59 months of age*. World Health Organization.

WHO. (2020). *Global health observatory data repository*. World Health Organization. Retrieved October 29, 2023, from https://apps.who.int/gho/data/view.main.60550?lang=en

World Health Organization (WHO). (2023). *Indicator Metadata Registry list*. Retrieved July 2, 2023, from https://www.who.int/data/gho/indicator-metadata-registry

World Health Organization (WHO). (2022). *Global spending on health: rising to the pandemic's challenges.* Geneva: World Health Organization.

Zeng, W., Shepard, D. S., Nguyen, H., Chansa, C., Das, A. K., Qamruddin, J., & Friedman, J. (2018). Cost-effectiveness of results-based financing, zambia: a cluster randomized trial. *Bulletin of the World Health Organization*, *96*(11), 760–771. https://doi.org/10.2471/BLT.17.207100

Zeng, W., Pradhan, E., Khanna, M., Fadeyibi, O., Fritsche, G., & Odutolu, O. (2022). Cost-effectiveness analysis of the decentralized facility financing and performance-based financing program in Nigeria. *Journal of Hospital Management and Health Policy, 6*, 13. https://doi.org/10.21037/jhmhp-20-82

Appendices

Appendix 1: Indicators and prices of PBF program and inclusion in CEA

Table A 1. Indicators and prices of PBF program and inclusion in CEA

No	Indicator	Definition	Price (ETB) *before* January 2021	Price (ETB) *from* January 2021	Included in CEA	Exclusion Criteria
1	First and repeated visits for FP modern methods (short term) – HC level	Number of women of reproductive age (15-49 years), who are not pregnant and are accepting a short-term modern contraceptive method: oral contraceptives and injectables (new and repeat)	48	72		Not in LiST
2	First and repeated visits for FP modern methods (long term) – HC level	Number of women of reproductive age (15-49 years), who are not pregnant and are accepting a long-term modern contraceptive method: implants and IUCD (new and repeat)	96	144		Not in LiST
3	First ANC visit before 16 weeks of pregnancy	Number of pregnant women that received antenatal care services at least 1 time before 16 weeks (i.e., up to and including 15 weeks)	80	120	Yes	
4	Four Antenatal Care Visits (ANC4)	Number of pregnant women received antenatal care services at least 4 visits	80	120	Yes	
5	Total number of births attended by skilled health personnel	Number of births attended by skilled health personnel (health officer/nurse/midwife)	160	240	Yes	
6	Postnatal care visit within first 7 days	Number of women with their newborn child who attended PNC at least once during the early postpartum period (within 7 days of delivery)	40	60		Not in LiST

7	HIV positive tested Pregnant Women put on PMTCT option B+	Number of HIV positive pregnant and lactating women who were initiated on ART for the first-time during ANC, labor, delivery, or PNC	48	72		Available in LiST but low volume
8	HIV exposed infants who received ARV prophylaxis for 6 weeks and 12 weeks	Number of HIV exposed infants who received ARV prophylaxis at labor, delivery, and PNC	48	72		Not in LiST
9	Immunization of Children < 1 year (fully vaccinated) – HC level	Number of children who received all the required vaccines/primary course doses or surviving infants who receive all doses of vaccines before their first birthday	64	96	Yes	
10	Growth monitoring for children < 2 years – HC level	Number of children measured for weight and height under 2 years by age	8	12		Not in LiST
11	Severe Acute Malnutrition (SAM) children < 5 years	Total number of children screened, who have severe acute malnutrition, are treated, and recovered	40	60		Available in LiST but low volume
12	Vitamin A supplementation (distribution) given to children 6-59 months	Number of children between 6 and 59 months supplemented with Vitamin A	8	12	Yes	
13	Testing for HIV/AIDS	Number of individuals who have been tested for HIV/AIDS and who received their results according to PICT protocol: pregnant, TB+, STI-case and other risk factors	8	12		Not in LiST
14	Cases of STIs treated - *male/female*	Number of male/female OPD visits for STI diagnosis and treated according to national protocols seen by a health officer / nurse	32	48		Not in LiST
15	Cases of Malaria diagnosed positive - *male/female*	Number of male/female malaria cases diagnosed positive (by slide or RDT)	64	96		Not in LiST
16	Cases of diabetic patients put on drug treatment	Number of cases of diabetic patients diagnosed and put on drug treatment	48	72		Not in LiST

17	Cases of hypertensive patients put on drug treatment	Number of cases of hypertensive patients diagnosed and put on drug treatment	48	72		Not in LiST
18	Outpatient Visits for children < 5 years - *male/female*	Number of male/female children under five years consulting health centre (OPD), seen by health officers/nurses	16	24		Not in LiST
19	Outpatient Visits - *male/female*	Number of male/female patients > 5 years consulting the health centre (OPD) investigated or seen by a Health Officer/Nurse	8	12		Not in LiST
20	Total length of stay (in days) in the reporting period	Number of bed days spent by in-patients in the health centre	80	120		Not in LiST
21	Number of people referred to other health facility	Total number of cases referred by the HC to the hospital (copy of the referral form to be checked at the facility)	64	96		Not in LiST
22	Bacteriologically confirmed New Pulmonary TB cases detected in the quarter	Number of bacteriologically confirmed new pulmonary TB cases (by microscopy or GeneXpert)	120	180		Not in LiST
23	Cured PTB+	Number of bacteriologically confirmed pulmonary TB cases (by microscopy or GeneXpert) that were cured	80	120		Not in LiST
24	CAC: Post Abortion Care	Number of women receiving post abortion care services	n/a	120		Available in LiST but low volume

Appendix 2: Health center technical quality indicator domains and weights

Table A 2. Technical quality domains

No	Indicator category Indicator	*Before* January, 2021		*from* January, 2021	
		Indicators	Points	Indicators	Points
1	General Appearance and Safety (Civil Service Reform)	10	17	10	17
2	Administration, financial management, HRM and planning	7	10	7	10
3	Health Management Information System (HMIS) and Supervision	4	7	5	10
4	Infection control and waste management	8	23	8	23
5	General Out-Patient Department (OPD)	6	11	6	12
6	Under 5 OPD	5	8	5	10
7	Emergency services	4	6	4	6
8	Antenatal Care (ANC)	3	8	4	13
9	Maternity services	13	27	13	36
10	Expanded Programme on Immunization (EPI) and growth monitoring (GM)	14	24	14	24
11	Nutrition services	4	5	4	5
12	Inpatient services	3	6	2	5
13	Referral services	4	6	4	6
14	Outreach and health post supervision	4	8	4	8
15	Laboratory service*	10	14	9	13
16	Logistics, medicines and supplies	8	20	8	20
	Total Points:		200		218

Note. Laboratory service is not applicable in health centers that do not have a laboratory.

Appendix 3: Description of catchment population estimates

The catchment population estimates are important during several steps of this CEA. They are important for allocating certain PBF program costs base on the share of the sample population to the total population in Jimma. Secondly, cost and effect outcomes need to be standardized for population size by taking per capita outcomes. Thirdly, they are important for measuring the service coverage of the modelled services as described in section 4.6.2. Therefore, the need for a transparent explanation in one place.

The catchment population represents an estimation of the population within a geographically outlined area that a specific health center serves and are available within each health facility. This data was retrieved by external evaluators for both the intervention and control group per facility for the third quarter of 2022, the last quarter of the study. For all PBF health centers, catchment populations data was also captured quarterly from January 2021 until present in the PBF DHIS2 system at Cordaid. To ensure reliability, the catchment data captured for Q3 2022 by external evaluators was compared with the data found in DHIS2. It matched with 99.99%, therefore the available catchment population data from DHIS2 was also used for catchment population estimates. Because it is unlikely that catchment populations remained the same during the period of the study (World Bank, 2023), estimates for periods where data is unavailable need to account for population growth.

The catchment population is calculated on an annual basis for each facility for all four years of the study. The catchment populations were first estimated for the sampled PBF health centers. The same estimation method was applied for measuring the population of the control health centers. The catchment populations for the non-sampled PBF facilities were calculated separately with distinct assumptions of the population growth rate. Afterwards, catchment data from the period Q1 2021 to Q3 2022 was checked for completeness. First, the available catchment data for January 2021 to August 2022 from the PBF DHIS2 were checked for completeness. 92% of the quarterly data was available. In case figures were missing the average between the catchment area before and after the respective quarter were used, if available. Otherwise, a population growth rate for that quarter was applied based on the average population growth for facilities that did have available data for that quarter. Once all missing value between January 2021 and August 2022 were filled for PBF facilities, the average quarterly growth rate for the seven quarters was calculated. This estimated average growth rate was used to estimate catchment populations prior to 2021 for PBF facilities. This amounted to a quarterly growth rate of 0.7535% for the sampled PBF facilities and represents an annual population grow rate of 3.05%. Hence, it is assumed that the growth rate remained constant during the period of the study. Finally, after quarterly catchment data was calculated, the data was annualized by taking the average from Q2 of a certain year to Q3 the next year.

As mentioned, for control facilities the catchment data was only available for Q3 2022. As Bedele Zone was chosen as a control group during the evaluation due to its sociodemographic similarities, an assumption is made that the quarterly population growth rate of the intervention zone is also applicable to the control zone. Therefore, the same 0.7535% quarterly growth rate is also applied for the control zone and the population annualized in an equivalent manner as was done for the intervention group. For the non-PBF health centers, the average quarterly population growth rates were estimated separately. The estimated quarterly growth rate was 0.4016% or 1.62% annually for this group. Hence, quite a bit lower than the sample group.

Table A 3 Estimated catchment population per health center in intervention and control sample

Health Center Group*	Woreda	Facility Name	Year 1: Sept 18 - Aug 19	Year 2: Sept 19 - Aug 20	Year 3: Sept 20 - Aug 21	Year 4: Sept 21 - Aug 22	4-year average
Intervention Group (Jimma Sample)							
Phase I	Mencho	Bilu Harsu HC	21,460	22,114	22,791	29,356	23,930
	Mencho	Darge Bortolo HC	35,843	36,936	37,942	36,062	36,696
	Mencho	Mole HC	24,678	25,430	27,313	33,738	27,790
	Omo Beyem	Dakano Elke HC	64,725	66,698	68,740	70,455	67,654
	Omo Beyem	Yela sasach HC	25,522	26,300	27,014	27,489	26,581
	Setema	Gatira HC	44,013	45,355	46,586	48,380	46,083
	Setema	Gesecha HC	25,032	25,795	26,584	27,734	26,287
	Setema	Sentema Kecha HC	23,051	23,753	24,454	25,241	24,125
		Total	264,324	272,382	281,424	298,454	279,146
Phase II*	Dedo	Korjo HC	16,185	16,679	16,873	16,528	16,566
	Dedo	Lalo HC	18,679	19,248	19,464	19,219	19,153
	Dedo	Meteso HC	40,538	41,773	42,261	40,963	41,384
	Dedo	Sheki HC	43,009	44,320	44,835	44,253	44,104
	Sokoru	Deneba HC	43,944	45,284	46,671	47,925	45,956
	Sokoru	Gebjiro HC	28,886	29,767	30,652	31,396	30,175
		Total	191,241	197,071	200,756	200,283	197,338
Total catchment population			455,565	469,452	482,180	498,737	476,484
Population growth rate				3.05%	2.71%	3.43%	
Control Group (Bedele Sample)							
	Bedele	Gamada HC	30,347	31,273	32,226	33,208	31,763
	Bedele	Haro Kamise HC	19,861	20,467	21,091	21,734	20,788
	Bedele	Haro Kera HC	11,412	11,760	12,119	12,488	11,945
	Chora	Ababora HC	18,651	19,220	19,806	20,410	19,522
	Chora	Abdella HC	20,580	21,207	21,854	22,520	21,540
	Chora	Gefo HC	25,850	26,638	27,450	28,287	27,056
	Chora	Kiltu Shibo HC	26,181	26,979	27,801	28,649	27,402
	Chora	Kumbabe HC	45,052	46,425	47,840	49,298	47,154
	Diddesa	Chalo HC	22,727	23,420	24,134	24,869	23,787
	Gechi	Hurufa HC	20,230	20,847	21,482	22,137	21,175
Total catchment population			240,892	248,235	255,802	263,599	252,132
Population growth rate				3.05%	3.05%	3.05%	

**Note.* Phase II facilities did not start with PBF until January 2021. The catchment populations marked in grey represent pre-intervention estimated catchment populations.

Appendix 4: Statistics used during Step 2 of the cost analysis

Table A 4 Share of catchment population of sampled health centers in Jimma Zone

Study Year	Year 1	Year 2	Year 3		Year 4
	Sept 18 - Aug 19	Sept 19 - Aug 20	Sept 20 - Dec 20	Jan 21 - Aug 21	Sept 21 - Aug 22
Share of catchment population of sampled health centers					
Phase I health centers	14.64%	14.81%	15.00%	7.87%	8.21%
Phase II health centers[a]				5.61%	5.51%
Total	14.64%	14.81%	15.00%	13.48%	13.71%
Catchment population in sampled health centers in Jimma Zone					
Phase I health centers	264,324	272,382	281,424	281,424	298,454
Phase II health centers[a]	191,241	197,071	200,756	200,756	200,283
Total	455,565	469,452	482,180	482,180	498,737
Catchment population in all health centers in Jimma Zone					
Phase I health centers	1,806,082	1,839,055	1,876,348	1,876,348	1,929,955
Phase II health centers	1,644,143	1,673,453	1,699,477	1,699,477	1,707,455
Total	3,450,224	3,512,508	3,575,825	3,575,825	3,637,410

**Note.* Phase II facilities did not start with PBF until January 2021. The catchment populations marked in grey represent pre-intervention estimated catchment populations.

Table A 5 Share of sampled health centers within sampled Woreda Health Office jurisdiction

Group	Woreda	Sampled Health Centers	Health Centers in Woreda	Share of health centers sampled
Phase I	Mencho	3	6	50.00%
	Omo Beyem	2	4	50.00%
	Setema	3	5	60.00%
Phase II	Dedo	4	8	50.00%
	Sokoru	2	6	33.33%

Appendix 5: Estimated (share of) subsidy payments for sampled health centers by phase

Table A 6. Estimated subsidy payment for sampled facilities including share of five study indicator quantity/equity subsidies to all PBF subsidies (actual EUR)

	Study Year		All 4 years	
Facility Group	Indicator	Quantity Subsidies	Equity Subsidy Bonus	Totals
Phase I and Phase II health centers	ANC1	20,398	1,641	22,039
	ANC4	12,087	1,162	13,248
	Skilled Deliveries	74,876	5,459	80,335
	Fully Immunized <1 age	10,021	689	10,710
	Vitamin A	7,986	588	8,574
	Totals (5 Study Indicators)	125,369	9,538	134,907
	Quality Subsidies (all indicators)			132,802
	Retention			-1,711
	Total Subsidies			265,998
Phase I health centers	ANC1	14,360	1,315	15,675
	ANC4	9,748	959	10,707
	Skilled Deliveries	50,485	4,160	54,645
	Fully Immunized <1 age	8,009	636	8,645
	Vitamin A	5,734	373	6,107
	Totals (5 Study Indicators)	88,336	7,443	95,779
	Quality Subsidies (all indicators)			89,811
	Retention			-984
	Total Subsidies Paid			184,606
Phase II health centersa	ANC1	6,038	326	6,364
	ANC4	2,338	203	2,541
	Skilled Deliveries	24,392	1,299	25,690
	Fully Immunized <1 age	2,012	53	2,065
	Vitamin A	2,253	215	2,467
	Totals (5 Study Indicators)	37,033	2,095	39,128
	Quality Subsidies (all indicators)			42,991
	Retention			-728
	Total Subsidies			81,392
Quantity + Equity Subsidy (all PBF Indicators)				376,559
Quantity + Equity Subsidy (Share 5 Study Indicators) - All Phases				35.83%
Quantity + Equity Subsidy (Share 5 Study Indicators) - Phase I Health Centers				25.44%
Quantity + Equity Subsidy (Share 5 Study Indicators) - Phase II Health Centers				10.39%

Table A 7 Estimated subsidy payment per year for sampled facilities including share of five study indicator quantity/equity subsidies to all PBF subsidies (actual EUR)

Study Year		Year 2			Year 3						Year 4		
		Sept 19 - Aug 20			Sept 20 - Dec 20			Jan 21 - Aug 21			Sept 21 - Aug 22		
Facility Group	Indicator	Quantity Subsidies	Equity Subsidy Bonus	Totals	Quantity Subsidies	Equity Subsidy Bonus	Totals	Quantity Subsidies	Equity Subsidy Bonus	Totals	Quantity Subsidies	Equity Subsidy Bonus	Totals
Phase I and Phase II health centers	ANC1	2,028	111	2,139	1,681	121	1,802	5,538	543	6,081	11,151	866	12,017
	ANC4	341	40	381	1,087	98	1,185	4,435	394	4,830	6,223	629	6,852
	Skilled Deliveries	7,539	580	8,119	4,077	287	4,364	20,560	1,839	22,399	42,700	2,753	45,453
	Fully Immunized <1 age	974	63	1,037	942	54	996	2,836	247	3,083	5,268	326	5,594
	Vitamin A	1,198	88	1,286	667	20	687	2,283	159	2,442	3,839	321	4,160
	Totals (5 Indicators)	12,080	882	12,962	8,455	579	9,034	35,652	3,182	38,834	69,181	4,895	74,076
	Quality Subsidies (all indicators)			11,132			7,363			37,091			77,216
	Retention			-112			-83			-1,516			-
	Total Subsidies			23,982			16,314			74,409			151,292
Phase I health centers	ANC1	2,028	111	2,139	1,681	121	1,802	4,352	471	4,823	6,299	612	6,911
	ANC4	341	40	381	1,087	98	1,185	3,971	365	4,336	4,350	456	4,805
	Skilled Deliveries	7,539	580	8,119	4,077	287	4,364	14,108	1,468	15,576	24,760	1,825	26,585
	Fully Immunized <1 age	974	63	1,037	942	54	996	2,397	233	2,630	3,696	286	3,982
	Vitamin A	1,198	88	1,286	667	20	687	1,695	107	1,802	2,174	158	2,332
	Totals (5 Indicators)	12,080	882	12,962	8,455	579	9,034	26,522	2,645	29,167	41,278	3,337	44,616
	All Quality Subsidies			11,132			7,363			27,572			43,744
	Retention			-112			-83			-789			-
	Total Subsidies Paid			23,982			16,314			55,950			88,360
Phase II health centers	ANC1							1,187	72	1,258	4,852	254	5,106
	ANC4							465	29	494	1,874	173	2,047
	Skilled Deliveries							6,452	1,468	7,920	17,940	928	18,867
	Fully Immunized <1 age							439	14	453	1,573	39	1,612
	Vitamin A							588	52	639	1,665	163	1,828
	Total (5 Indicators)							9,130	538	9,667	27,903	1,558	29,461
	All Quality Subsidies									9,519			33,472
	Retention									-728			-
	Total Subsidies									18,459			62,933
Quantity + Equity Subsidy (all PBF Indicators)				44,895			23,287			114,286			194,091
Quantity + Equity Subsidy (Share 5 Study Indicators) - All Phases				28.87%			38.80%			33.98%			38.17%
Quantity + Equity Subsidy (Share 5 Study Indicators) - Phase I				28.87%			38.80%			25.52%			22.99%
Quantity + Equity Subsidy (Share 5 Study Indicators) - Phase II										8.46%			15.18%

Appendix 6: Number of services per year for each indicator

Table A 8 Number of services per year for each indicator

		Year 1 (Sep 18 - Aug 19)		Year 2 (Sep 19 - Aug 20)		Year 3 (Sep 20 - Aug 21)		Year 4 (Sep 21 - Aug 22)	
Group	Indicator	PBF	Control	PBF	Control	PBF	Control	PBF	Control
Phase I	ANC1	2,255		2,688		3,243		3,687	
	ANC4	5,541		3,670		3,984		5,090	
	Skilled Deliveries	5,068		3,317		4,879		6,471	
	Immunization1year	2,176		2,092		2,601		2,632	
	VitaminA	6,727		10,618		17,855		16,613	
Phase II	ANC1	898		628		1,484		2,593	
	ANC4	3,348		2,531		2,243		2,591	
	Skilled Deliveries	3,429		2,816		3,113		4,408	
	Immunization1year	1,132		1,068		1,126		1,235	
	VitaminA	2,348		3,708		6,619		9,075	
Total	ANC1	3,153	1,580	3,316	1,809	4,727	1,572	6,280	1,381
	ANC4	8,889	5,364	6,201	5,244	6,227	5,063	7,681	4,048
	Skilled Deliveries	8,497	6,022	6,133	6,292	7,992	7,340	10,879	5,677
	Immunization1year	3,308	4,961	3,160	4,722	3,727	4,823	3,867	3,841
	VitaminA*	9,075	27,960	14,326	38,911	24,474	38,221	25,688	27,241

Note. Health center, Kumbabe, in control group dropped for VitaminA indicator due to an extreme outlier.

Appendix 7: Service coverage rates per year for each indicator

Table A 9 Service coverage rates per year for each indicator

		Year 1 (Sep 18 - Aug 19)		Year 2 (Sep 19 - Aug 20)		Year 3 (Sep 20 - Aug 21)		Year 4 (Sep 21 - Aug 22)	
Group	Indicator	PBF	Control	PBF	Control	PBF	Control	PBF	Control
Phase I	ANC1	23.31%		29.11%		36.31%		39.57%	
	ANC4	61.37%		38.77%		43.96%		54.45%	
	Skilled Deliveries	54.15%		34.52%		53.83%		69.49%	
	Immunization1year	25.57%		23.59%		31.06%		32.10%	
	VitaminA	8.17%		14.84%		26.13%		24.01%	
Phase II	ANC1	13.12%		9.45%		23.10%		40.55%	
	ANC4	48.50%		40.34%		34.16%		39.67%	
	Skilled Deliveries	46.06%		42.74%		45.53%		66.34%	
	Immunization1year	16.66%		16.98%		17.68%		19.98%	
	VitaminA	3.89%		6.11%		12.80%		19.31%	
Total	ANC1	18.95%	41.82%	20.68%	47.33%	30.65%	32.97%	39.99%	22.40%
	ANC4	55.85%	72.45%	39.44%	54.76%	39.76%	58.02%	48.12%	43.40%
	Skilled Deliveries	50.68%	84.51%	38.05%	72.23%	50.27%	84.36%	68.14%	60.21%
	Immunization1year	21.75%	63.99%	20.76%	62.91%	25.33%	65.80%	26.90%	45.22%
	VitaminA	6.34%	51.07%	11.10%	64.65%	20.42%	69.18%	22.00%	52.17%

Appendix 8: Difference-in-Difference Model 1 and 2 output

Table A 10 DiD Model 1 for ANC1 Service Coverage

N =	76
Adj R-squared =	0.8084

X	Model 1 Coefficients (β)	P-value	[95% confidence intervals]	
inter_pbf_post	0.1878	0.0040	0.064056	0.311593
Abdella HC	-0.1000	0.5820	-0.462714	0.262714
Bilu Harsu HC	-0.0439	0.7830	-0.362806	0.275039
Chalo HC	-0.0630	0.6680	-0.356857	0.230827
Dakano Elke HC	-0.0514	0.7470	-0.370306	0.267539
Darge Bortolo HC	-0.0614	0.7010	-0.380306	0.257539
Deneba HC	0.0131	0.9320	-0.294619	0.320764
Gamada HC	-0.0255	0.8620	-0.319357	0.268327
Gatira HC	0.0436	0.7850	-0.275306	0.362539
Gebjiro HC	0.0081	0.9580	-0.299619	0.315764
Gefo HC	0.0211	0.8960	-0.302508	0.344664
Gesecha HC	0.1436	0.3700	-0.175306	0.462539
Haro Kamise HC	-0.0942	0.6270	-0.480965	0.292533
Haro Kera HC	-0.0800	0.6600	-0.442714	0.282714
Hurufa HC	-0.1179	0.4580	-0.434844	0.199111
Kiltu Shibo HC	1.1620	0.0000	0.868143	1.455827
Korjo HC	-0.0094	0.9510	-0.317119	0.298264
Lalo HC	0.0406	0.7920	-0.267119	0.348264
Meteso HC	-0.1844	0.2340	-0.492119	0.123264
Mole HC	0.0511	0.7490	-0.267806	0.370039
Sentema Kecha HC	0.0236	0.8820	-0.295306	0.342539
Sheki HC	-0.0344	0.8230	-0.342119	0.273264
Yela sasach HC	0.1461	0.3620	-0.172806	0.465039
Year 2	-0.0721	0.1770	-0.177777	0.033567
Year 3	-0.0785	0.2410	-0.211448	0.054481
Year 4	-0.0342	0.6110	-0.168431	0.099998
Contant	0.1942	0.1840	-0.095256	0.483689

Table A 11 DiD Model 2 for ANC1 Service Coverage

N =	76
Adj R-squared =	0.8078

	Model 2			
X	Coefficients (β)	P-value	[95% confidence intervals]	
inter_pbfi_posti	0.1434	0.0730	-0.0138502	0.3006651
inter_pbfii_postii	0.2188	0.0030	0.0776340	0.3599912
Abdella HC	-0.1000	0.5830	-0.4634503	0.2634503
Bilu Harsu HC	-0.0108	0.9470	-0.3383876	0.3167772
Chalo HC	-0.0632	0.6680	-0.3576885	0.2311893
Dakano Elke HC	-0.0183	0.9110	-0.3458876	0.3092772
Darge Bortolo HC	-0.0283	0.8630	-0.3558876	0.2992772
Deneba HC	-0.0027	0.9860	-0.3128673	0.3075556
Gamada HC	-0.0257	0.8610	-0.3201885	0.2686893
Gatira HC	0.0767	0.6400	-0.2508876	0.4042772
Gebjiro HC	-0.0077	0.9610	-0.3178673	0.3025556
Gefo HC	0.0156	0.9230	-0.3088384	0.3400827
Gesecha HC	0.1767	0.2840	-0.1508876	0.5042772
Haro Kamise HC	-0.0842	0.6650	-0.4723847	0.3038994
Haro Kera HC	-0.0800	0.6600	-0.4434503	0.2834503
Hurufa HC	-0.1180	0.4590	-0.4356309	0.1996110
Kiltu Shibo HC	1.1618	0.0000	0.8673115	1.4561890
Korjo HC	-0.0252	0.8710	-0.3353673	0.2850556
Lalo HC	0.0248	0.8730	-0.2853673	0.3350556
Meteso HC	-0.2002	0.2010	-0.5103673	0.1100556
Mole HC	0.0842	0.6080	-0.2433876	0.4117772
Sentema Kecha HC	0.0567	0.7290	-0.2708876	0.3842772
Sheki HC	-0.0502	0.7470	-0.3603673	0.2600556
Yela sasach HC	0.1792	0.2770	-0.1483876	0.5067772
Year 2	-0.0515	0.3720	-0.1664965	0.0634827
Year 3	-0.0682	0.3150	-0.2033166	0.0668711
Year 4	-0.0242	0.7220	-0.1604710	0.1119857
Contant	0.1842	0.2090	-0.1066288	0.4751141

Table A 12 DiD Model 1 for ANC4 Service Coverage

N =	94
Adj R-squared =	0.5726

	Model 1			
X	Coefficients (β)	P-value	[95% confidence intervals]	
inter_pbf_post	-0.0016	0.9830	-0.1456030	0.1424438
Abdella HC	-0.2000	0.1170	-0.4513386	0.0513386
Bilu Harsu HC	-0.3213	0.0220	-0.5948822	-0.0477484
Chalo HC	-0.5025	0.0000	-0.7538386	-0.2511614
Dakano Elke HC	-0.4463	0.0020	-0.7198822	-0.1727484
Darge Bortolo HC	-0.2738	0.0500	-0.5473822	-0.0002484
Deneba HC	-0.3442	0.0110	-0.6056615	-0.0827589
Gamada HC	-0.6525	0.0000	-0.9038386	-0.4011614
Gatira HC	-0.2463	0.0770	-0.5198822	0.0272516
Gebjiro HC	-0.1992	0.1330	-0.4606615	0.0622411
Gefo HC	-0.1254	0.3630	-0.3984308	0.1476969
Gesecha HC	-0.1688	0.2220	-0.4423822	0.1047516
Haro Kamise HC	-0.1425	0.2620	-0.3938386	0.1088386
Haro Kera HC	-0.4320	0.0020	-0.7050974	-0.1589697
Hurufa HC	-0.3175	0.0140	-0.5688386	-0.0661614
Kiltu Shibo HC	0.4100	0.0020	0.1586614	0.6613386
Korjo HC	-0.3117	0.0200	-0.5731615	-0.0502589
Kumbabe HC	0.0200	0.8740	-0.2313386	0.2713386
Lalo HC	-0.3517	0.0090	-0.6131615	-0.0902589
Meteso HC	-0.4717	0.0010	-0.7331615	-0.2102589
Mole HC	-0.0938	0.4960	-0.3673822	0.1797516
Sentema Kecha HC	-0.3138	0.0250	-0.5873822	-0.0402484
Sheki HC	-0.4442	0.0010	-0.7056615	-0.1827589
Yela sasach HC	-0.2338	0.0930	-0.5073822	0.0397516
Year 2	-0.1807	0.0030	-0.2996635	-0.0618166
Year 3	-0.1783	0.0130	-0.3174299	-0.0390938
Year 4	-0.2066	0.0040	-0.3457632	-0.0674271
Contant	0.9039	0.0000	0.7054955	1.1023030

Table A 13 DiD Model 2 for ANC4 Service Coverage

N =	94
Adj R-squared =	0.5666

	Model 2			
X	Coefficients (β)	P-value	[95% confidence intervals]	
inter_pbfi_posti	-0.0200	0.8340	-0.2100354	0.1699492
inter_pbfii_postii	0.0134	0.8800	-0.1625079	0.1892895
Abdella HC	-0.2000	0.1190	-0.4531615	0.0531615
Bilu Harsu HC	-0.3075	0.0380	-0.5979764	-0.0169589
Chalo HC	-0.5025	0.0000	-0.7556615	-0.2493385
Dakano Elke HC	-0.4325	0.0040	-0.7229764	-0.1419589
Darge Bortolo HC	-0.2600	0.0790	-0.5504764	0.0305411
Deneba HC	-0.3517	0.0110	-0.6196988	-0.0836920
Gamada HC	-0.6525	0.0000	-0.9056615	-0.3993385
Gatira HC	-0.2325	0.1150	-0.5229764	0.0580411
Gebjiro HC	-0.2067	0.1280	-0.4746988	0.0613080
Gefo HC	-0.1264	0.3620	-0.4014892	0.1487593
Gesecha HC	-0.1550	0.2910	-0.4454764	0.1355411
Haro Kamise HC	-0.1425	0.2650	-0.3956615	0.1106615
Haro Kera HC	-0.4330	0.0030	-0.7081559	-0.1579074
Hurufa HC	-0.3175	0.0150	-0.5706615	-0.0643385
Kiltu Shibo HC	0.4100	0.0020	0.1568385	0.6631615
Korjo HC	-0.3192	0.0200	-0.5871988	-0.0511920
Kumbabe HC	0.0200	0.8750	-0.2331615	0.2731615
Lalo HC	-0.3592	0.0090	-0.6271988	-0.0911920
Meteso HC	-0.4792	0.0010	-0.7471988	-0.2111920
Mole HC	-0.0800	0.5840	-0.3704764	0.2105411
Sentema Kecha HC	-0.3000	0.0430	-0.5904764	-0.0094589
Sheki HC	-0.4517	0.0010	-0.7196988	-0.1836920
Yela sasach HC	-0.2200	0.1350	-0.5104764	0.0705411
Year 2	-0.1743	0.0080	-0.3015590	-0.0469467
Year 3	-0.1755	0.0160	-0.3168760	-0.0341584
Year 4	-0.2039	0.0050	-0.3452094	-0.0624918
Contant	0.9009	0.0000	0.7000745	0.7000745

Table 14. DiD Model 1 for Skilled Delivery Service Coverage

N =	94
Adj R-squared =	0.6609

	Model 1			
X	Coefficients (β)	P-value	[95% confidence intervals]	
inter_pbf_post	0.0605	0.4680	-0.1049646	0.2260343
Abdella HC	-0.4200	0.0050	-0.7088169	-0.1311831
Bilu Harsu HC	-0.8354	0.0000	-1.1497610	-0.5210412
Chalo HC	-1.1200	0.0000	-1.4088170	-0.8311831
Dakano Elke HC	-1.0204	0.0000	-1.3347610	-0.7060412
Darge Bortolo HC	-0.8729	0.0000	-1.1872610	-0.5585412
Deneba HC	-0.8603	0.0000	-1.1607050	-0.5598298
Gamada HC	-1.3075	0.0000	-1.5963170	-1.0186830
Gatira HC	-0.9529	0.0000	-1.2672610	-0.6385412
Gebjiro HC	-0.8228	0.0000	-1.1232050	-0.5223298
Gefo HC	-0.6331	0.0000	-0.9469285	-0.3193650
Gesecha HC	-0.8629	0.0000	-1.1772610	-0.5485412
Haro Kamise HC	-0.6025	0.0000	-0.8913169	-0.3136831
Haro Kera HC	-0.9931	0.0000	-1.3069290	-0.6793650
Hurufa HC	-0.8150	0.0000	-1.1038170	-0.5261831
Kiltu Shibo HC	0.0150	0.9180	-0.2738169	0.3038169
Korjo HC	-0.8628	0.0000	-1.1632050	-0.5623298
Kumbabe HC	-0.4825	0.0010	-0.7713169	-0.1936831
Lalo HC	-1.0053	0.0000	-1.3057050	-0.7048298
Meteso HC	-0.9503	0.0000	-1.2507050	-0.6498298
Mole HC	-0.7229	0.0000	-1.0372610	-0.4085412
Sentema Kecha HC	-0.9554	0.0000	-1.2697610	-0.6410412
Sheki HC	-0.9478	0.0000	-1.2482050	-0.6473298
Yela sasach HC	-0.9579	0.0000	-1.2722610	-0.6435412
Year 2	-0.1470	0.0350	-0.2837027	-0.0103893
Year 3	-0.0576	0.4750	-0.2175165	0.1023237
Year 4	-0.0776	0.3360	-0.2375165	0.0823237
Contant	1.4506	0.0000	1.2225710	1.6785480

Table A 15. DiD Model 2 for Skilled Delivery Service Coverage

N =	94
Adj R-squared =	0.6573

	Model 2			
X	Coefficients (β)	P-value	[95% confidence intervals]	
inter_pbfi_posti	0.0209	0.8480	-0.1970120	0.2388798
inter_pbfii_postii	0.0926	0.3630	-0.1091350	0.2944224
Abdella HC	-0.4200	0.0050	-0.7104091	-0.1295909
Bilu Harsu HC	-0.8057	0.0000	-1.1389520	-0.4724491
Chalo HC	-1.1200	0.0000	-1.4104090	-0.8295909
Dakano Elke HC	-0.9907	0.0000	-1.3239520	-0.6574491
Darge Bortolo HC	-0.8432	0.0000	-1.1764520	-0.5099491
Deneba HC	-0.8763	0.0000	-1.1837570	-0.5688871
Gamada HC	-1.3075	0.0000	-1.5979090	-1.0170910
Gatira HC	-0.9232	0.0000	-1.2564520	-0.5899491
Gebjiro HC	-0.8388	0.0000	-1.1462570	-0.5313871
Gefo HC	-0.6353	0.0000	-0.9508907	-0.3196841
Gesecha HC	-0.8332	0.0000	-1.1664520	-0.4999491
Haro Kamise HC	-0.6025	0.0000	-0.8929091	-0.3120909
Haro Kera HC	-0.9953	0.0000	-1.3108910	-0.6796841
Hurufa HC	-0.8150	0.0000	-1.1054090	-0.5245909
Kiltu Shibo HC	0.0150	0.9180	-0.2754091	0.3054091
Korjo HC	-0.8788	0.0000	-1.1862570	-0.5713871
Kumbabe HC	-0.4825	0.0010	-0.7729091	-0.1920909
Lalo HC	-1.0213	0.0000	-1.3287570	-0.7138871
Meteso HC	-0.9663	0.0000	-1.2737570	-0.6588871
Mole HC	-0.6932	0.0000	-1.0264520	-0.3599491
Sentema Kecha HC	-0.9257	0.0000	-1.2589520	-0.5924491
Sheki HC	-0.9638	0.0000	-1.2712570	-0.6563871
Yela sasach HC	-0.9282	0.0000	-1.2614520	-0.5949491
Year 2	-0.1331	0.0730	-0.2791689	0.0129045
Year 3	-0.0517	0.5260	-0.2138667	0.1104472
Year 4	-0.0717	0.3800	-0.2338667	0.0904472
Contant	1.4441	0.0000	1.2137590	1.6745170

Table A 16. DiD Model 1 for Immunization Service Coverage

N =	93
Adj R-squared =	0.886

	Model 1			
X	Coefficients (β)	P-value	[95% confidence intervals]	
inter_pbf_post	0.0979	0.0500	0.0000701	0.1957791
Abdella HC	-0.0800	0.3590	-0.2528293	0.0928293
Bilu Harsu HC	-0.6759	0.0000	-0.8637097	-0.4881772
Chalo HC	-1.0000	0.0000	-1.1728290	-0.8271707
Dakano Elke HC	-0.8784	0.0000	-1.0662100	-0.6906772
Darge Bortolo HC	-0.7709	0.0000	-0.9587097	-0.5831772
Deneba HC	-0.7940	0.0000	-0.9735836	-0.6143409
Gamada HC	-1.0000	0.0000	-1.1728290	-0.8271707
Gatira HC	-0.8759	0.0000	-1.0637100	-0.6881772
Gebjiro HC	-0.8815	0.0000	-1.0610840	-0.7018409
Gefo HC	-0.7648	0.0000	-0.9781857	-0.5513539
Gesecha HC	-0.8809	0.0000	-1.0687100	-0.6931772
Haro Kamise HC	-0.1525	0.0830	-0.3253293	0.0203293
Haro Kera HC	-0.3186	0.0010	-0.5063516	-0.1309356
Hurufa HC	-0.9775	0.0000	-1.1503290	-0.8046707
Kiltu Shibo HC	0.0900	0.3020	-0.0828293	0.2628293
Korjo HC	-0.8240	0.0000	-1.0035840	-0.6443409
Kumbabe HC	-0.0100	0.9080	-0.1828293	0.1628293
Lalo HC	-0.9065	0.0000	-1.0860840	-0.7268409
Meteso HC	-0.9890	0.0000	-1.1685840	-0.8093409
Mole HC	-0.5234	0.0000	-0.7112097	-0.3356772
Sentema Kecha HC	-0.9509	0.0000	-1.1387100	-0.7631772
Sheki HC	-0.8240	0.0000	-1.0035840	-0.6443409
Yela sasach HC	-0.7834	0.0000	-0.9712097	-0.5956772
Year 2	-0.0756	0.0660	-0.1563519	0.0050890
Year 3	-0.0720	0.1330	-0.1666716	0.0225785
Year 4	-0.1486	0.0020	-0.2415020	-0.0556954
Contant	1.0741	0.0000	0.9383892	1.2097490

Table A 17 DiD Model 2 for Immunization Service Coverage

N =	93
Adj R-squared =	0.8844

	Model 2			
X	Coefficients (β)	P-value	[95% confidence intervals]	
inter_pbfi_posti	0.1111	0.0910	-0.0180346	0.2401887
inter_pbfii_postii	0.0870	0.1550	-0.0336877	0.2075949
Abdella HC	-0.0800	0.3620	-0.2540909	0.0940909
Bilu Harsu HC	-0.6858	0.0000	-0.8850172	-0.4865983
Chalo HC	-1.0000	0.0000	-1.1740910	-0.8259091
Dakano Elke HC	-0.8883	0.0000	-1.0875170	-0.6890983
Darge Bortolo HC	-0.7808	0.0000	-0.9800172	-0.5815983
Deneba HC	-0.7885	0.0000	-0.9727218	-0.6042318
Gamada HC	-1.0000	0.0000	-1.1740910	-0.8259091
Gatira HC	-0.8858	0.0000	-1.0850170	-0.6865983
Gebjiro HC	-0.8760	0.0000	-1.0602220	-0.6917318
Gefo HC	-0.7659	0.0000	-0.9810458	-0.5508407
Gesecha HC	-0.8908	0.0000	-1.0900170	-0.6915983
Haro Kamise HC	-0.1525	0.0850	-0.3265909	0.0215909
Haro Kera HC	-0.3180	0.0010	-0.5070878	-0.1288320
Hurufa HC	-0.9775	0.0000	-1.1515910	-0.8034091
Kiltu Shibo HC	0.0900	0.3060	-0.0840909	0.2640909
Korjo HC	-0.8185	0.0000	-1.0027220	-0.6342318
Kumbabe HC	-0.0100	0.9090	-0.1840909	0.1640909
Lalo HC	-0.9010	0.0000	-1.0852220	-0.7167318
Meteso HC	-0.9835	0.0000	-1.1677220	-0.7992318
Mole HC	-0.5333	0.0000	-0.7325172	-0.3340983
Sentema Kecha HC	-0.9608	0.0000	-1.1600170	-0.7615983
Sheki HC	-0.8185	0.0000	-1.0027220	-0.6342318
Yela sasach HC	-0.7933	0.0000	-0.9925172	-0.5940983
Year 2	-0.0803	0.0680	-0.1667886	0.0062148
Year 3	-0.0738	0.1290	-0.1698318	0.0221517
Year 4	-0.1504	0.0020	-0.2445951	-0.0561131
Contant	1.0761	0.0000	0.9388326	1.2134080

Table A 18 Model 1 for Vitamin A Service Coverage

N =	90
Adj R-squared =	0.7455

	Model 1			
X	Coefficients (β)	P-value	[95% confidence intervals]	
inter_pbf_post	0.1081	0.1590	-0.0436246	0.2598367
Abdella HC	-0.0875	0.5070	-0.3498207	0.1748207
Bilu Harsu HC	-0.6436	0.0000	-0.9295203	-0.3576388
Chalo HC	-0.8575	0.0000	-1.1198210	-0.5951793
Dakano Elke HC	-0.9411	0.0000	-1.2270200	-0.6551388
Darge Bortolo HC	-0.8711	0.0000	-1.1570200	-0.5851388
Deneba HC	-0.9116	0.0000	-1.1846240	-0.6384822
Gamada HC	-0.8600	0.0000	-1.1223210	-0.5976793
Gatira HC	-0.9561	0.0000	-1.2420200	-0.6701388
Gebjiro HC	-0.8366	0.0000	-1.1096240	-0.5634822
Gefo HC	-0.2700	0.0440	-0.5323207	-0.0076793
Gesecha HC	-0.8861	0.0000	-1.1720200	-0.6001388
Haro Kamise HC	-0.2950	0.0280	-0.5573207	-0.0326793
Haro Kera HC	0.0519	0.7170	-0.2331492	0.3369049
Hurufa HC	-0.7885	0.0000	-1.0729910	-0.5039762
Kiltu Shibo HC	0.1000	0.4490	-0.1623207	0.3623207
Korjo HC	-0.8016	0.0000	-1.0746240	-0.5284822
Lalo HC	-0.8766	0.0000	-1.1496240	-0.6034822
Meteso HC	-0.8916	0.0000	-1.1646240	-0.6184822
Mole HC	-0.5886	0.0000	-0.8745203	-0.3026388
Sentema Kecha HC	-0.8961	0.0000	-1.1820200	-0.6101388
Sheki HC	-0.9441	0.0000	-1.2171240	-0.6709822
Yela sasach HC	-0.8411	0.0000	-1.1270200	-0.5551388
Year 2	0.0208	0.7400	-0.1037787	0.1453126
Year 3	0.0578	0.4340	-0.0889947	0.2045602
Year 4	-0.0311	0.6790	-0.1804689	0.1183011
Contant	0.9156	0.0000	0.7081669	1.1231000

Table A 19. DiD Model 2 for Vitamin A Service Coverage

N =	90
Adj R-squared =	0.7414

	Model 2			
X	Coefficients (β)	P-value	[95% confidence intervals]	
inter_pbfi_posti	0.1105	0.2720	-0.0886551	0.3097129
inter_pbfii_postii	0.1061	0.2570	-0.0794384	0.2916619
Abdella HC	-0.0875	0.5110	-0.3520081	0.1770081
Bilu Harsu HC	-0.6454	0.0000	-0.9491751	-0.3416183
Chalo HC	-0.8575	0.0000	-1.1220080	-0.5929919
Dakano Elke HC	-0.9429	0.0000	-1.2466750	-0.6391183
Darge Bortolo HC	-0.8729	0.0000	-1.1766750	-0.5691183
Deneba HC	-0.9106	0.0000	-1.1908620	-0.6302493
Gamada HC	-0.8600	0.0000	-1.1245080	-0.5954919
Gatira HC	-0.9579	0.0000	-1.2616750	-0.6541183
Gebjiro HC	-0.8356	0.0000	-1.1158620	-0.5552493
Gefo HC	-0.2700	0.0460	-0.5345081	-0.0054919
Gesecha HC	-0.8879	0.0000	-1.1916750	-0.5841183
Haro Kamise HC	-0.2950	0.0290	-0.5595081	-0.0304919
Haro Kera HC	0.0520	0.7190	-0.2354773	0.3394918
Hurufa HC	-0.7885	0.0000	-1.0753490	-0.5015873
Kiltu Shibo HC	0.1000	0.4530	-0.1645081	0.3645081
Korjo HC	-0.8006	0.0000	-1.0808620	-0.5202493
Lalo HC	-0.8756	0.0000	-1.1558620	-0.5952493
Meteso HC	-0.8906	0.0000	-1.1708620	-0.6102493
Mole HC	-0.5904	0.0000	-0.8941751	-0.2866183
Sentema Kecha HC	-0.8979	0.0000	-1.2016750	-0.5941183
Sheki HC	-0.9431	0.0000	-1.2233620	-0.6627493
Yela sasach HC	-0.8429	0.0000	-1.1466750	-0.5391183
Year 2	0.0199	0.7670	-0.1136867	0.1534901
Year 3	0.0574	0.4440	-0.0916735	0.2065491
Year 4	-0.0314	0.6800	-0.1831310	0.1202785
Contant	0.9160	0.0000	0.7058296	1.1262140

Appendix 9: LiST inputs: Coverage rates for four scenarios (Model 1)

Table A 20. Year effects (control group trends) from Model 1 difference-in-difference estimates used to estimate comparison coverage rates

Indicators	Year 2	Year 3	Year 4
ANC1	-7.21%	-7.85%	-3.42%
ANC4	-18.07%	-17.83%	-20.66%
Skilled Deliveries	-14.70%	-5.76%	-7.76%
Full Immunization of infants <1	-7.56%	-7.20%	-14.86%
Vitamin A for children <5	2.08%	5.78%	-3.11%

Table A 21. LiST inputs: Service coverage for PBF and 'no-PBF' scenarios for Phase I facilities

Indicators	Scenario	Pre-PBF	Post-PBF		
		Year 1	Year 2	Year 3	Year 4
ANC1	PBF Phase I	24.79%	30.02%	36.10%	36.93%
	Comparison	24.79%	17.58%	16.94%	21.36%
ANC4	PBF Phase I	65.27%	40.12%	43.86%	50.54%
	Comparison	65.27%	47.19%	47.44%	44.61%
Skilled Deliveries	PBF Phase I	57.69%	35.77%	53.74%	64.00%
	Comparison	57.69%	42.98%	51.93%	49.93%
Full Immunization of infants <1	PBF Phase I	27.42%	24.45%	31.27%	28.86%
	Comparison	27.42%	19.86%	20.22%	12.56%
Vitamin A for children <5	PBF Phase I	8.69%	15.69%	26.47%	21.24%
	Comparison	8.69%	10.77%	14.47%	5.58%

Note. The comparison or 'counterfactual' scenario is constructed using the fixed effects from the difference-in-differences estimates from Model 1

Table A 22. LiST inputs: Service coverage for PBF and 'no-PBF' scenarios for Phase II facilities

Indicators	Scenario	Pre-PBF	Post-PBF	
		Year 2	Year 3	Year 4
ANC1	PBF Phase II	9.46%	22.72%	39.90%
	Comparison	9.46%	8.82%	13.25%
ANC4	PBF Phase II	40.40%	33.58%	39.08%
	Comparison	40.40%	40.65%	37.81%
Skilled Deliveries	PBF Phase II	42.83%	44.76%	65.53%
	Comparison	42.83%	51.78%	49.78%
Full Immunization of infants <1	PBF Phase II	17.02%	17.39%	19.74%
	Comparison	17.02%	17.38%	9.72%
Vitamin A for children <5	PBF Phase II	6.15%	12.58%	19.19%
	Comparison	6.15%	9.85%	0.97%

Note. The comparison or 'counterfactual' scenario is constructed using the fixed effects from the difference-in-differences estimates from Model 1